Parameter Estimation consistency between MCMC and Fisher Information Matrix

Peter Shaffery

University of Florida

Seoul National University and

University of Massachusetts Lowell

Abstract

We test the agreement between the standard deviation predicted by effective Fisher Matrix and by MCMC calculated, 1-D posterior for the symmetric mass ratio and chirp mass parameters of a non-spinning, inspiral, BH-NS binary. We had planned to obtain MCMC data for 6 different total masses, but computational difficulties prevented any runs up to the time of writing. Instead we obtained MCMC data from another paper for an identical injection to our 11.4 total mass binary. We find that, in this case, the effective Fisher Matrix and the MCMC results are in close agreement.

I. INTRODUCTION

As the next generation of detectors come online, the work of gravitational wave physics will begin to shift from just making a detection to actually extracting information about the generating source of some detected signal. Thus, while detector sensitivity is still a crucial area of development, it is equally important to be able to assess and improve the accuracy, precision, and efficiency of a detector and it's data extraction pipeline.

Fisher Information Matrices provide a useful statistical tool for meeting this goal and are a natural choice given that they naturally extend from the matched filtering techniques used to extract a signal from the noisy detector output. Matched filtering pairs a mathematically constructed template waveform from known parameter values with the detector output and takes their inner product (referred to as an overlap function), normalized by the detector's noise curve. Various models exist for the purpose of template construction. Each model uses different parameter sets, or extends a previous model to a higher post-Newtonian order. The parameters used in each model describe the physical quantities of the generating source (binary mass, orbital phase, distance to source); a typical model uses between 9 and 15 parameters. Abstractly, the inner product assigns a scalar to a vector pair corresponding to the "distance" between them, and this provides an intuitive understanding of matched filtering: it quantitatively assesses the dissimilarity between the template and the signal. The resulting scalar output is referred to as the Signal to Noise Ratio (SNR).

If a template-signal pair generates a sufficiently high SNR then the template and the corresponding parameter values used to generate it are flagged as a measurement and is passed on for parameter estimation. Parameter estimation makes use of the overlap function as part of the "posterior", a multidimensional probability density function which assigns a probability to each set of parameter values. The probability distribution generated by this function serves as a physical representation both of the measurement and of its precision. The maxima of the distribution corresponds to the most probable parameter values of the generating source, while the standard deviation of the distribution represents the precision of the measurement. That is to say, a wider curve implies that a large range of values are probably the "true" value of the GW source, while a narrower curve implies that a small range of such values of probable. Generally it is assumed that this distribution is multivariate Gaussian, however depending on correlations between parameters and the specific noise realization this is not always exactly true. It is not uncommon for the posterior distribution to have several local maxima and minima, and may not even have a single global maximum, this obvious undesirable.

Calculating the probability density curve over a representation patch of parameter space presents some difficulties, however. It requires a random walk through a high-dimensional space, and a full calculation of the posterior at each point. In the case of a non-spinning BH-NS binary, 9 parameters are required to define the system, and when spinning cases are allowed as many as 15 parameters are needed; thus, calculating a posterior can be very computationally expensive, particularly if the posterior has many local maxima and minima.

As the parameter estimation pipeline is developed, several template generating models need to be compared. A model is examined by first generating a simulated waveform from known parameter values, and then injecting it into real detector noise. This is then run through the detector pipeline and extracted using templates generated by the model under consideration. The posterior generated by these templates is the best metric to determine the quality of the model. Ideally the global maximum of the posterior will correspond closely to the parameter values of the simulated waveform, and furthermore the standard deviation will be reasonably low. For our current generation of models, the deflection of the global maximum from the true parameter values is depends primarily on the noise from the detector and the SNR of the measurement rather than the model chosen, however the standard deviation of the distribution, ie. the precision of the measurement, does depend strongly on the model used. Thus there is a need for a fast and effective method of comparing the precision two different models would allow. This the role of the Fisher Information Matrix; the standard definition of the FIM is motivated by the expression for a multivariate Gaussian probability function, but more practically it can be described as the inverse of the covariance matrix of some distribution. The FIM can be calculated analytically, requiring only the template generating model. In doing so, the precision of the parameter estimation pipeline can quickly be found without actually having make a measurement of an injected, simulated signal: providing a fast, simple method of comparing waveform models.

