

# Identifying Correlations in Precessing Gravitational-Wave Signals with Machine Learning

KAREN KANG,<sup>1</sup> ADVISORS: SIMONA MILLER,<sup>2</sup> KATERINA CHATZIOANNOU,<sup>2</sup> AND DEBORAH FERGUSON<sup>3</sup>

<sup>1</sup>*Department of Physics and Astronomy, Amherst College, Amherst, MA 01002, USA*

<sup>2</sup>*Department of Physics, California Institute of Technology, Pasadena, California 91125, USA*

<sup>3</sup>*Center for Gravitational Physics and Department of Physics, The University of Texas at Austin, Austin, TX 78712*

## ABSTRACT

Binary binary hole (BBH) spins provide unique and important insights into the formation environments, evolutionary history, and dynamics of these objects. We would like to gain a better understanding of merger-dominated signals for highly massive highly spinning BBH systems, which are prone to spurious measurements due to their very short duration and low bandwidth. Astrophysical parameters from gravitational wave (GW) sources are extracted by match-filtering signals with numerical relativity (NR) waveform templates. The degeneracies in waveforms, where dissimilar parameters yield similar waveforms, further complicates source identification. Using machine learning, we can visualize these degeneracies in the 14-dimensional BBH parameter space and develop models to quantify parameter correlations. We also propose enhancing existing mismatch-prediction neural networks with higher order modes and precession effects, thereby refining our ability to model these degeneracies. The results produced by this network will inform us about the measurability of spin parameters from inferred waveform signals of highly massive, precessing BBHs.

## 1. INTRODUCTION

GW190521 ( $M \sim 150\odot$ ) is the heaviest BH binary detected to date and one of the few BBHs measured to be highly precessing. It is the first strong observational evidence of intermediate-mass BH (IMBH), which is believed to be the missing link for explaining the formation of supermassive BHs (Abbott et al. 2020). The detected waveform is dominated by the merger phase where effects of precession remain elusive. In the upcoming LIGO’s fourth observing run (O4) as well as in the advent of the space-based Laser-Interferometer Space Antenna (LISA) and the ground-based Einstein Telescope (ET), we expect to observe large sample of events similar to GW190521 as a result of increased detector sensitivity. Spin configurations are indicative of the compact progenitor’s orbital dynamics and therefore help illuminate the formation channels of BH mergers in the pair-instability (PI) mass gap and aid population modeling of BBHs (Abbott et al. 2020; Mandel & Farmer 2022).

We would like to respond to signals from such systems with maximal accuracy, which requires a thorough phenomenological understanding of the measurability of spin parameters from inferred waveforms.

## 2. OBJECTIVES

The objective of the project is to identify degeneracies in the parameter space for highly massive, precessing

BBH systems using machine learning. I will 1) determine whether a certain set of parameters can be recovered from detected waveform, 2) investigate the correlations between degenerate parameters, and 3) quantify and produce visualizations of such correlations. Understanding the degeneracies and correlations between spin measurements will inform us as to which spin parameters are actually independently measurable in GW data.

## 3. BACKGROUND & APPROACH

### 3.1. Theoretical Modeling

My project will be largely based on the neural network `mismatch_prediction` presented in (Ferguson 2023). The model is currently trained on NR waveforms from the SXS GW catalog. The network predicts the mismatch of the GW emitted by two BBH systems with initial input parameters  $\lambda = \eta, \mathbf{a}_1, \mathbf{a}_2$ , consisting of the symmetric mass ratio  $\eta$

$$\eta = m_1 m_2 / (m_1 + m_2)^2 \quad 0 \leq \eta \leq 0.25 \quad (1)$$

and the dimensionless spin vectors  $\mathbf{a}_1, \mathbf{a}_2$ , which are 3-dimensional vectors with  $x, y, z$  components.

$$\mathbf{a} = \frac{\mathbf{J}}{m^2} \quad 0 < a < 1$$

The model uses a mismatch metric  $\mathcal{MM}$  to evaluate how different a waveform, generated by specific initial

parameters, is from a preexisting waveform. Mathematically, it's represented as:

$$\mathcal{MM} = 1 - \max_{t, \phi} \mathcal{O}[h_1, h_2] \equiv 1 - \max_{t, \phi} \frac{\langle h_1 | h_2 \rangle}{\sqrt{\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle}}$$

where we take the normalized inner product of the frequency domain strains  $h_1, h_2$ , optimized over time and phase, to be the overlap of two waveforms,

$$\langle h_1 | h_2 \rangle = 2 \int_{f_0}^{\infty} \frac{h_1^* h_2 + h_1 h_2^*}{S_n} df$$

and  $S_n$  the one-sided power spectral density of the detector. The metric  $\mathcal{MM}$  is normalized to 1, with  $\mathcal{MM} = 0$  representing two completely identical waveforms. In the existing network, the mismatch is computed on a flat noise curve at  $f_{ref}$  using `PyCBC` and `LALSimulation`. It is trained with  $l = 2, m = 2$  modes for systems of  $M_{tot} = 1M_{\odot}$ , corresponding to  $f_{ref} = 1840$  Hz. Since the frequency of the emitted gravitational wave scales inversely with the total mass of the system, heavy binaries merge at lower frequencies, which leaves fewer cycles in the sensitive band of ground-based detectors. I will be adjusting the frequency cutoff to adapt the network specifically to highly massive, precessing BBHs with fewer observable cycles (currently `mass = 100 M⊙`, we might use 150 for the training data since we are interested in 190521-like systems). To account for effects of precession, I will be computing mismatches between these waveforms using `PyCBC`, which requires optimization over both time and coalescence when the system is phase-on. We will make use waveform model `NRSur7dq4`, which is the only model calibrated to numerical simulations of precessing BH binaries, including the full six-dimensional spin degrees of freedom and higher harmonics ([Biscoveanu et al. 2021](#); [Varma et al. 2019](#)). The waveforms will be split randomly into training, development, and test sets with ratios of 0.8/0.1/0.1 respectively following how the network was trained previously.

### 3.2. Degeneracy Mapping

Since it is difficult to infer component spin parameters from GW, we use  $\chi_{eff}, \chi_p$  to characterize the signal detected. The effective inspiral spin  $\chi_{eff}$  is the mass-weighted average of the components of the BBH's spins aligned with the system's angular momentum

$$\chi_{eff} = \frac{m_1 a_1 \cos(\theta_1) + m_2 a_2 \cos(\theta_2)}{m_1 + m_2} \quad -1 \leq \chi_{eff} \leq 1 \quad (2)$$

The effective precession spin can be modeled with a single parameter  $\chi_p$ , which is defined to be the mass-weighted in-plane spin component that contributes to

precession of the orbital plane at some (arbitrary) instant during the inspiral phase ([Schmidt et al. 2015](#)).

$$\chi_p = \max \left( a_1 \sin(\theta_1), \frac{4m_2^2 + 3m_1 m_2}{4m_1^2 + 3m_1 m_2} a_2 \theta_2 \right) \quad 0 < \chi_p < 1 \quad (3)$$

The GW waveforms for systems with precession effects have richer morphology.

Expectedly, characterizing an extremely complex GW waveform with few parameters results in degeneracy in the parameter space (i.e.  $\mathcal{MM} \ll 0.05$ ). [Fig. 9a](#) provides a visualization of the degeneracies between the parameters  $\chi_{eff}$  and  $\eta$  in the case where BH spins are aligned/anti-aligned with the orbital angular momentum, and figure shows the degeneracy when precession is included.

