Detailed comparison of LIGO and Virgo inspiral pipelines in preparation for a joint search

F Beauville[1], M-A Bizouard[2], L Blackburn[3], L Bosi[4], L Brocco[5], D A Brown[6,7], D Buskulic[1], F Cavalier[2], S Chatterji[6], N Christensen[8,9], A-C Clapson[2], S Fairhurst[7], D Grosjean[1], G Guidi[10], P Hello[2], S Heng[11], M Hewitson[11], E Katsavounidis[3], S Klimenko[12], M Knight[8], A Lazzarini[6], N Leroy[2], F Marion[1], J Markowitz[3], C Melachrinos[3], B Mours[1], F Ricci[5], A Viceré[10], I Yakushin[13] and M Zanolin[3] (The joint LIGO/Virgo working group)

[1] Laboratoire d'Annecy-le-Vieux de Physique des Particules, Chemin de Bellevue, BP 110, 74941 Annecy-le-Vieux Cedex, France
[2] Laboratoire de l'Accélérateur Linéaire, IN2P3/CNRS-Université de Paris XI, BP 34, 91898 Orsay Cedex, France
[3] LIGO—Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4] INFN Sezione di Perugia and/or Università di Perugia, Via A Pascoli, I-06123 Perugia, Italy
[5] INFN Sezione di Roma and/or Università "La Sapienza", P.le A Moro 2, I-00185, Roma, Italy
[6] LIGO—California Institute of Technology, Pasadena, CA 91125, USA
[7] University of Wisconsin—Milwaukee, Milwaukee, WI 53201, USA
[8] Carleton College, Northfield, MN 55057, USA
[9] European Gravitational Observatory (EGO), Cascina (Pi), Italy
[10] INFN Sezione Firenze/Urbino Via G Sansone 1, I-50019 Sesto Fiorentino; and/or Università di Firenze, Largo E.Fermi 2, I-50125 Firenze and/or Università di Urbino, Via S Chiara 27, I-61029 Urbino, Italy
[11] University of Glasgow, Glasgow, G12 8QQ, UK
[12] University of Florida—Gainesville, FL 32611, USA
[13] LIGO Livingston Observatory, Livingston, LA 70754, USA

Abstract

Presented in this paper is a detailed and direct comparison of the detection pipelines used by LIGO and Virgo in their attempt to observe gravitational waves from binary neutron star systems. In order to test the search programs, numerous inspiral signals were added to 24 h of simulated detector data. The efficiencies of the different pipelines were tested, and found to be very similar. Parameter estimation routines were also tested. We demonstrate that there are definite benefits if LIGO and Virgo conduct a joint coincident analysis; these advantages include increased detection efficiency and information on source sky location.

PACS numbers: 95.55.Ym, 04.80.Nn, 07.05.Kf

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The Laser Interferometer Gravitational Wave Observatory (LIGO) [1] detectors have reached their design sensitivity, while Virgo [2] is quickly approaching its target sensitivity. This achievement will ultimately be rewarded through the observation of gravitational waves. Numerous potential sources exist, producing signals of differing character. The goal of the data analysis efforts of LIGO and Virgo are to detect and identify these signals. The inspiral of binary compact objects, such as neutron stars or black holes, is one of the most promising sources of gravitational waves. The observation of the coalescence of these binary systems will expand our knowledge of the astrophysics of compact objects and provide unique tests of general relativity and cosmology [3]. LIGO has already conducted searches for binary neutron star inspiral signals, and has placed upper limits on source distributions [4, 5].

LIGO and Virgo have each developed methods (analysis pipeline software) for finding binary inspiral signals. In order to maximize the probability of observing gravitational waves LIGO and Virgo have embarked on a collaborative detection effort, and a necessary beginning to such a working relationship is the validation and understanding of each group's detection strategies. In this paper, we present the results of a comprehensive study where the LIGO and Virgo inspiral search pipelines were compared side by side. We have also conducted a similar LIGO–Virgo comparison study for a gravitational wave burst search [6]. The results presented in this paper show that the LIGO and Virgo binary inspiral detection pipelines operate equally well. In addition, we show that by working together there are undeniable benefits in the quest for detecting gravitational waves from binary inspirals, and that more astrophysical information can be extracted from the signals. For example, the results summarized in this paper demonstrate that a detection strategy based on a two-detector coincidence is improved considerably with Virgo included, as opposed to just using LIGO data from the Livingston and Hanford observatories.

In order to conduct a study that compares inspiral detection capabilities it was necessary for both groups to ensure that they were looking for identical signals. The LIGO and Virgo groups demonstrated to one another that they were, in fact, looking for the exact same form of the binary inspiral signals. This was a non-trivial commencement to the study, as the mathematical description of the inspiral signal template is complicated and only approximate. Only after this signal generation validation study was complete we were able to mutually verify the detection capability of our search pipelines.

A previous LIGO–Virgo study showed that both detection pipelines were equally successful in finding inspiral signals from optimally oriented sources [7]. In the study that we are presenting here the detection validation was accomplished using simulated data containing realistic source orientation. Specifically, the binary inspiral sources were simulated to come from galaxies M87 and NGC 6744. Due to the effects of the earth's rotation, and the random orientation of the binary orbital plane, the (simulated) signals impinging on the (simulated) interferometers were non-optimum for detection. The present study also incorporates the time delays that will be present in the response of a network of detectors to gravitational waves from some particular sky direction. In this way we conducted a study whereby the response of the network of the LIGO and Virgo detectors was as realistic as possible. For this exercise we created simulated noise, with the noise spectral density matching the target expectations for LIGO and Virgo; the simulated noise was Gaussian distributed.