The Fisher Matrix has limitations, however. It assumes a model with linearly correlated parameters, a detector with Gaussian noise, and a high SNR. Rodriguez et al. have shown that at total binary mass higher than 10.0 solar mass, the standard deviation predicted by Fisher Matrices does not agree with the standard deviation of a fully calculated posterior

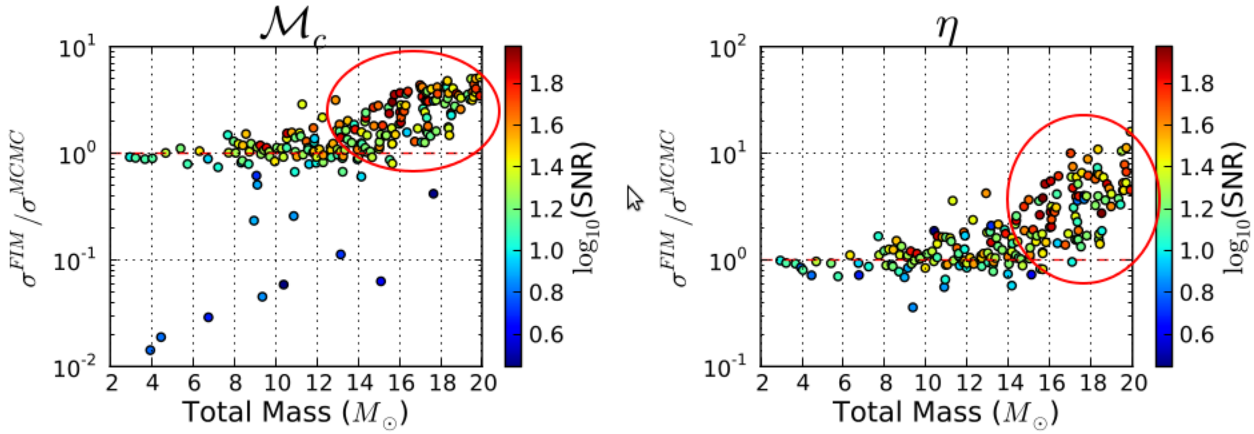


FIG. 1. Results from Rodriguez et al. showing disagreement between the standard Fisher Matrix and MCMC results at high total mass (red circles).

(Fig. 1). Motivated by this, Cho et al. have defined an alternative, effective Fisher Matrix, calculated from a quadratic fit to the posterior at some scale of interest.

Cho et al. showed that the effective Fisher Matrix and MCMC are consistent for a non-spinning BH-NS binary inspiral, however they used only a single injection for a 4.5 solar mass binary . In this work, we vary the BH mass and distance to the binary to study the consistency between the effective Fisher Matrix and MCMC results for non-spinning BH-NS inspiral signals with regard to predicting the precision of measurement of the chirp mass and symmetric mass ratio parameters.

II. BACKGROUND

A. Overlap and Ambiguity Function

Matched filtering relies on an inner product referred to as an overlap function. The overlap function accepts two input signals and returns a scalar proportionate to how similar the inputs are (ie. it is an inner product weighted by the LIGO noise curve) and is defined:

$$\langle a(t)|b(t)\rangle \equiv \int \frac{\tilde{a}(f)\tilde{b}(f)^*}{S_n} \tag{1}$$

Where $a(t)$ and $b(t)$ are two time domain waveforms and $\tilde{a}(f)$ and $\tilde{b}(f)$ are their Fourier transforms.