## 4. CURRENT PROGRESS

Since the data set is high dimensional, I started with experimenting various methods to visualize degeneracy correlations in 2 dimensions. Once I got a good grasp on manipulating the data, I started testing different interpolation schemes and statistical models (See [Preliminary Results](#) for details) to identify the degeneracy in the parameter space. I have been working with non-precessing GW waveforms and correlations with two parameters ( $\eta$  and  $\chi_{eff}$ ). As of now, I have a working code prototype that can “walk” through the degenerate region of the 2D parameter space by applying Bayesian Gaussian Mixture Model / Principal Component Analysis to the data points. I have tested the script in 3 Dimensions ( $q, a_1, a_2$ ). It seems that the correlation recovered is consistent with the expected correlation in 2D ( $\eta$  and  $\chi_{eff}$ ); however, the recovered path in 3D does not span the 2D parameter space.

We initially planned to use SXS catalog NR waveforms to compute mismatches for higher order modes and thought we had a code that worked by aligning inclination of the coalescence and then optimizing over phase to maximize the overlap. Nonetheless, there turns out to be errors executing the code with some systems. In the interest of time, we opt for using surrogate model `NRSur7dq4`, which would allow us to control most variables and avoid sources of errors, to generate training data for the updated network. If time permits, we hope to get preliminary results on running the updated network with 190521 posteriors.

## 5. METHOD OUTLINE

To identify the correlations in the BBH parameter space, we design an algorithm that finds local variance with respect to the injected reference simulation. Given

the local variance, algorithm would decided on the placement of the next injection to re-generate the parameter space with respect to the updated injection and recalculate the variance.

### 5.1. Model Selection

We evaluated various statistical models for discerning correlations within the parameter space (detailed in 6). Notably, the Gaussian Mixture Model (GMM) and Principal Component Analysis (PCA) emerged as the most suitable due to their capability to pinpoint 'principal directions' of movement. This is achieved either by computing the eigenvectors and eigenvalues of the covariance matrix (in GMM) or by transforming data points along principal components (in PCA). We verified the effectiveness of these methods on dummy 2D data, supplemented with a Gaussian value in the third dimension, using a defined covariance matrix and mean with noise introduced. The fits are shown in Fig. 2. Fig. 1a and 1b show that the results from the two methods are consistent (near identical) for `n_components = 1` for GMM and `n_components = 2` for PCA in a two-dimensional parameter space.

### 5.2. Walker Algorithm

The algorithm is outlined as [Degeneracy Mapping Walker](#) with functions outlined in 1a.

---

**Algorithm 1** Degeneracy Mapping in  $n$ -Dimensional Parameter Space based on GMM Eigendecomposition

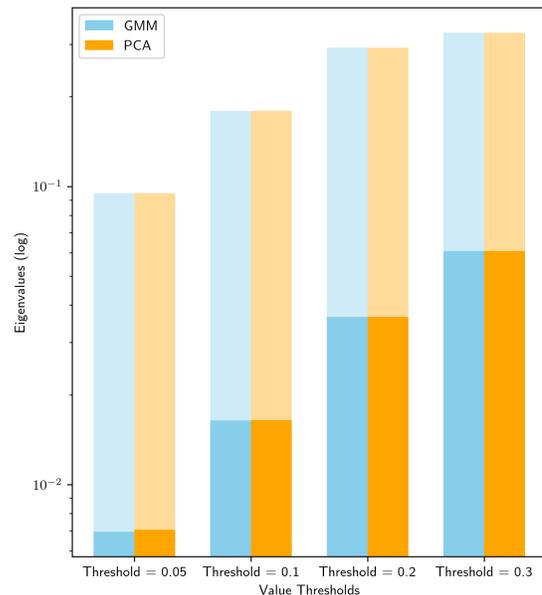
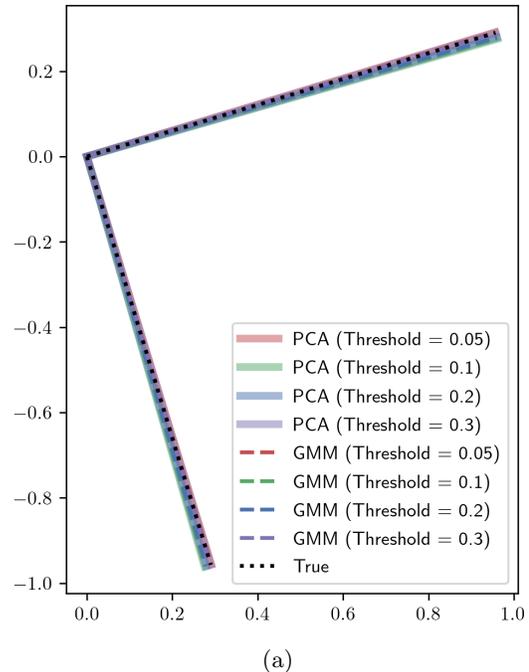
---

```

1: Initialization:
2: Set initial point as start from given or default values
3: Set initial threshold and define bounds for the space
4: Initialize lists: eigenvalues, eigenvectors, points, steps
5: Set max iterations if required
6: while start is within bounds do
7:   Display current reference point
8:   Get parameter space using external function
9:   filtered_data ← FILTERDATA(threshold, data)
10:  fitted_model ← FITMODEL(filtered_data)
11:  direction ← DETERMINEMOVEMENT(fitted_model)
12:  Propose a new point in the parameter space based on
     direction
13:  if proposed point is within bounds then
14:    Update start
15:  else
16:    Exit loop
17:  end if
18:  Calculate and display predicted mismatches
19:  if mismatch crosses threshold or max iterations
     reached then
20:    Exit loop
21:  end if
22: end while

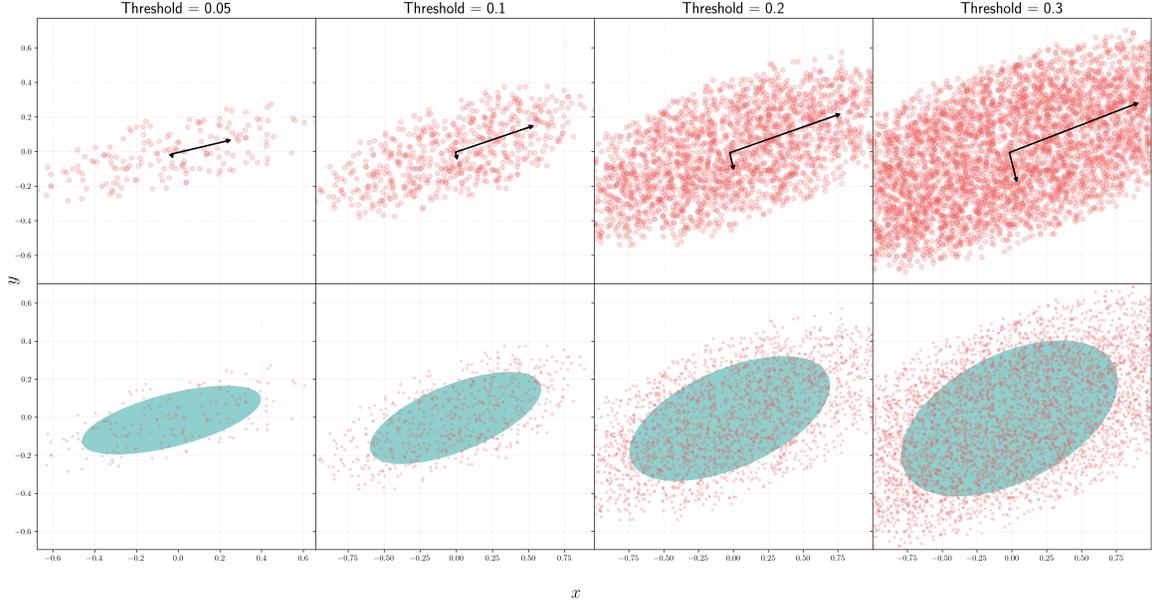
```

---



**Figure 1:** Comparisons of GMM and PCA results. (a) Eigenvectors. (b) Eigenvalues.

Since I used GMM for writing the first working prototype of the walker algorithm in 2D, I will continue to



**Figure 2:** Comparison between modeling the distribution with PCA (top) and GMM (bottom): principal components in black arrows, scaled for best visualization, means and covariance plotted as ellipse.

use GMM in higher dimensions for consistency unless the method fails.