The paper is organized as follows. In section 2, we review the results of our previous analysis where the LIGO and Virgo inspiral pipelines were used to detect (fictitious) gravitational wave signals from optimally oriented binary neutron star systems. In section 3, we discuss the comparison of the generation of simulated binary inspiral signals, and ensure

that LIGO and Virgo are searching for signals of the exact same form. In section 4, all of the LIGO and Virgo inspiral search pipelines are discussed, and their response to the data from this study are presented; the two Virgo pipelines are presented in sections 4.1 and 4.2, while the LIGO pipeline is described in section 4.3. The results from a LIGO developed Bayesian parameter estimation technique are given in section 5. A side-by-side comparison of the Virgo and LIGO pipeline results is summarized in section 6. The benefits of a combined LIGO and Virgo inspiral search are presented in section 7; we find that there is an increase in two-way coincident detection probability, and that there is also the means to gather information on the sky location of the source. Concluding remarks, and an outline for future goals are presented in section 8.

2. Review of initial comparison project

LIGO and Virgo recently initiated a study to compare their methods for detecting gravitational waves from binary inspiral systems. Each group tested their binary inspiral pipelines on simulated data. A similar study was also conducted on the LIGO and Virgo burst search pipelines [8]. For the inspiral study signals were created from optimally oriented sources, directly above the interferometers; the detector noise levels were comparable to the target sensitivities for LIGO and Virgo. Each collaboration created signals in order to confirm that the other collaboration's pipeline was able to detect them.

The simulated LIGO signals were created using two mass pairs, [$1.4M_\odot$, $1.4M_\odot$] and [$1.0M_\odot$, $1.0M_\odot$], at various distances (20, 25, 30 and 35 Mpc), and then inserted into the noise. The LIGO $h(t)$ strain signal was created with a sampling rate of 16 384 Hz, and a lower frequency cutoff of 40 Hz. A total of 26 signals were spread over 3 h of data. The synthesized Virgo signals were from a [$1.4M_\odot$, $1.4M_\odot$] mass pair at a distance of 25 Mpc. The $h(t)$ strain signal had a sampling rate of 20 000 Hz, and a low frequency cutoff of 24 Hz; nine signals were injected into 2.5 h of data.

The results from the initial study confirmed the ability of each group to correctly detect and characterize a binary inspiral signal. The LIGO and Virgo groups both analyzed the data created by each group. The LIGO and Virgo pipelines were able to detect the same events, and produced comparable parameter estimates for the chirp mass ($m_c = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$), effective distance and arrival time. This initial study provided confidence to both groups, and also encouraged us to engage in even more comprehensive and challenging tests.

3. Signal generation

For the study presented in this paper the LIGO and Virgo groups each created 24 h of simulated data. The modeled detectors were the 4 km LIGO systems at Hanford, WA (H1) and Livingston, LA (L1), and the Virgo detector at Cascina, Italy (V1). The noise of the data matched the target sensitivities for the LIGO and Virgo interferometers. For the simulated Virgo noise the only spectral features were violin modes; this noise was Gaussian since the features were produced via simulated thermal noise. For the Virgo data it was also assumed that the power lines were absent since they should be canceled by the data preconditioning, as is the case for the real Virgo data. The generation of the narrow spectral features was made using a thermal noise model, which was Gaussian 1/f noise filtered through a second-order filter to reproduce a thermal mode with a given resonant frequency, quality factor and effective mass. In a similar fashion the lines in the simulated LIGO noise were created using random phase modulation, with the result being colored stationary noise. The lines in the simulated
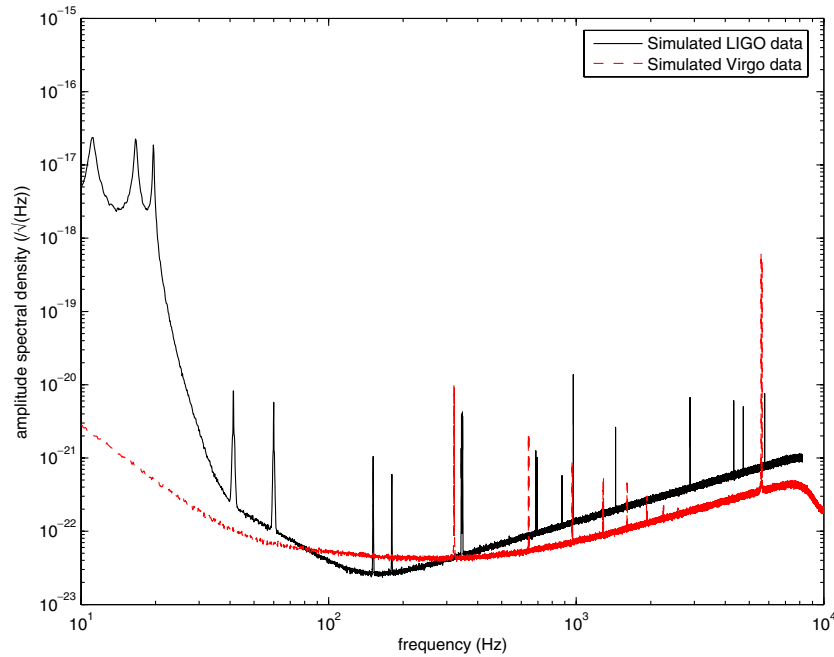
Figure 1. The noise spectrum (at design sensitivity) of the simulated data for LIGO (solid line) and Virgo (dashed line). The distortion close to the Nyquist frequency in the Virgo spectrum is due to the use of a low-pass filter applied before downsampling the Virgo data generated at 40 kHz down to 20 kHz.

LIGO noise were not based on any particular physical models. Figure 1 shows the noise spectral densities for the simulated data.