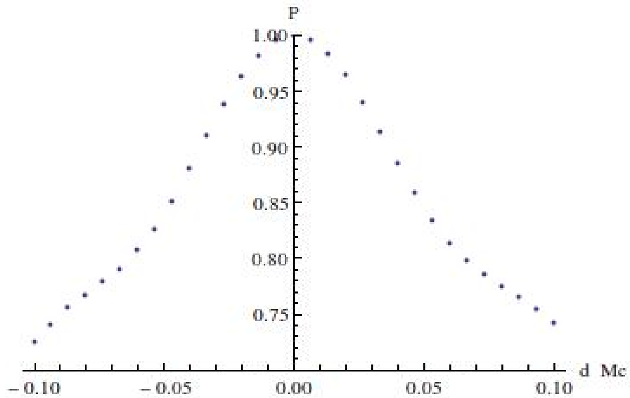


FIG. 2. The overlap (P) between a template waveform and a waveform with varying chirp mass (M_c). Note how the structure of the function changes at different levels of SNR

In some cases the overlap function can have scale-dependent structure (see Fig. 2). As standard Fisher Matrices rely on differentiating a waveform by each of its parameters they respond strongly to fine scale structure. Therefore the presence of this structure will have ramifications for the effectiveness of a standard Fisher Matrix as we will discuss shortly.

Furthermore from Cho et al. we define a normalized ambiguity function between two signals $a(\lambda_1)$ and $b(\lambda_2)$ as:

$$P(a(\lambda_1), b(\lambda_2)) \equiv max_{t_c,\psi} \frac{|\langle a(\lambda_1)|b(\lambda_2)\rangle|}{\sqrt{\langle a(\lambda_1)|a(\lambda_1)\rangle\langle b(\lambda_2)|b(\lambda_2)\rangle}} \tag{2}$$

With this we can state that two different signals are distinguishable if the ambiguity function between them satisfies:

$$1 - P \geq 1/\rho^2 \tag{3}$$

This will have relevance later in our discussion of the effective Fisher Matrix.

B. MCMC and Calculating Posteriors

Calculation of posteriors relies upon Bayes Theorem which states:

$$p(\hat{\lambda}|s) \propto p(\hat{\lambda})L(s|\hat{\lambda}) \tag{4}$$

where $p(\hat{\lambda}|s)$ is the probability (posterior) that the parameter vector $\hat{\lambda}$ contains the actual values of the parameters of the generating source given some measured signal s, $p(\hat{\lambda})$ is the "prior" which reflects our previous knowledge of the probability of certain parameter values (in this context it is a constant over all physically possible values, representing total ignorance), and where $L(s|\hat{\lambda})$ is the "likelihood function" which expresses how likely we are to have measured such signal given some parameter vector.

The detector output S can be represented as the sum of a noise realization N_0 and a "true" signal $h(\lambda)$, thus:

$$S = N_0 + h(\lambda) \tag{5}$$

If we assume that the noise is a stationary, Gaussian process then the probability of some noise realization is:

$$p(N_0) \propto e^{\frac{-\langle N_0|N_0\rangle}{2}} \tag{6}$$

Combined with the expression for detector output we have an expression for the likelihood function:

$$L(S|\hat{\lambda}) = e^{\frac{-\langle S-h(\hat{\lambda})|S-h(\hat{\lambda})\rangle}{2}} \tag{7}$$

Since the log likelihood has the same maximum as the likelihood itself it can also serve in Bayes' theorem. With this then, given some normative constant C, we can write our posterior function:

$$p(\hat{\lambda}|S) = C * \frac{-\langle S-h(\hat{\lambda})|S-h(\hat{\lambda})\rangle}{2} \tag{8}$$

This expression of the posterior probability of some parameter vector is used by MCMC algorithms to calculate the posterior distribution in the parameter space of the model. In our usage MCMC refers specifically to the Metropolis-Hastings algorithm. In this algorithm a Markov Chain is generated, where each step in the chain is probabilistically independent from the previous step. The chain begins at some random point in the parameter space and from here a candidate point is selected from a small neighbourhood. The posterior at the candidate point is calculated, and the candidate point is then accepted with some probability, which is weighted to favor points with a higher posterior probability. In this way the random walk will cover a representative area of the parameter space while never straying far from the global maximum. A problem arises when the posterior has several local maxima. These can attract the Markov Chain and cause it to converge prematurely, ie. before sufficiently exploring the global maximum. To prevent this the MCMC run is passed between several computing threads. In each thread the posterior is 'flattened' by some temperature factor T (eg. thread 0 is the unchanged posterior, thread 1 is slightly more flattened, thread 2 even more, etc.) As the run is passed between them the Markov