---

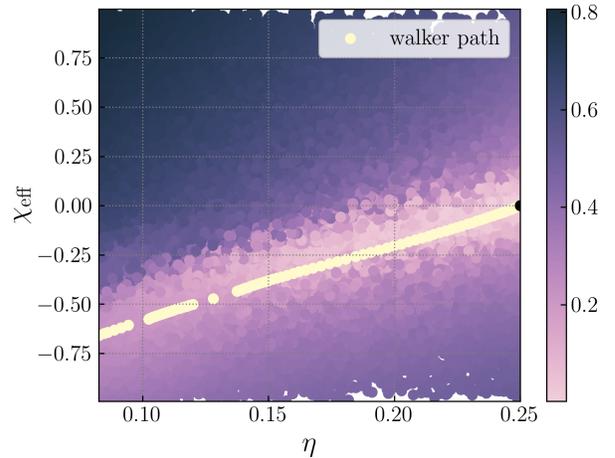
**Algorithm 1a** Supporting Procedures for the Main Algorithm

---

- 1: **procedure** FILTERDATA(threshold, data)
  - 2:   Use *threshold* to filter data points
  - 3:   **while** sample size below criteria and threshold not exceeded **do**
  - 4:     Adjust *threshold* and re-filter
  - 5:   **end while** **return** filtered data
  - 6: **end procedure**
  - 7: **procedure** FITMODEL(data)
  - 8:   Fit Gaussian Mixture Model on data
  - 9:   Save model's mean to global list **return** fitted model
  - 10: **end procedure**
  - 11: **procedure** DETERMINEMOVEMENT(Model)
  - 12:   Calculate and order eigenvectors and eigenvalues of Model's covariance
  - 13:   Construct direction vector from primary eigenvectors **return** direction vector
  - 14: **end procedure**
- 

## 6. PRELIMINARY RESULTS

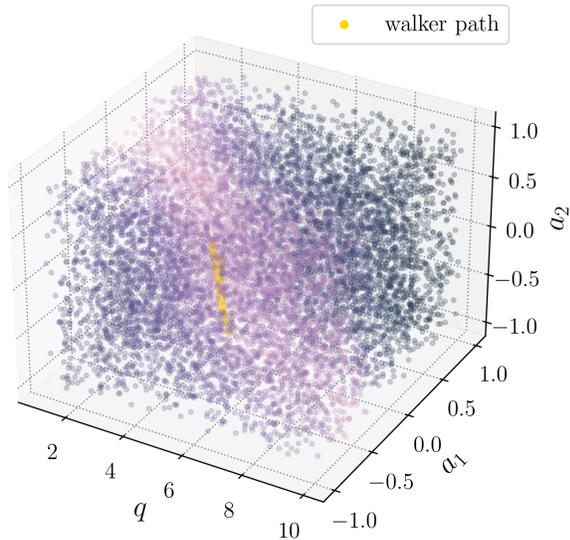
Using the iterative walker algorithm, we are able to map out the degenerate spaces based on the path taken by the walker. Fig. 3 shows the path calculated in 2D  $\eta, \chi_{\text{eff}}$  space and see fig. for the complete record of the evolution of the parameter space and the fitted Gaussian in Fig. 15.



**Figure 3:** Path calculated by the walker algorithm for starting reference point at  $\lambda = [0.25, 0, 0, 0, 0, 0, 0]$ . The background data points are simulations with respect to the starting reference point.

It is worth noting that in the 2-dimensional case, we are able to compute the the step to take by taking the direction of the largest eigenvector and the means of the Gaussian coincides with the injected simulation. The same method would fail when additional degrees of freedom is added— the changes in the parameter space can no longer be captured by a single eigenvector. Instead, we combined the primary eigenvectors (discarding the smallest eigenvector) to calculate the direction vector.

We successfully identified a path that appears credible in three dimensions in Fig. 4, 5. To verify that the path aligns with the  $\chi_{\text{eff}}$  vs.  $\eta$  correlation, we reverse plotted three points (table 1) calculated from the two-dimensional parameter space.



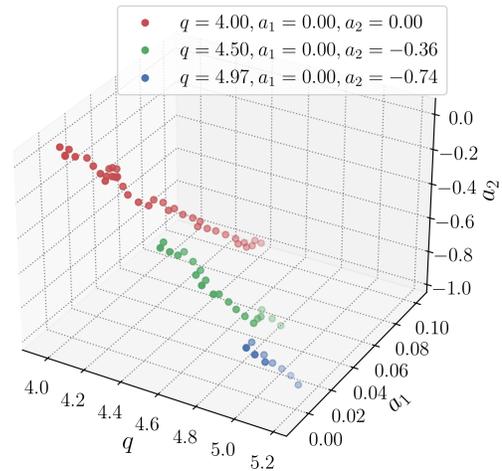
**Figure 4:** Path calculated by walker in 3D  $(q, a_1, a_2)$  with starting reference simulation at  $\lambda = [0.16, 0, 0, 0, 0, 0, 0]$ . Two possible directions exist when walking along the combined eigenvectors: positive and negative. In the plot, we only display the negative direction.

$\lambda$	$\eta$	$\chi_{\text{eff}}$	$q$	$a_1$	$a_2$
1	0.16	0.0	4.0	0	0
2	0.14881366	-0.066	4.497	0	-0.361
3	0.139	-0.124	4.972	0	-0.740

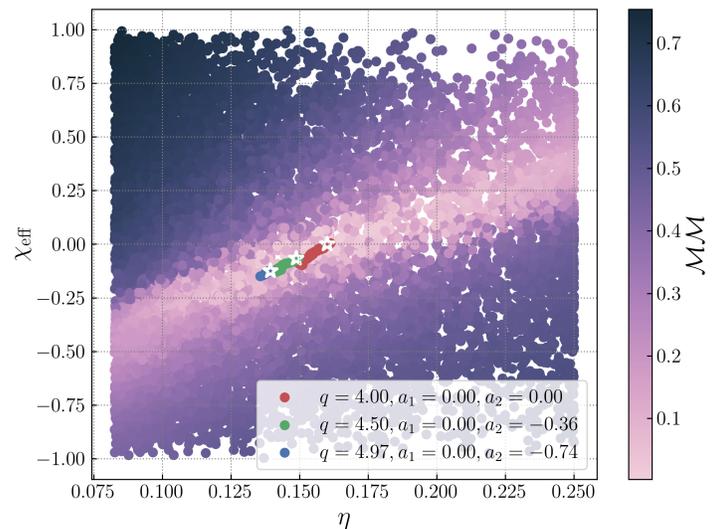
**Table 1:** Simulation Points for  $\eta$ ,  $\chi_{\text{eff}}$ , and  $q, a_1, a_2$ . We set  $a_1 = 0$  and calculate  $a_2$  based on  $\chi_{\text{eff}}$ . This is not the best practice, and I plan to redo this after submitting this preliminary result in my report.