In order to initiate our study of the LIGO and Virgo inspiral detection capacities we ensured that the 2.0 post-Newtonian (PN) [10] binary inspiral signals for $1.4M_\odot$–$1.4M_\odot$ mass pairs produced by both groups were identical. This exercise highlighted three important subjects that need to be closely monitored in studies dependent on simulated inspiral signal generation. Not surprisingly, it is important to use the same values for physical constants, such as Newton's constant G. Second, the lengths of the inspiral chirps must be accurately calculated. Finally, the definition of the termination frequency of the inspiral signal must be specified consistently. Since the calculated signals come from approximation methods in general relativity it is not surprising that slight discrepancies can occur. One can write the frequency of the waveform as a function of time $f(t)$, or the time before coalescence as a function of the frequency $t(f)$; these two functions used in this study are inverses of each other only at 2.0 PN order. In the LIGO code, the length of the chirp was ascertained by solving $f(t) = f_{\text{low}}$ using a root finder. However, in the Virgo code, the length of the chirp was found by determining $t(f_{\text{low}})$. For a $1.4M_\odot$–$1.4M_\odot$ inspiral beginning at 30 Hz, the LIGO code generated a signal that was 54.6789 s, while the inspiral generated by the Virgo software was 54.6799 s, so the calculated length of the chirp differed by 1 ms. In the test we were performing, this led to an offset between the two chirps. It was possible to obtain agreement between the two waveforms by simply sliding the Virgo waveform backwards by 1 ms. This shift was validated through an inspection of the signals over the entire time period; the two signals were seen to always overlap. For the signal termination there was a slight difference

found between the LIGO and Virgo methods, which was due to the two codes using slightly different ending frequencies. LIGO uses the test mass innermost stable circular orbit (ISCO). Virgo uses the last stable orbit (LSO, the occurrence of the minimum of the dynamical energy, see [10]) calculated at 2.0 PN order, or when the phasing formula defined in [11] breaks down at that same 2.0 PN order (which was the case for the Virgo generated signals in the present study). This resulted in an additional cycle at the very end of the Virgo waveform, but we found that this contributed negligibly to the signal-to-noise ratio for the binary neutron star coalescence, and the detection results.

The results of our previous study [7] were encouraging, but the examination of more realistic signals was needed. We decided to create simulated signals that produce an authentic detection scenario for LIGO and Virgo. When discussing the inspiral of binary neutron star pairs, $[1.4M_\odot–1.4M_\odot]$ systems are a convenient reference. For our definition of the sensitivity of a detector we use the distance of a $[1.4M_\odot–1.4M_\odot]$ binary pair that is in the optimal direction (with the gravitational waves arriving orthogonal to the plane defined by the interferometer arms) and orientation (the orbital plane being parallel to the interferometer plane); such a system producing a signal-to-noise ratio (SNR) of 8 defines the measurement sensitivity for the inspiral signal search. In reality, actual gravitational waves will impinge on the detectors from a variety of directions, with different orbital plane orientations, and disparate polarizations. This produces a decrease in the signal amplitude. When one averages over all directions, orientations and polarizations, the average effective distance to sources is 2.3 times greater than the actual distance [9]. LIGO and Virgo were designed such that at their target sensitivities their inspiral ranges would extend up to 35 Mpc (for an optimally oriented $1.4M_\odot–1.4M_\odot$ binary system). Since the Virgo cluster of galaxies is within this distance (at 16 Mpc), we decided to have this LIGO–Virgo inspiral study respond to signals from the M87 galaxy, as it is the largest galaxy in the Virgo cluster. In addition, other signals were created to simulate emission from NGC 6744 at 10 Mpc. The random orientation of the sources, plus the rotation of the Earth, created signals that produced a wide variety of responses from the detectors. The angle parameters for the signals (phase at coalescence, polarization, the cosine of the orbital plane inclination) were chosen randomly from uniform distributions. The sky locations were determined by the galaxies. The masses were selected from a small set (1.0, 1.4, 2.0 and $3.0M_\odot$). The injections were spaced in time so that they would not significantly affect the power spectrum estimation; the injection times chosen randomly with an average of one injection for every 600 s. Within the 24 h of data 144 inspiral signals were injected. The signals generated by LIGO and Virgo in this present study both used the same values for the physical constants: $G = 6.672\,59 \times 10^{-11}\,\mathrm{m^3 kg^{-1}\,s^{-2}}$, $c = 299\,792\,458\,\mathrm{m\,s^{-1}}$. The LIGO and Virgo signals were adjusted so that the coalescence times and phases were the same. In the end we were satisfied with the overall similarity of the LIGO and Virgo signals, and proceeded with the detection study.

4. Virgo and LIGO inspiral detection routines

Virgo has developed two search pipelines for binary neutron star inspiral signals, the purpose of which is to experiment with different analysis solutions and cross-check their outputs; the plan is to keep developing both methods, because it is anticipated that the two may have different merits when applied to different portions of the parameter space, and/or to different kinds of binary systems (black holes, inclusion of spin, etc). One Virgo method is a multi-band templated analysis (MBTA), whereby the templates are split for efficiency into low- and high-frequency parts [12]. The templates are then subsequently combined together in a hierarchical way. Virgo also has a standard flat-search pipeline, called Merlino [13], that is similar to the

Table 1. Common search parameters for the LSC and Virgo pipelines.

	LIGO dataset	Virgo dataset
Mass range	1–$3M_\odot$	1–$3M_\odot$
Grid minimal match	95%	95%
Starting frequency f_{low}	40 Hz	30 Hz
Longest template duration	$\sim$45 s	$\sim$96 s
SNR threshold	6	6

Table 2. Configuration and computing cost of each analysis.