Chain will still be attracted to the maxima of the posterior, however the probability of it converging on some local maxima is reduced. To further ensure that a good portion of the parameter space is covered a single injection is typically run between three and five times, and the results are combined.

Cho et al. also provides this alternate expression for log likelihood:

$$\ln L = -\rho^2(1 - P) \tag{9}$$

Where ρ is the SNR and P is again the ambiguity function. For a derivation of this expression see Cho et al.

C. Standard Fisher Matrices

Assuming that $\hat{\lambda} \approx \hat{\lambda}_0$ we may develop a Taylor approximant to first order of $h(\hat{\lambda})$ about $\hat{\lambda}_0$:

$$h(\hat{\lambda}) \approx h(\hat{\lambda}_0) + \delta\lambda^i h_i \tag{10}$$

Where λ^i represents the i$^{\text{th}}$ parameter of $\hat{\lambda}$, $\delta\lambda^i = |\lambda^i - \lambda_0^i|$, and h_i is the derivative of h with respect to λ^i and evaluated at λ_0^i. Note that we are employing the summation convention, where repeated sub and superscripts are summed over. This is called the Linear Signal Approximation (LSA), and if we assume that $h(\hat{\lambda})$ is close to $h(\hat{\lambda}_0)$ then we can recast (3):

$$S = N_0 + h(\hat{\lambda}_0) \tag{11}$$

From here it is simple to combine (6) and (9):

$$\begin{aligned} p(\hat{\lambda}|S) = \exp[\frac{-1}{2}\langle N_0 + h(\hat{\lambda}_0) - h(\hat{\lambda})| \\ N_0 + h(\hat{\lambda}_0) - h(\hat{\lambda})\rangle \end{aligned} \tag{12}$$

It is then obvious that by using the LSA (8) and by exploiting the linearity of the inner product we can arrive at our final formulation of (6):

$$p(\hat{\lambda}|S) = N * \exp[\frac{1}{2}\delta\lambda^i\delta\lambda^j\langle h_i|h_j\rangle] \tag{13}$$

As noted above, this is all derived from a Gaussian distribution, and thus (11) represents a multivariate, Gaussian distribution which gives a probability density function for some measurement error around parameters λ_i and λ_j. Therefore we define each element of our Fisher Matrix:

$$\Gamma_{ij} \equiv \langle h_i | h_j \rangle \tag{14}$$

So that the covariance matrix Σ_{ij} can be found as the inverse of the Fisher Matrix:

$$\Sigma_{ij} = {\Gamma_{ij}}^{-1} \tag{15}$$

D. Effective Fisher Matrices

While the Fisher Matrix is an extremely useful tool in the case of purely Gaussian data, ie. when the SNR is very high and the LSA holds completely true. However in many applications these assumptions do not hold so well, and the Fisher Matrix will frequently fall prey to the scale dependence discussed above. As such, the σ_{ij} predicted by the FM are typically overly optimistic when compared to a full computation of the posterior. It is therefore more useful to define an effective Fisher Matrix (effFM). Using the definition of the log likelihood given by (7) and an alternate expression of the Fisher Matrix relating it to the log likelihood. We can also define a normalized Fisher Matrix, $\hat{\Gamma}_{ij}$ (Cho et al. 26, 39):

$$\begin{aligned}\Gamma_{ij} &= \frac{-\partial^2 \ln L(\hat{\lambda})}{\partial \lambda_i \partial \lambda_j} \\ &= \rho^2 \frac{\partial^2 (1-P)}{\partial \lambda_i \partial \lambda_j} = \rho^2 \hat{\Gamma}_{ij}\end{aligned} \tag{16}$$

With this we can motivate a step-by-step method of calculating the effFM.