We found that although the correlations in three dimensions align with the theoretical  $\chi_{\text{eff}}, \eta$  relation, it is difficult to characterize a transformed 2D space as the walker path does not span the space effectively (Fig. 6). Fig. 6 shows the paths plotted in two dimensions. We can observe that the paths plotted from degenerate starting reference simulation have similar topologies, subject to scaling, rotation, and translation, which is likely related to how the parameters are transformed

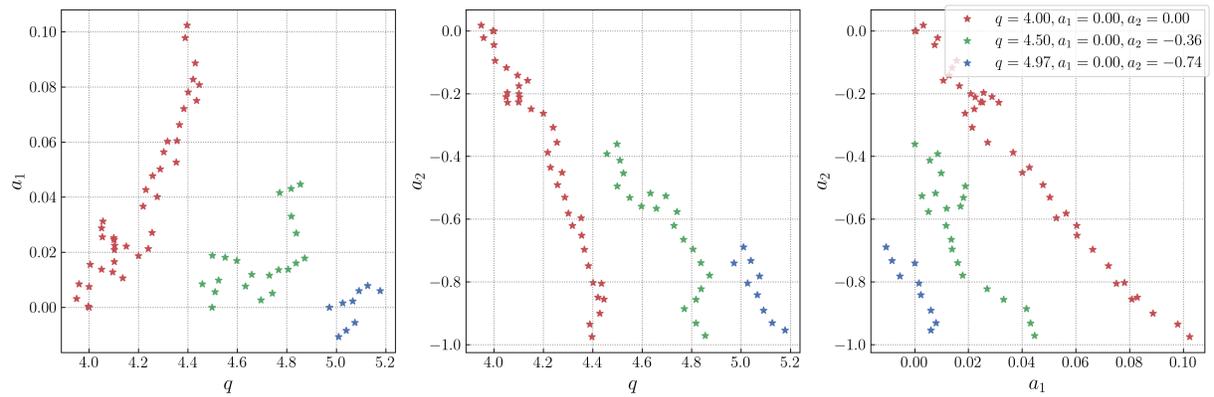
into the  $\chi_{\text{eff}}, \eta$  space. We would be addressing this issue with the remaining time of the program and hopefully generalize to higher dimensions so that we can fully characterize the spin parameters of BBH (including precession, see Fig. 8b, 11 for what the space looks like).



**Figure 5:** 3D walker path for degenerate simulations to  $\lambda = [0.16, 0, 0, 0, 0, 0, 0]$



**Figure 6:** 3D walker path transformed to  $\chi_{\text{eff}}$  vs.  $\eta$  2D parameter space



**Figure 7:** 2D projection of 3D walker paths: reference points chosen at along the 2D degeneracy line with starting reference simulation at  $\lambda = [0.16, 0, 0, 0, 0, 0, 0]$

## APPENDIX

## WORK PLAN: UPDATED 08/04/2023

**Before arrival:** Start background reading on project. Download required software and libraries. Familiarize myself with the mismatch-prediction model by [Ferguson \(2023\)](#) and understand the working principles of the APIs used (TensorFlow and Keras).

**Week 1-2:** Orientation; Reproducing existing results. Write plotting routines. Investigate algorithms for identifying degeneracies in parameter space: Hessian (failed), KDE (in progress).

**Week 3-5:** Trip to LHO; report I. Finished KDE plots (failed) and tested Gaussian Mixture Model for mapping out the parameter space. Experimented with Principal Component Analysis and binning data. Mapped out degeneracy contours for the binned data.

**Week 6-8:** Compute mismatches for existing waveforms; report II. Compare methods accuracy in recovering correlations. Verify correlation recovered in 3D with 2D. Prepare training data to retrain the network on high mass, precessing systems. Test methods with precessing systems in 2,2 mode.

**Week 9-10:** Follow-up with interesting results; *clean up scripts!!*; final report.

## A. MISMATCH.PREDICTION

The training data for the published model at [https://github.com/deborahferguson/mismatch\\_prediction](https://github.com/deborahferguson/mismatch_prediction) consists of BBH systems with symmetric mass ratios  $0.0826 \leq \eta \leq 0.25$  and spin magnitudes  $0 \leq a_1, a_2 \leq 0.9695$  in various directions. For this project, we sample points in the parameter space where the mass ratio  $1 \leq q \leq 10$  with  $q \equiv m_1/m_2, m_2 \leq m_1$  and  $0 \leq a_1, a_2 \leq 1$ . By setting the secondary mass of the system to 1, we have from equations 1, 2

$$\eta = \frac{q}{(q+1)^2} \quad (\text{A1})$$

$$\chi_{\text{eff}} = \frac{qa_1 + a_2}{q+1} \quad (\text{A2})$$

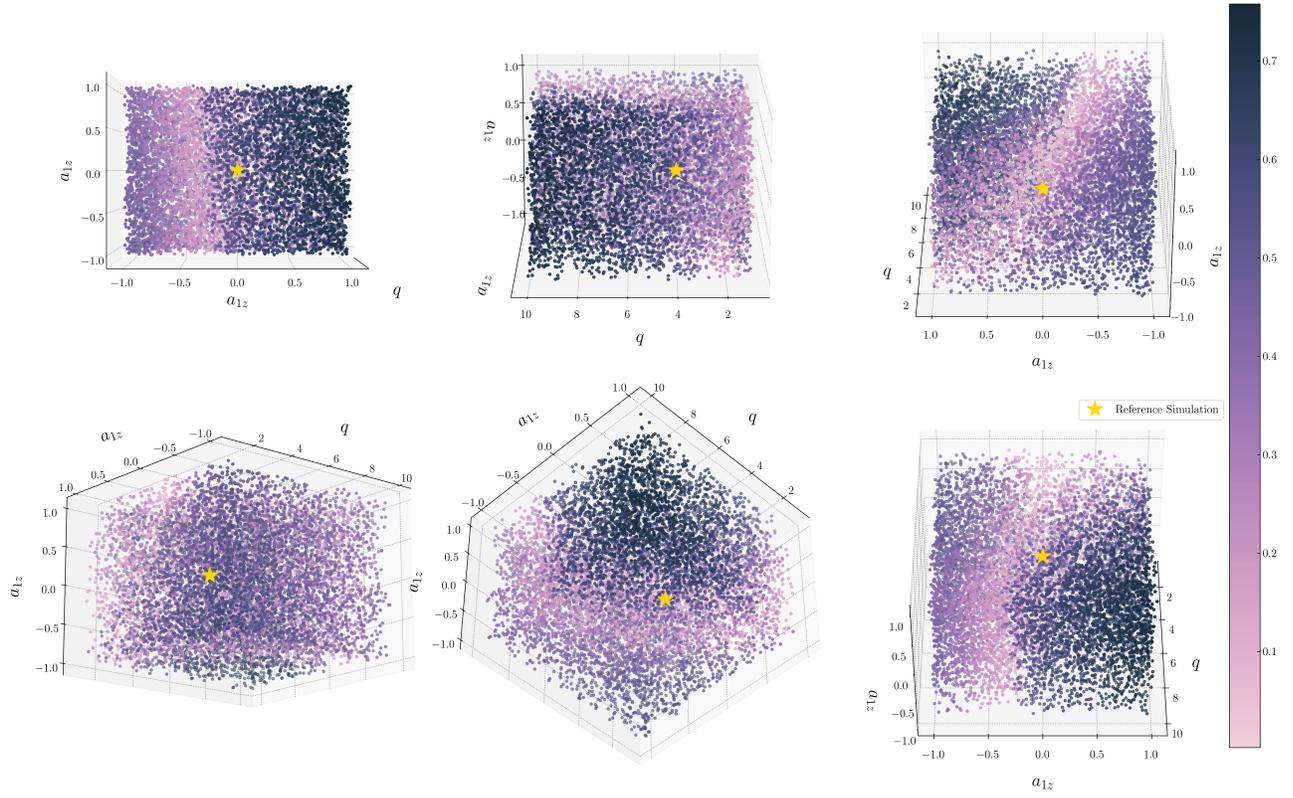
In terms of the mass and spin parameters, we can rewrite the effective precession parameter  $\chi_p$  as

$$\chi_p = \max \left( \sqrt{a_{1x}^2 + a_{1y}^2}, \frac{4+3q}{4q^2+3q} \sqrt{a_{2x}^2 + a_{2y}^2} \right) \quad (\text{A3})$$