Pipeline	LIGO	MBTA	Merlino
Data set	LIGO	LIGO	LIGO
Number of templates	$\sim$2 900	$\sim$1 900	$\sim$2 000
Type of processor	1 GHz Pentium II	2.2 GHz Opteron	2.2 GHz Opteron
Total processing time	$\sim$368 h	$\sim$55 h	$\sim$50 h
$\frac{\text{processing time} \times \text{processor speed}}{\text{number of templates}}$	$\sim$457 s GHz	$\sim$229 s GHz	198 s GHz
Pipeline	LIGO	MBTA	Merlino
Data set	Virgo	Virgo	Virgo
Number of templates	$\sim$10 900	$\sim$7 000	$\sim$6 800
Type of processor	2.66 GHz Xeon	2.2 GHz Opteron	2.2 GHz Opteron
Total processing time	$\sim$704 h	$\sim$231 h	250 h
$\frac{\text{processing time} \times \text{processor speed}}{\text{number of templates}}$	$\sim$618 s GHz	$\sim$261 s GHz	291 s GHz

LIGO inspiral pipeline [4, 5]. Both of these Virgo inspiral detection pipelines, as well as the single LIGO pipeline, were applied to the data in this study. A summary of the basic details of the Virgo and LIGO inspiral detection pipelines is given in table 1; detailed descriptions are in the sub-sections below. The range of masses covered by the templates was 1–$3M_\odot$. The template banks were created with a minimal match criterion of 95%, ensuring that no event in that mass space would be detected with a SNR loss greater than 5%. The starting frequency f_{low} for the analysis of the data, in order to be consistent with the SNR accumulation (defined by the sensitivity curves of both experiments) was set to 40 Hz for LIGO data and to 30 Hz for Virgo data. Triggers were recorded when the SNR exceeded a threshold of 6. No χ^2 test [16] was used as a veto for any of the three pipelines since it has little effect with stationary and Gaussian noise. With the trigger lists from each of the pipelines, an event was labeled as *true* if the end time of the inspiral event matched the end time of the injected inspiral event within ± 10 ms.

The mass parameter space layout for the LIGO grid is described in [14]. The Virgo MBTA pipeline creates the grid according to a 2D contour reconstruction technique based on the parameter space metric [15]. The Virgo Merlino template placement is explained in [13]. Table 2 provides information about the way the production was done for the three pipelines and each data set, which type of processor was used and other resources needed. Table 3 summarizes the detection efficiency results of the three inspiral pipelines from their application to the data sets used in this study. Table 4 provides the information on the accuracy of each of the three pipelines for parameter determination for the signals detected in the V1 data. These results are explained in the sections below.

Table 3. The signal detection results for the LIGO, MBTA and Merlino inspiral detection pipelines applied to the L1, H1 and V1 data. The false alarm rate (FAR) for the pipelines applied to the data is also listed.

	% detected in L1 data	% detected in H1 data	% detected in V1 data	FAR for LIGO data	FAR for Virgo data
LIGO pipeline	64%	66%	62%	0.07 Hz	0.77 Hz
MBTA pipeline	62%	61%	56%	0.02 Hz	0.1 Hz
Merlino pipeline	55%	59%	55%	0.03 Hz	0.1 Hz

Table 4. Parameter determination accuracy results for the LIGO, MBTA and Merlino inspiral detection pipelines applied to V1 data. The values in the table are derived from the distribution of all events detected by the respective pipeline. The parameter difference is defined as the actual parameter value subtracted from the pipeline's recovered value; results are given for the chirp mass (in units of $M_\odot$) and end time (in units of s). Also listed is the effective distance fractional error, which is defined as (recovered effective distance $-$ actual effective distance)/(actual effective distance).

	MBTA	Merlino	LIGO
Chirp mass difference mean ($M_\odot$)	2.46×10^{-4}	1.60×10^{-4}	2.79×10^{-4}
Chirp mass difference RMS ($M_\odot$)	9.13×10^{-4}	1.00×10^{-3}	9.04×10^{-4}
End time difference mean (ms)	0.340	0.083	0.966
End time difference RMS (ms)	0.876	1.03	1.29
Effective distance fractional error mean	-0.0031	0.01	-0.026
Effective distance fractional error RMS	0.10	0.11	0.112

4.1. Results for Virgo multi-band analysis of inspiral signals

The Virgo MBTA inspiral detection pipeline was designed to reduce the computational cost of a binary inspiral search. A large number of templates are needed to sufficiently cover the parameter space, especially for binary systems containing relatively smaller masses. The required number of templates depends on the duration of the longest possible signal, and this is affected by the relatively slow frequency evolution of the binary inspiral at lower frequencies. The computational time of a fast Fourier transform (FFT) depends on the size of the data set (hence the frequency span of that data set). The matched filtering technique uses FFTs as part of the computation. The goal of the MBTA technique is to split the analysis into low- and high-frequency parts; these results are subsequently combined in a hierarchical way. A description of the method can be found in [12].