First we must plot the overlap P between some fiducial waveform and a template varied over some scale of interest (typically $P \geq .99$) . From here we can use a least squares fitting to create a quadratic fitting P^* (a quadratic curve is a decent approximation of Gaussian curve, particularly at high values along the horizontal axis):

$$P^* = P_{\max} + p_1 \delta\lambda_1^2 + p_2 \delta\lambda_2^2 + p_{12} \delta\lambda_1 \delta\lambda_2 \tag{17}$$

Where p_1, p_2, and p_{12} are fitting constants and we have chosen to set $P_{\max} = 1$. If we then substitute P^* for P in (14) we have our definition off ${}_{\text{eff}}\hat{\Gamma}_{ij}$, furthermore we can find each element of the effective Fisher Matrix by way of the fitting constants. For example, ${}_{\text{eff}}\hat{\Gamma}_{11} = -2p_1$, this can be shown:

$$
\begin{aligned}
{}_{\text{eff}}\hat{\Gamma}_{ij} &= \frac{\partial^2(1-P^*)}{\partial\lambda_i\partial\lambda_j} \\
&= \frac{\partial^2(-p_1\delta\lambda_1^2 - p_2\delta\lambda_2^2 - p_{12}\delta\lambda_1\delta\lambda_2)}{\partial\lambda_i\partial\lambda_j} \\
{}_{\text{eff}}\hat{\Gamma}_{11} &= \frac{\partial^2(-p_1\delta\lambda_1^2 - p_2\delta\lambda_2^2 - p_{12}\delta\lambda_1\delta\lambda_2)}{\partial^2\lambda_1} \\
&= -2p_1
\end{aligned}
\tag{18}
$$

Thus by using a least squares regression to find suitable fitting constants the effective Fisher Matrix can be easily computed. To find the fitting scale P we can use:

$$
1 - P \leq \frac{1}{\rho^2} \tag{19}
$$

By basing our fitting scale on the SNR, ρ, we are calculating the precision of the measurement where the signal is strongest.

III. RESULTS

To provide a simple metric for the consistency between the standard deviation predicted by the FM and by MCMC parameter estimation we define a standard deviation ratio Rodriguez et al.:

$$
\Lambda \equiv \frac{\sigma_{FM}}{\sigma_{MCMC}} \tag{20}
$$

Where σ_{FM} and σ_{MCMC} are the standard deviations predicted by FM and MCMC estimation respectively. Note that $\Lambda \approx 1$ indicates the desired agreement between FM and MCMC.

A. Injection Parameters

To test the effective Fisher Matrix we used the SpinTaylorT4 waveform model for a non-spinning, BH-NS binary. Furthermore, Cho et al. have shown that the effective Fisher Matrix is particularly suitable for the chirps mass and symmetric mass ratio (η) parameters. Therefore these are the two parameters we generated effective Fisher Matrices for. They are

m_1	M_c	η	*Distance*
2.8	1.7034	0.222222	15.0
5.0	2.173	0.170898	18.0
6.0	2.4028	0.153397	20.0
10.0	2.9943	0.107725	23.1
20.0	4.0014	0.06114	26.0
30.0	4.7268	0.043598	26.3

FIG. 3. The injected values for each BH mass and the SNR corrected distances.

each given:

$$\eta = \frac{m_1 \times m_2}{(m_1 + m_2)^2}$$
$$M_c = (m_1 + m_2) \times \eta^{3/5} \tag{21}$$