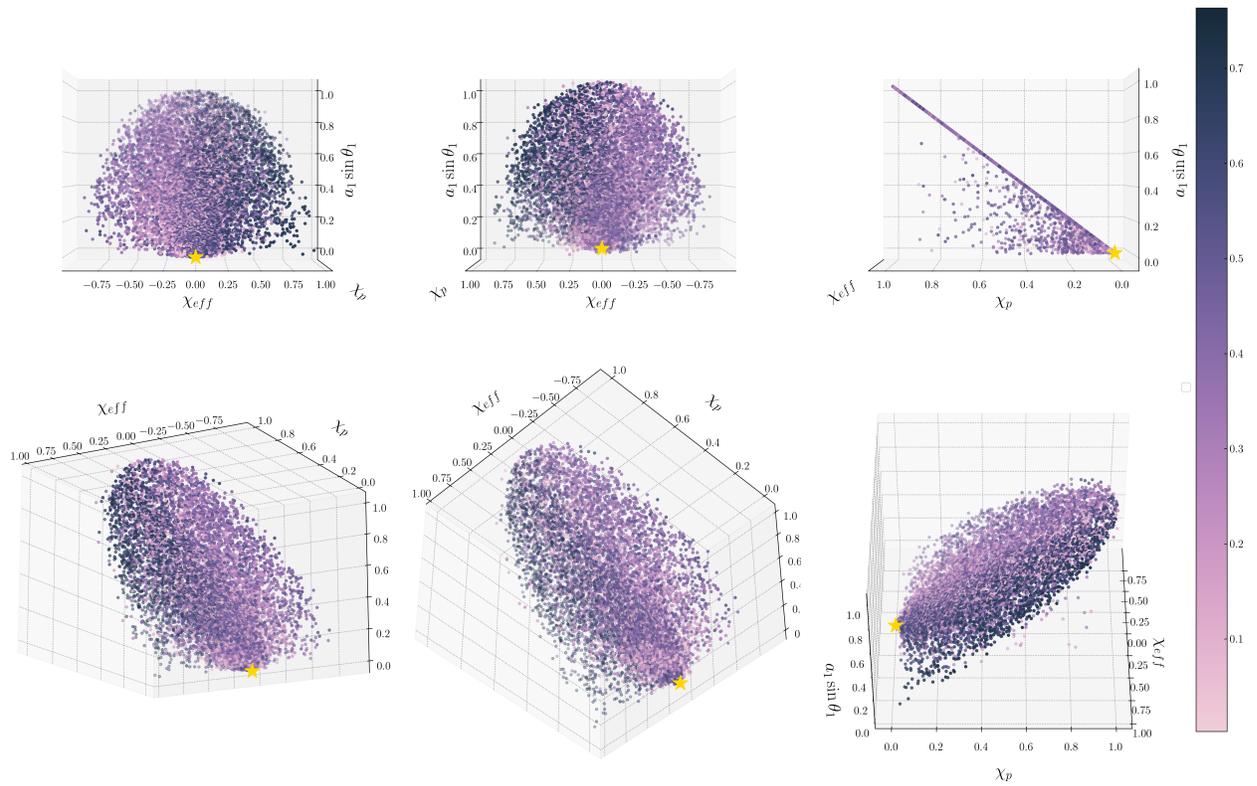
We have not yet included in effects of precession ( $\chi_p = 0$ ) and are considering systems where spin vectors are aligned or antialigned with the orbital angular momentum ( $a_{1x} = a_{1y} = a_{2x} = a_{2y} = 0$ ) in the data analysis process. Unless indicated otherwise, the degeneracy visualization plots are generated by uniformly sampling over  $q, a_{1z}, a_{2z}$ . A key challenge we encounter in this project is that our modeling of the effective spin parameter  $\chi_{\text{eff}}$  and the symmetric mass ratio  $\eta$  is not uniform, due to the nature of our sampling in terms of the spin magnitudes  $a_1, a_2$  and mass ratio  $q$ . These parameters represent ‘physical’ characteristics of the binary black hole (BBH) systems, but their transformation into the derived quantities  $\eta$  and  $\chi_{\text{eff}}$  introduces non-uniformities in the distribution of the latter, which impacts the reliability of using density-based data analysis methods such as [kernel density estimate](#).

## B. VISUALIZING DEGENERACIES

In the simplest case that we are considering,  $\mathcal{MM}$  is an output from three input variables ( $q, a_1, a_2$ ) when a reference point ( $q', a'_1, a'_2$ ) is chosen in the three-dimensional parameter space. To map out the degeneracies, we sample 10,000 random points in the parameter for each reference point and assign a color map to the mismatch values associated with each point, as shown in [fig. 8a](#). When precession is added, we are able to plot the distributions shown in [fig. 8b](#).



(a) Simulations without precession (aligned spins only)

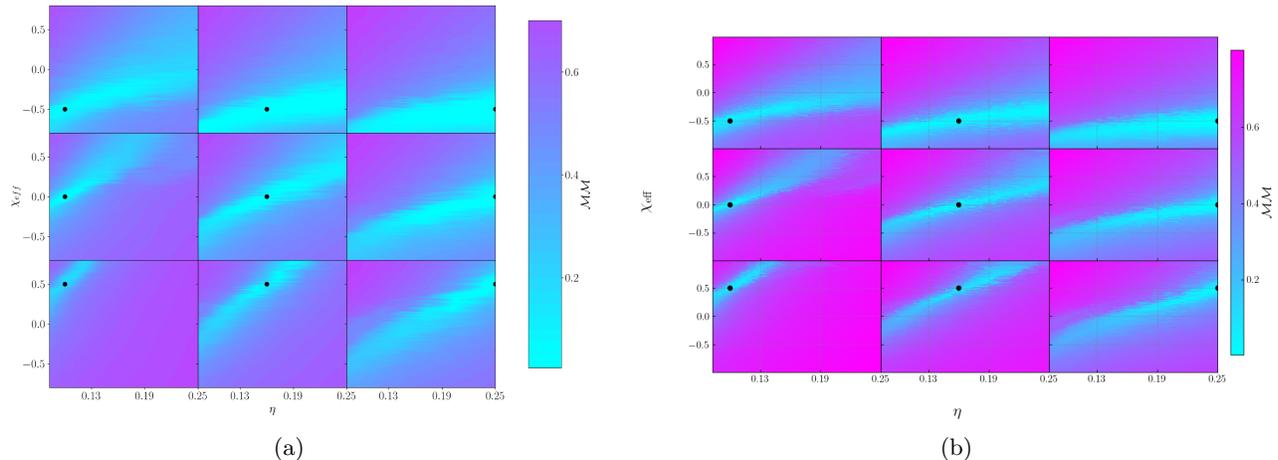


(b) Simulations with precession

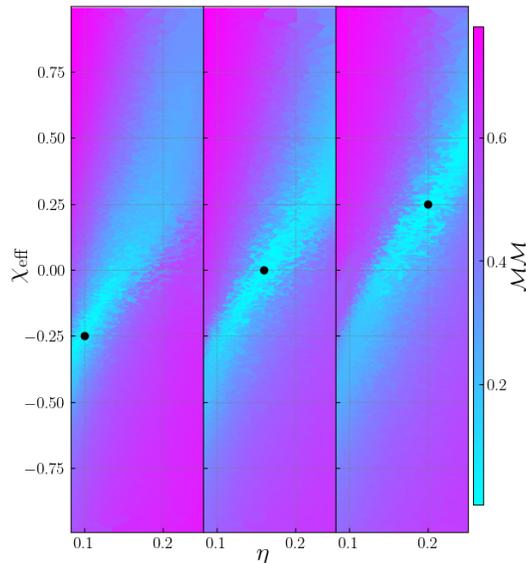
**Figure 8:** Visualizations of the three-dimensional parameter space with respect to reference simulation with  $q = 4$ ,  $a_1 = a_2 = 0$  or  $\lambda = [0.16, 0, 0, 0, 0, 0]$ . Reference simulation indicated with gold star.