The MBTA pipeline was applied to data from H1, L1 and V1. Detected events with SNR > 6 were recorded, and clustered both in time and over the template bank. Triggers with matching end times (within ±10 ms) were considered the same event, and the trigger with the highest SNR was recorded; if this trigger was within ±10 ms of an injection event end time then it was specified as a detection. There was no χ^2 test [16] used with the MBTA pipeline. The template bank spanned the range from $1M_\odot$ to $3M_\odot$ and had a minimal match of 0.95; a total of 7000 templates were used to analyze the Virgo data set, and 1900 templates for the LIGO data. The Virgo MBTA code was run with a splitting frequency between the low- and high-frequency bands chosen so as to share in an approximately equal way the SNR between the two bands; the band splitting frequency was 130 Hz when applied to the LIGO data, and 95.3 Hz for the Virgo data. With these settings the single instrument false alarm rate was 0.1 Hz for the Virgo data, and 0.02 Hz for the LIGO data. In general, the higher false alarm
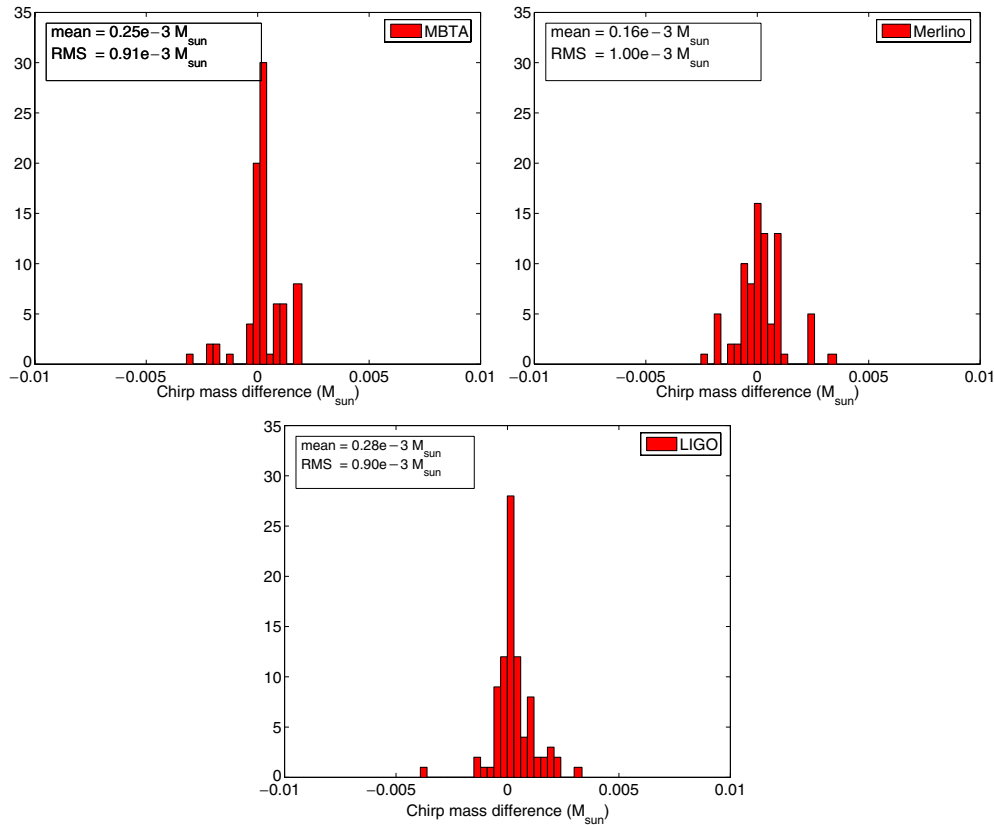
Figure 2. Histograms of the chirp mass determination accuracy for the Virgo MBTA, Virgo Merlino and LIGO inspiral analysis pipelines. These three pipelines were applied to the Virgo data, and displayed are the results from the signals recovered. The *difference* is defined as the actual chirp mass subtracted from the recovered chirp mass.

rate with the Virgo data (seen with all the pipelines) was due to the larger number of templates used, which itself is due to the lower frequency cutoff of the search. Based on the number of signals found in the data set, the single interferometer detection efficiencies for the MBTA pipeline were 61% for signals in the H1 data, 62% for events detected in the L1 data and 56% for signals in the V1 data. The actual value of the efficiency depends entirely on the source population (and resultant effective distance distribution) chosen. Because we have chosen a population for which the efficiency is roughly 50%, we are particularly sensitive to small differences in the pipelines and algorithms, which was the goal of this study. There were a total of 144 injections, so the statistical error on the detection efficiencies was around 4% for each detection pipeline. The injections that were not found had effective distances exceeding the range of the instrument, typically greater than 50 Mpc.

The ability to quantify the accuracy with which we can recover various injected parameters was an important goal for this study of the inspiral detection pipelines. Using the Virgo signal data, histograms were created for the parameter determination difference: see figure 2 for the chirp mass and figure 3 for the end time. The difference is defined as (recovered − injected). Figure 4 displays the effective distance parameter determination in terms of fractional error,
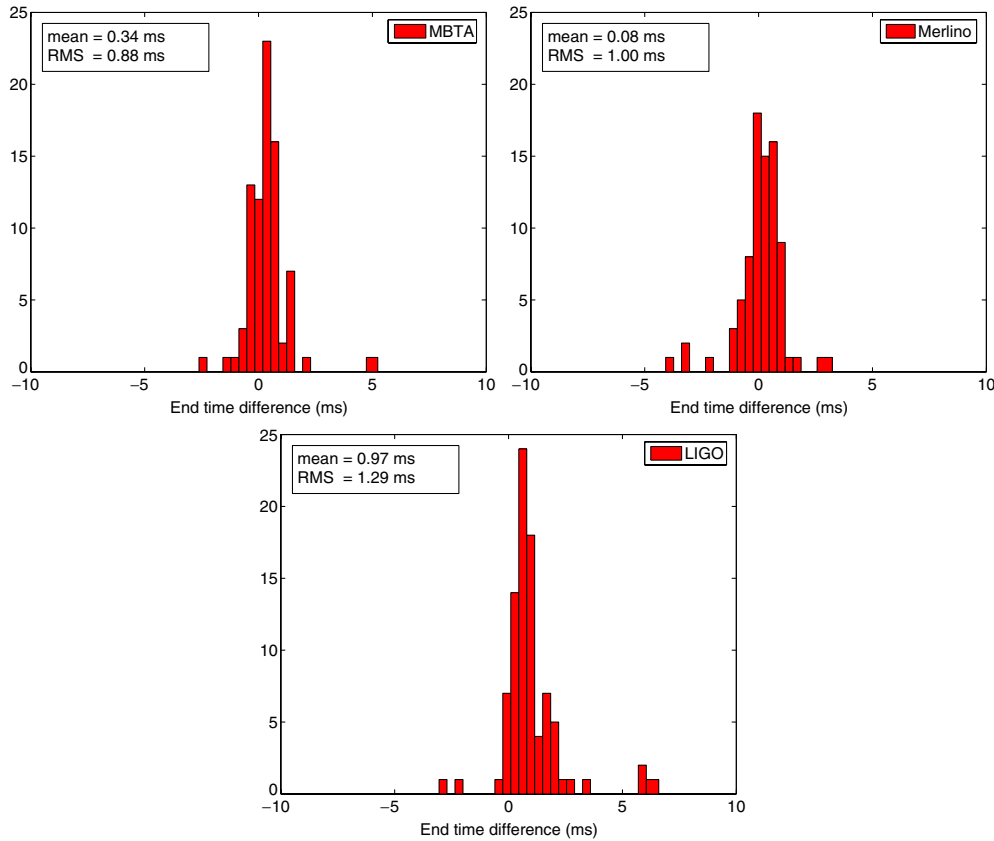
Figure 3. Histograms of the end-time determination accuracy for the Virgo MBTA, Virgo Merlino and LIGO inspiral analysis pipelines. These three pipelines were applied to the Virgo data, and displayed are the results from the signals recovered. The *difference* is defined as the actual end time subtracted from the recovered end time.