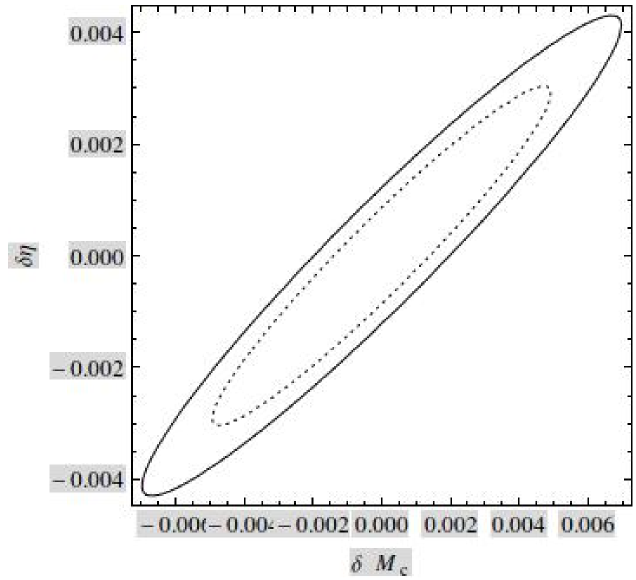
Where m_1 and m_2 are the individual masses of the binary components.

Cho et al. calculated an effective Fisher Matrix for these two parameters only for an 11.4 solar mass binary. Motivated by Rodriguez et al. we fixed $m_2 = 1.4$ (m_2 being the neutron star mass) and varied the black hole mass (m_1) from 2.8 to 30.0 solar mass. For each value of m_1 we calculated M_c and η (see Fig. 3).

Furthermore, MCMC calculation requires that we also specify the distance from the detector to the binary. Because heavier total mass binaries give a higher SNR at the same distance as a lighter total mass, we adjusted the the distances such that each injection would have SNR$\approx$ 20. These values were found, essentially, by guess and check.

B. Effective Fisher Matrices

Using a python script a single simulation waveform was generated for each injection. A template waveform was then generated for varied values of M_c and η, and an overlap between the two was taken. The resulting data was then plotted with Mathematica and a quadratic fitting was found and used to generate an effective Fisher Matrix. Because of the relatively few number of data points $\geq .995$ it was difficult to create a fitting to this scale, and thus difficult to create an effective Fisher Matrix at an SNR$=$ 20. Therefore each effective Fisher Matrix was calculated at an SNR$=$ 10 ($P \geq .99$) (Fig. 4) and then the resulting standard



—	m_1	η
M_c	5082.4	-7904.1
η	-7904.1	13373.0

FIG. 4. effective Fisher Matrix and confidence ellipse (contour plot of posterior at P = .99 [solid line] and .995 [dotted line]) for reference injection (BH Mass = 10.0 solar mass and NS mass =1.4 solar mass).