### B.1. Mass-Effective Spin Degeneracy

By transforming the coordinate system of the parameter space, we observe the approximate linear relationship between  $\chi_{\text{eff}}$  and  $\eta$  as expected. In fig. 9, we were able to recreate the features of the degeneracy plot from (Ferguson 2023) subjected to differences in color scheme and figure scaling. We further experimented by varying the reference points along the line where the  $\chi_{\text{eff}}$  and  $\eta$  are most degenerate (See appendix A. of Ng et al. (2018)). The resulting plot is shown in fig. 10, which maps very similar space.



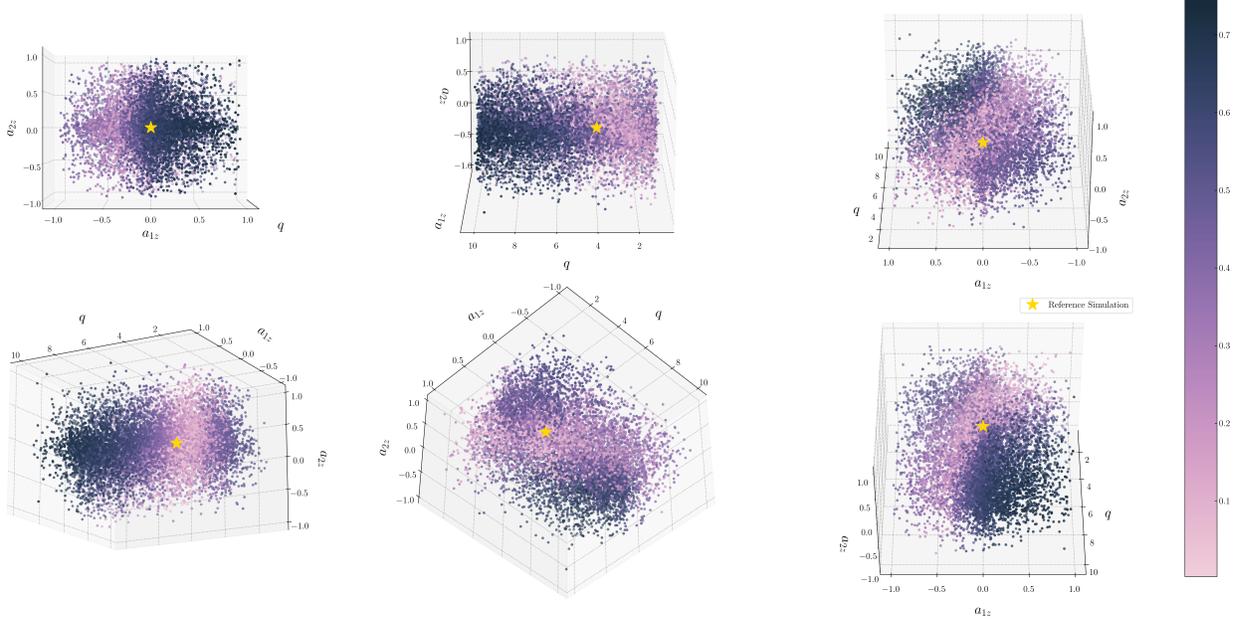
**Figure 9:** Mismatch between waveform of reference binary (black dot) and of other binaries throughout the parameter space plotted as  $\chi_{\text{eff}}$  vs.  $\eta$ . From left to right, the reference simulation has  $\eta = 0.1, 0.16, 0.25$ . From top to bottom, the reference has  $\chi_{\text{eff}} = -0.5, 0, 0.5$ . The original plot from Ferguson (2023) is shown on the left, and the recreation of the plot is shown on the right. Each subplot in the figure below is interpolated using the `nearest` method in `scipy.interpolate.griddata`.



**Figure 10:** Reference simulations chosen at  $\chi_{\text{eff}} = -0.25, 0, 0.25$  and  $\eta = 0.1, 0.16, 0.2$  along the line  $\chi_{\text{eff}} = 4.934\eta - 0.757$  where the two parameters are approximately degenerate. The degeneracy exhibited is globally consistent for the points chosen with some local variations due to SNR.

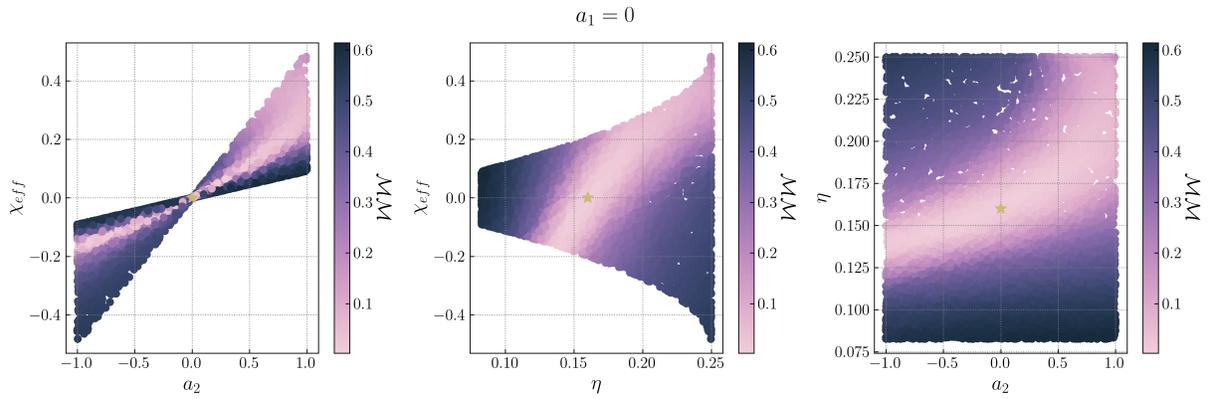
### B.2. Spin Degeneracies

In precessing Binary Black Hole (BBH) systems, the parameter space spans seven dimensions, encapsulating in-plane spins, the effective spin  $\chi_{\text{eff}}$ , precession parameters  $\chi_p$ , and the symmetric mass ratio  $\eta$ . At present, our focus is on refining our algorithm to discern correlations in three or more dimensions, particularly translating the  $q, a_1, a_2$  parameter space into  $\eta$  and  $\chi_{\text{eff}}$  representations. Consequently, we have momentarily sidelined the analysis of in-plane spins. However, preliminary observations, as illustrated in Fig. 11, suggest intriguing correlations when precession is incorporated.



**Figure 11:** Parameter space with respect to reference simulation at  $\lambda = [0.16, 0, 0, 0, 0, 0, 0]$  including systems with precession

In the no-precession case, we explored the spin degeneracies by setting  $a_1$  and varying the initial mass  $q$  and secondary spin component  $a_2$  as illustrated in Fig. 12. The resulting parameter spaces for  $a_1 > 0$  and  $a_1 < 0$  are depicted in Figs. 13 and 14, respectively. It is easily to identify qualitatively that a positive primary spin component better constrains the parameters.