while figure 5 displays the fractional error in the detected effective distance versus the actual effective distance. The effective distance fractional error is defined as (recovered effective distance − actual effective distance)/(actual effective distance). In figure 5, one can see that for large injected distances, the recovered distance tends to be less than the injected distance; this was because these events were near to the threshold of the search. If the noise acts to make the signal weaker (i.e. to increase the effective distance) the event will not be detected. However, if the noise acts to make the signal stronger (i.e. to decrease the effective distance) the pipeline will detect the injected signal above threshold; this effect was observed in all the inspiral detection pipelines that we studied.

4.2. Results for Virgo flat band search pipeline of inspiral signals

One of the main goals for Virgo is the realization of a reliable real-time observation strategy in order to use the interferometer as a gravitational wave observatory. With this aim, the computational strategy was carefully designed for a binary inspiral search by addressing the computational size of the problem, available computational resources, communication
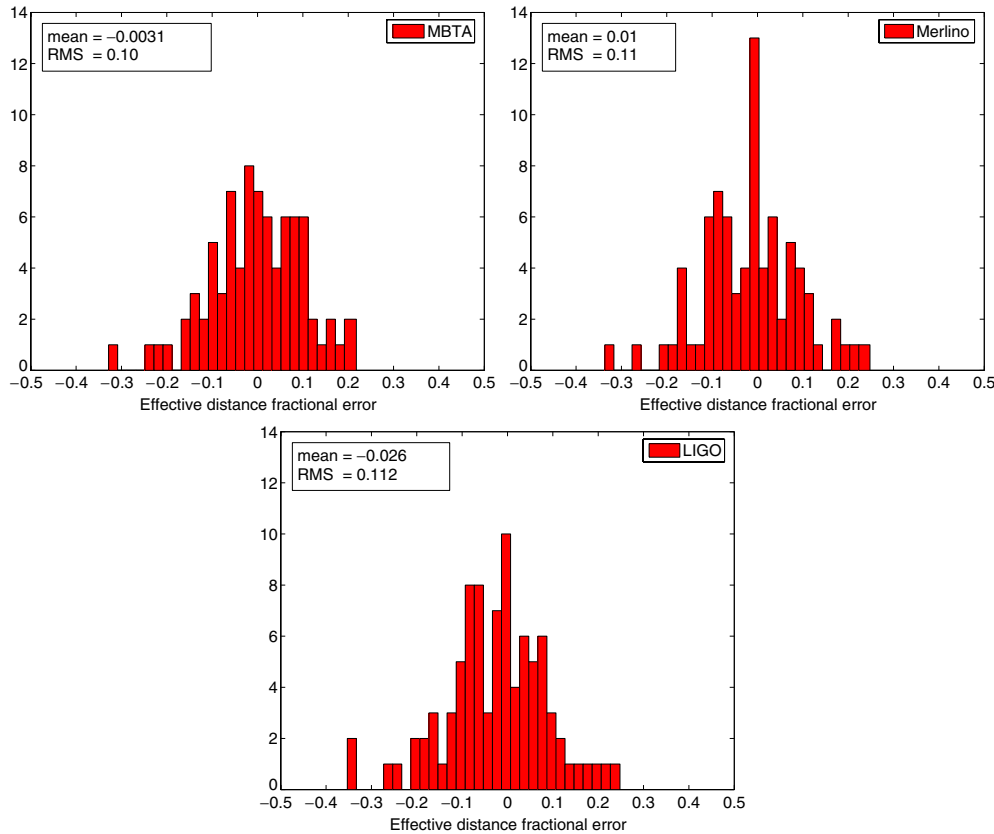
Figure 4. Histogram of the effective distance fractional error for the Virgo MBTA, Virgo Merlino and LIGO inspiral analysis pipelines. These three pipelines were applied to the Virgo data, and displayed are the results from the signals recovered. The effective distance fractional error is defined as (recovered effective distance − actual effective distance)/(actual effective distance).

requirements, data handling and the constraints due to real-time analysis conditions. The distributed signal analyzer (DiSA) [13], named Merlino, is a particular binary inspiral search solution implemented in Virgo, based on a parallel-distributed applications environment. This framework is composed of several processes communicating via message passing interface (MPI), distributing and controlling user algorithms and the data. The algorithms can be dynamically changed and inserted in the Merlino logic data flow using a plug-in strategy. Specifically, the coalescing binaries plug-ins have been used to analyze the data examined in the study presented in this paper; the overlap-add data handling method and the storage in memory of the template bank are implemented in order to further speed-up the analysis [13].