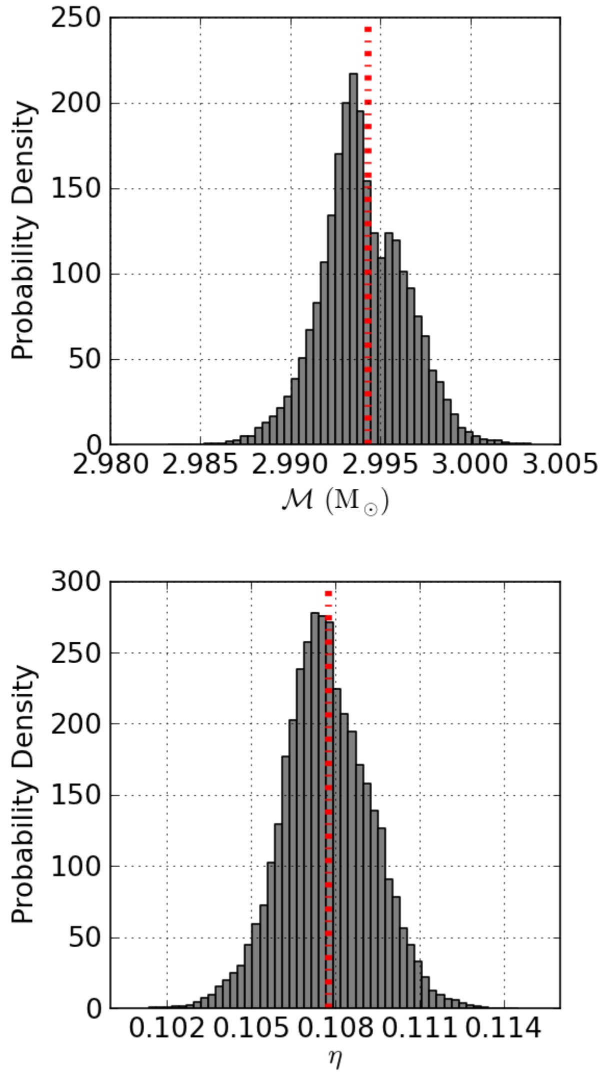
M_c	η
0.00247	0.00152

FIG. 5. Standard Deviation calculated by effective Fisher Matrix and scaled by .5

deviations were scaled by .5, thus giving us the standard deviation of the posterior for an SNR= 20 (Fig. 5).

C. MCMC Results

Difficulties with the KISTI computing cluster prevented MCMC runs for all 6 injection parameters up to the time of writing. Another paper, O'Shaughnessy et al. (in prep.), however, performed an MCMC run with injected parameters identical to our reference injection (11.4 total mass). With this data we were able to perform the analysis we wanted for at least one case.



M_c	η
0.00235	0.00156

FIG. 6. Marginalized, 1-dimensional PDFs for the chirp mass and symmetric mass ratio (the red line indicates the injected value of the parameter) and their standard deviations.

Using the cbcBayesPostProc.py script from the Ligo Algorithm Library we marginalized the full, 9-dimensional posterior, down to a 1-dimensional PDF for each parameter from which we found a standard deviation for the chirp mass and symmetric mass ratio (Fig. 6).

Using these values for standard deviation we computed the desired ratios for each parameter (Fig. 7).

	M_c	η
σ_{effFM}	0.00247	0.00152
σ_{MCMC}	0.00235	0.00156
$\frac{\sigma_{effFM}}{\sigma_{MCMC}}$	1.0508	0.969

FIG. 7. Standard Deviation calculated from effective Fisher Matrix and MCMC results and their ratio.

IV. CONCLUSION

The $\frac{\sigma_{effFM}}{\sigma_{MCMC}} \approx 1$ for each parameter indicates close agreement between the effFM and MCMC results at SNR$\approx$ 20. This result is is similar to that of Rodriguez et al. (Fig. 1) at 11.4 total mass. In the absence of more data, however, particularly at higher total mass values, this result can only confirm that the effective Fisher Matrix functions at least as well as the standard Fisher Matrix with regards to agreement with MCMC.

At time of writing the KISTI computing cluster is undergoing repairs and upgrades. When it is fully operation we plan to submit our full set of 12 jobs (6 parameter injections both amplitude corrected to 3.5 pN order and uncorrected) for at least 3 runs each.

V. REFERENCES

[1] Cho, H-S., Ochsner, E., O'Shaughnessy, R., Kim, C., Lee, C-H., 2013, arXiv:1209.4494

[2] Rodriguez, C., Farr, B., Farr, W., Mandel, I., in prep., "Inadequacies of the Fisher Information Matrix in gravitational-wave parameter estimation"

[3] O'Shaughnessy et al., in prep.

[4] Ochsner, E., 2010, "Improving Analytical Templates and Searching for Gravitational Waves from Coalescing Black Hole Binaries"

VI. ACKNOWLEDGEMENTS

I would like to thank Chunglee Kim for her tireless assistance in this project, she has been a fantastic mentor both as relates directly to the project as well as explaining to me ”the rules of the game” so to speak. Her reviews of my presentation and report have been tireless, ruthless, and incredibly helpful. Furthermore I would like to thank Prof. Hyung Mok Lee and the his research group for making me feel like a welcome part of the Astronomy department.

In addition I want to extend my thanks to Dr. Whiting and Dr. Mueller of the University of Florida for giving me this opportunity and for organizing this program. Kristin Nichola and Antonis Mytidis also have gone above and beyond, and I want to thank them as well.