**Figure 12:** Parameter space with  $a_{1z} = 0$

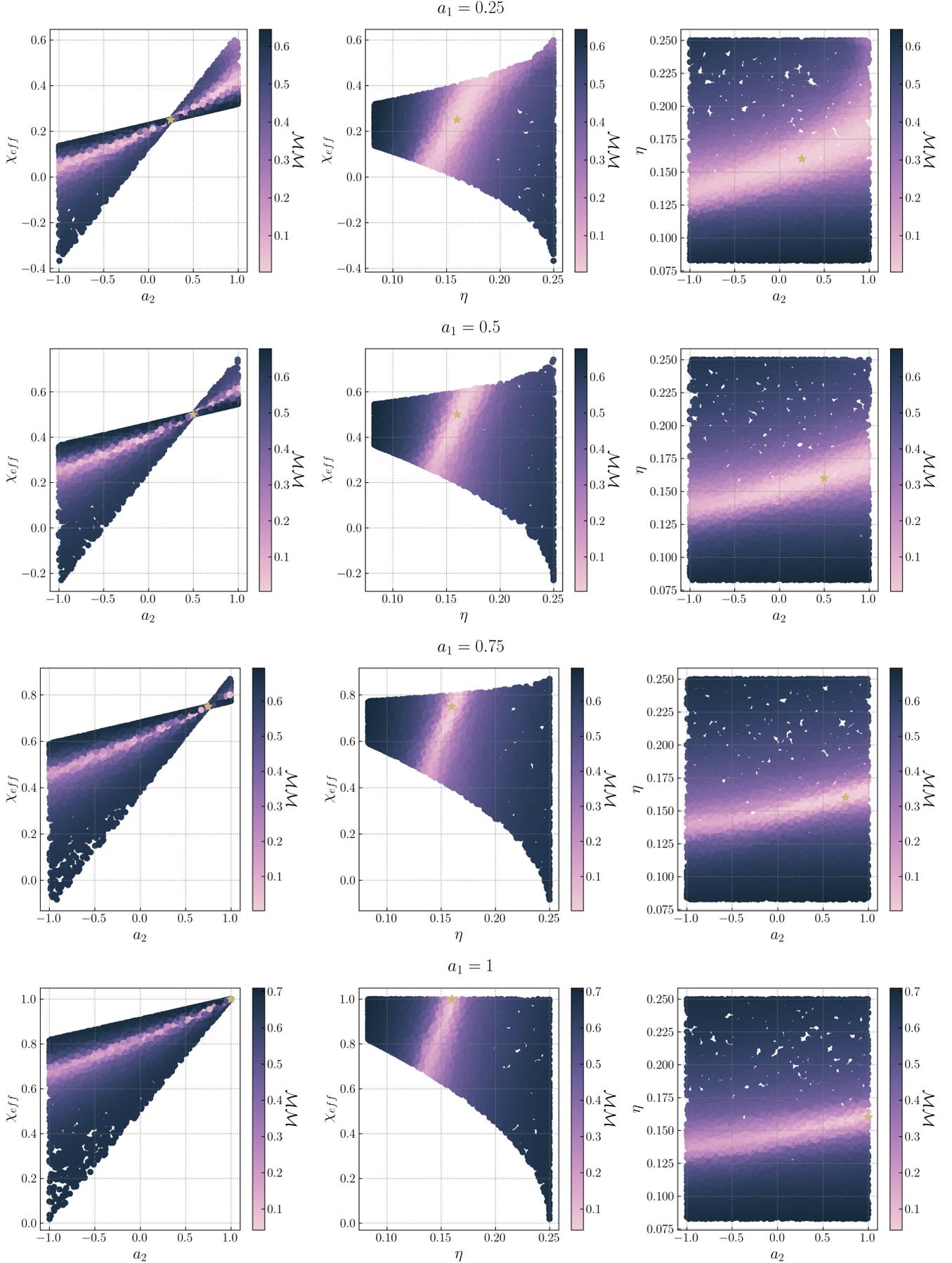


Figure 13: Parameter space with fixed positive primary spin component

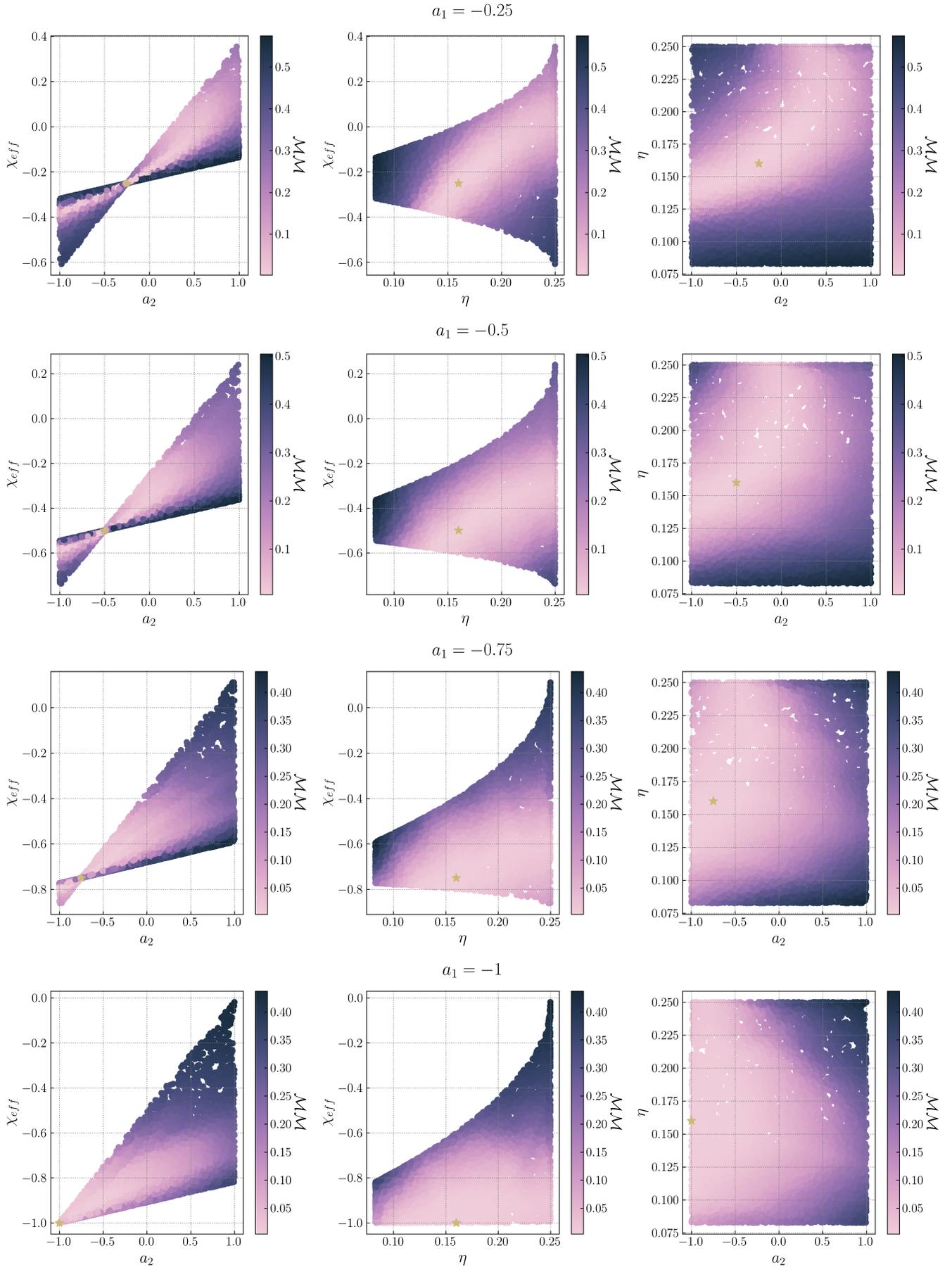


Figure 14: Parameter space with fixed negative primary spin component

## C. IDENTIFYING CORRELATIONS

### C.1. Kernel Density Estimate (KDE)

Initially, we attempted to identify the degeneracies in the parameter space by computing the eigenvalues of Hessians at each point. However, since the parameter space consists of discrete points, there is too much noise about the associated Hessian from interpolation. We then instead applied kernel density estimation (KDE) to characterize discretized parameter space by using `PESummary`. We set `weights = MM` and `method = "Reflection"` to account for the boundary conditions of the parameter space. As the KDE function only provides a visualization of the correlations, we plan to fit the injected simulations to a Gaussian mixture model next, which would enable us to quantify the correlations between parameters.