A total of 6800 templates were used to cover the $1.0M_\odot$ to $3.0M_\odot$ range in the Virgo data, while 2000 templates were used for the LIGO data; the template bank had a minimal match of 0.95. A χ^2 value [16] was calculated as part of the Merlino pipeline, however it was not used to veto events. In general, the Merlino χ^2 is calculated using 15 bands. The number of bands was selected by performing qualitative tests on data at the Virgo target sensitivity; various detection tests using a different number of bands were conducted, and in the end 15 bands was determined to be the proper number. Triggers from the Merlino pipeline were required to have
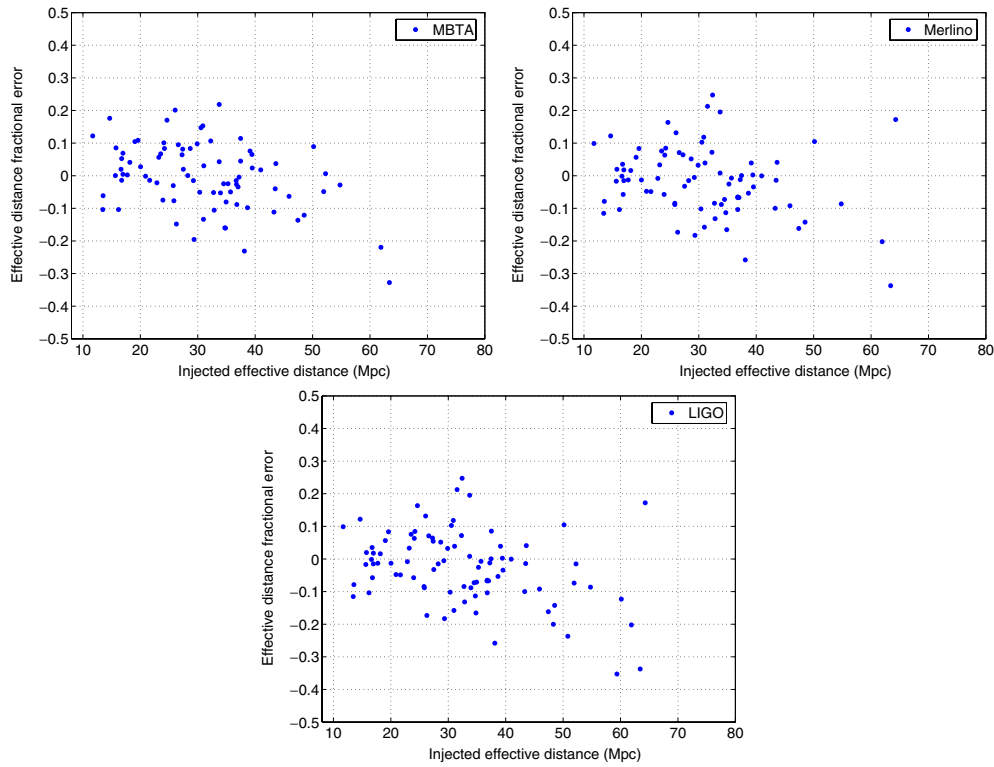
Figure 5. The effective distance determination accuracy as a function of actual effective distance, for the Virgo MBTA, Virgo Merlino and LIGO inspiral analysis pipelines. These three pipelines were applied to the Virgo data, and displayed are the results from the signals recovered. The effective distance fractional error is defined as (recovered effective distance − actual effective distance)/(actual effective distance).

a SNR > 6. All events within ±10 ms were clustered, and the event with the largest SNR was selected as the trigger; if this trigger was within ±10 ms of an injection event end time then it was specified as a detection. This flat search code found 55% of the binary inspiral signals injected into the V1 data, 59% of the signals in the H1 data and 55% of the signals in the L1 data. The false alarm trigger rate was 0.1 Hz for the Virgo data, and 0.03 Hz for the LIGO data.

The performance of the Virgo Merlino pipeline to resolve signal parameters was comparable to the other pipelines, and example results are displayed here for the analysis of the Virgo data set. Figure 2 displays the ability of the Merlino code to accurately determine the chirp mass. The accuracy of the end-time parameter is presented in figure 3, while figure 4 displays the effective distance parameter estimation. Figure 5 also displays the fractional error in the detected effective distance versus the actual effective distance.

4.3. Results for LIGO inspiral detection routine

The LIGO inspiral detection pipeline has been used to search for signals and set upper limits with data from LIGO's first two scientific data runs [4, 5]. These publications also present detailed descriptions of the LIGO inspiral pipeline. This same LIGO inspiral pipeline was

applied to the 24 h of data for this present study. The simulated signals from H1, L1 and V1 were all examined with the LIGO code.

For the initial single detector test a threshold of SNR > 6 was used, with no χ^2 threshold, or mass consistency check. The χ^2 signal consistency test [16] was not used as it has little effect with stationary, Gaussian noise. The data was high-pass filtered (including the injections) for each interferometer individually. The template bank spanned the range from $1M_\odot$ to $3M_\odot$ and had a minimal match of 0.95; a total of 10 900 templates were used to analyze the Virgo data set, and 2900 templates for the LIGO data. When calculating the detection efficiency we required that a candidate trigger occur within ±10 ms of the injected signal's end time. There was no clustering over the template bank. For the H1 data 66% of the inspiral signals were detected, while the pipeline found 64% of the L1 signals. The efficiency of signal detection efficiency was 62% for the V1 data. For the LIGO pipeline the false alarm trigger event rate was 0.07 Hz when analyzing the LIGO data and 0.77 Hz for the Virgo data set.