### C.2. Gaussian Mixture Model (GMM)

In this section, we detail the method utilized for local estimation of a single Gaussian distribution from multidimensional data through the `sklearn.mixture` implementation. The Gaussian Mixture Model (GMM) is a probabilistic model representing a mixture of multiple Gaussian distributions. Through the `sklearn.mixture` framework, we apply the GMM to fit data points in our BBH parameter space. The foundational assumption of this model is that the observed data is generated from multiple Gaussian distributions. Each component of the GMM, or cluster, is defined by its mean ( $\boldsymbol{\mu}_k$ ) and covariance matrix ( $\boldsymbol{\Sigma}_k$ ):

$$p(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = (2\pi)^{-\frac{D}{2}} |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right)$$

We used the Bayesian approach `sklearn.mixture.BayesianGaussianMixture` integrated in the module to counteract overfitting issues, especially when a GMM with a single component is used. This is represented using Bayes' theorem:

$$p(\theta \mid d) = \frac{L(d \mid \theta)\pi(\theta)}{Z(d)}$$

Gaussian Mixture Models uses an Expectation - Maximization (EM) Algorithm for parameter estimation. The likelihood function for GMM, which offers insight into the probability of observed data given our model parameters, is expressed as:

$$p(\mathbf{X} \mid \boldsymbol{\theta}) = \prod_{n=1}^N \sum_{k=1}^K \pi_k \frac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_k}} \exp\left(-\frac{(\mathbf{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k)}{2}\right),$$

$$\mathcal{L} = \log p(\mathbf{X} \mid \boldsymbol{\theta}) = \sum_{n=1}^N \log \left[ \sum_{k=1}^K \pi_k \frac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_k}} \exp\left(-\frac{(\mathbf{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k)}{2}\right) \right].$$

Here,  $\boldsymbol{\theta}$  encapsulates the full set of model parameters, and the function  $\mathcal{L}$  denotes the log-likelihood, crucial for the iterative optimization within the EM algorithm.

### C.3. Principal Component Analysis (PCA)

Principal Component Analysis is an unsupervised learning method for reducing dimensionality of data. Since we are only fitting the data to one Gaussian component, the variances calculated for the two methods are comparable. The fundamental goal of PCA is to rotate and project the original data onto a new coordinate system such that the largest variance by any projection of the data is placed on the first axis, labeled the first principal component. The subsequent axis captures the next largest variance, and this continues in decreasing order for the other dimensions.

Mathematically, the principal components of PCA can be derived from the eigenvectors  $\mathbf{e}_i$  of the data's covariance matrix  $\boldsymbol{\Sigma}$  :

$$\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i\mathbf{e}_i$$

where  $\lambda_i$  is the corresponding eigenvalue, indicating the variance captured by the  $i^{\text{th}}$  principal component.

## D. FRAMES FROM WALKER

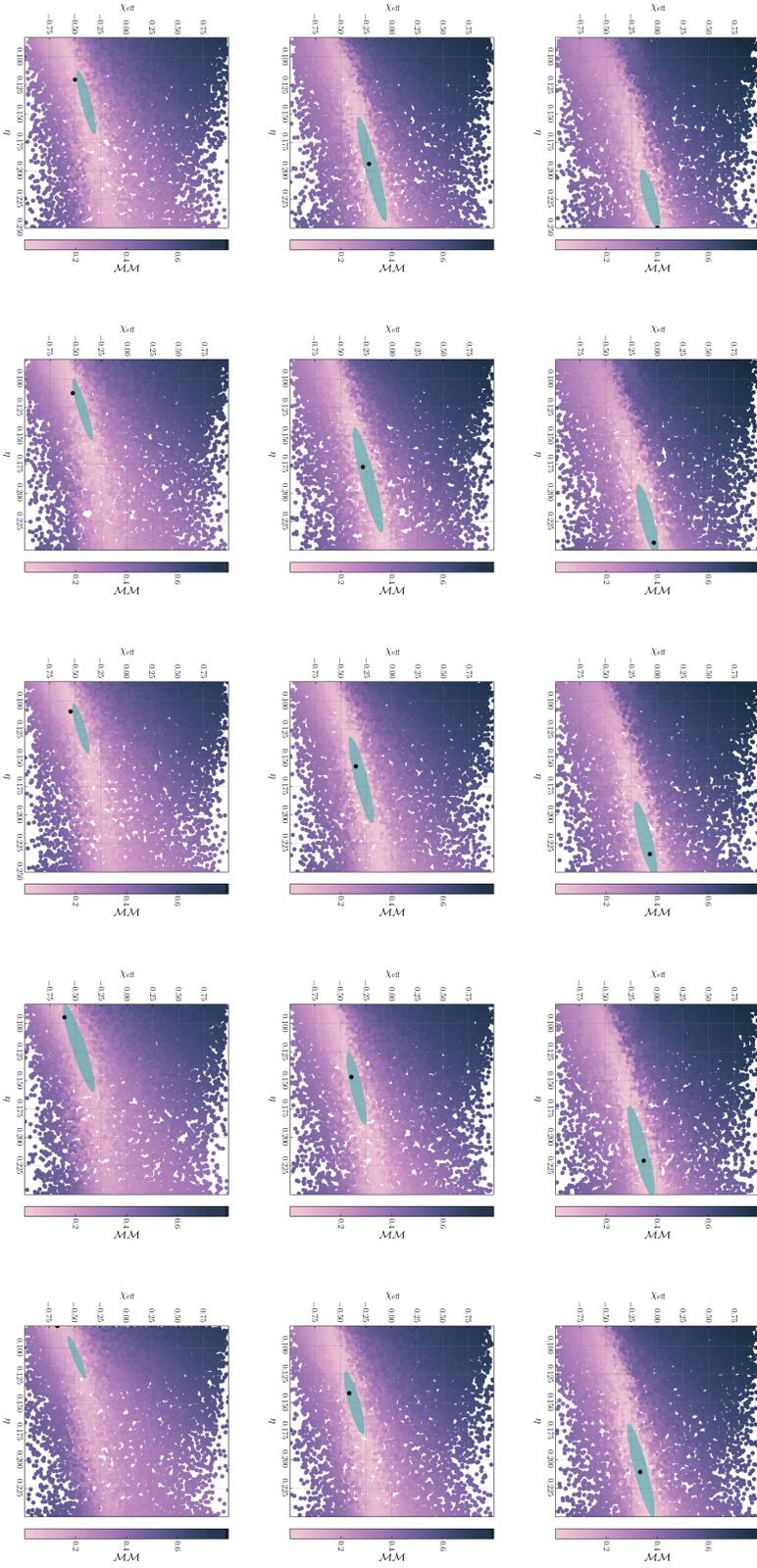


Figure 15: 2D walker path and parameter space evolution for starting reference simulation at  $\lambda = [0.25, 0, 0, 0, 0, 0]$

## REFERENCES

- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020, The Astrophysical Journal, 900, L13, doi: [10.3847/2041-8213/aba493](https://doi.org/10.3847/2041-8213/aba493)
- Biscoveanu, S., Isi, M., Varma, V., & Vitale, S. 2021, Physical Review D, 104, doi: [10.1103/physrevd.104.103018](https://doi.org/10.1103/physrevd.104.103018)
- Ferguson, D. 2023, Physical Review D, 107, doi: [10.1103/physrevd.107.024034](https://doi.org/10.1103/physrevd.107.024034)
- Mandel, I., & Farmer, A. 2022, Physics Reports, 955, 1, doi: [10.1016/j.physrep.2022.01.003](https://doi.org/10.1016/j.physrep.2022.01.003)
- Ng, K. K. Y., Vitale, S., Zimmerman, A., et al. 2018, Physical Review D, 98, 083007, doi: [10.1103/PhysRevD.98.083007](https://doi.org/10.1103/PhysRevD.98.083007)
- Schmidt, P., Ohme, F., & Hannam, M. 2015, Phys. Rev. D, 91, 024043, doi: [10.1103/PhysRevD.91.024043](https://doi.org/10.1103/PhysRevD.91.024043)
- Varma, V., Field, S. E., Scheel, M. A., et al. 2019, Phys. Rev. Res., 1, 033015, doi: [10.1103/PhysRevResearch.1.033015](https://doi.org/10.1103/PhysRevResearch.1.033015)