The recovered parameter values correspond to the template producing the largest SNR trigger within 10 ms of the actual end time. Using the Virgo signal data, we created histograms of the parameter determination difference (figure 2 for the chirp mass, figure 3 for the end time and figure 4 for the effective distance), as well as a plot, figure 5 that displays the fractional error in the detected effective distance versus the actual effective distance.

5. MCMC parameter estimation

Parameter estimation, and the generation of *a posterior* probability density function (PDF) for each parameter, was also done utilizing a Markov chain Monte Carlo (MCMC) routine. These MCMC methods are part the LIGO binary inspiral data analysis effort. The basic operation of the inspiral MCMC code is described in [17], which also contains a description of MCMC techniques. The purpose of this code is to take triggers generated from the LIGO inspiral pipeline, and then examine that section of the data about the event. This MCMC code was applied to those sections of data in this study where coincident events were found. The MCMC is not a search pipeline, as the program is too computationally taxing to be applied to all of the data. The MCMC searched for events that had a binary coalescence end time within a ±50 ms window of coincident triggers from the LIGO inspiral search pipeline.

The MCMC code used in the present study was written in C, and looked for inspiral signals based on 2.0 PN signals in the frequency domain. For the present study the prior for the masses of the compact objects was uniform from $0.9M_\odot$ to $3.1M_\odot$ range. For this problem there are five parameters to estimate: the binary masses, m_1 and m_2, the effective distance d_L, the phase at coalescence ϕ_c and the time at coalescence t_c. The program reparametrizes the masses in terms of the chirp mass $m_c = (m_1m_2)^{3/5}/(m_1+m_2)^{1/5}$ and the mass ratio parameter $\eta = m_1m_2/(m_1+m_2)^2$. This inspiral MCMC code begins with a method called *importance resampling* [18, 19]; at the start the code first generates a large sample of parameter space points from a distribution covering the whole prior, and then draws the actual sample out of these with correspondingly assigned weights depending on the posterior density. The Markov chains were started in regions of parameter space that were *likely* to be close to the true parameter values. Simulated annealing [20] was used to optimize the initial *burn-in* of the Markov chains. During the burn-in period, the effect of the noise in the likelihood function was arbitrarily increased (an effective temperature increase); this simulated annealing technique was introduced in [21] and allows scanning of the whole parameter space by permitting larger steps. The candidate generating function [17] for the parameters was designed so that correlations between chains were measured as the program progressed, and correlation values were fed back to the generating function.
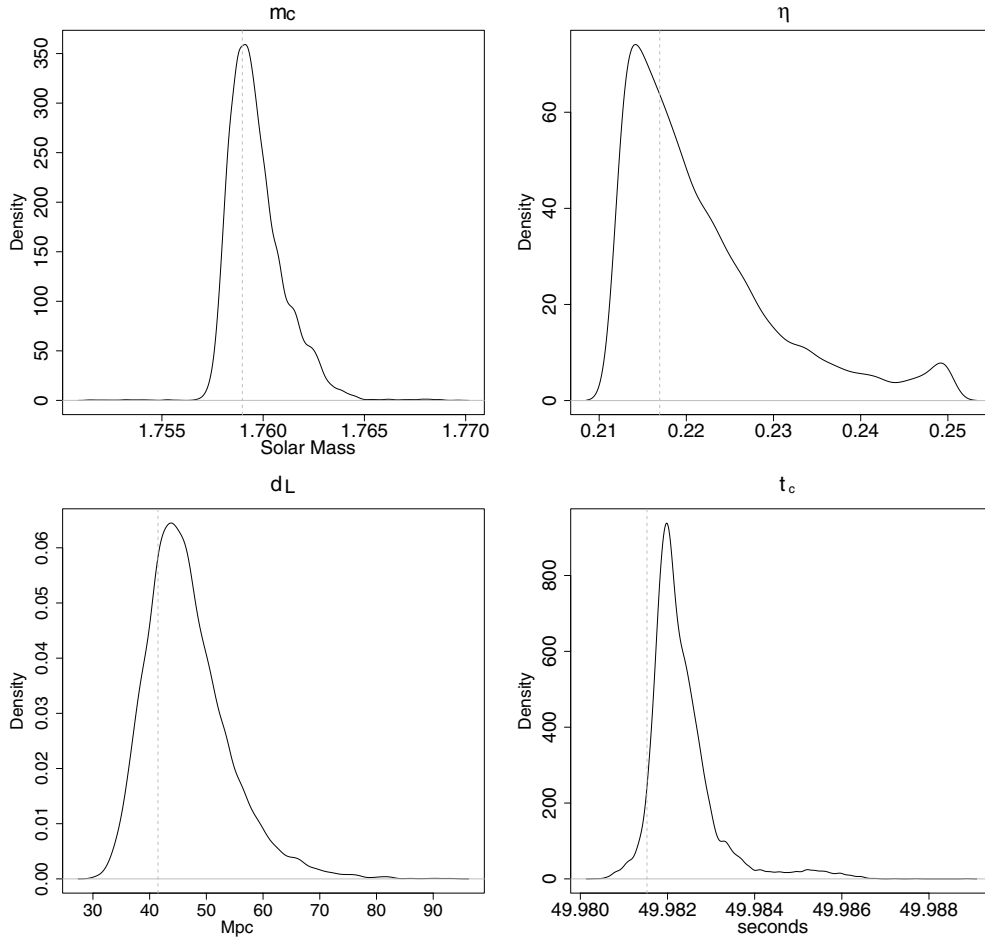
Figure 6. MCMC produced posterior PDFs for chirp mass m_c, mass ratio η, effective distance d_L and the coalescence time t_c for one event. In this example, the binary inspiral signal is embedded in the H1 data. The actual parameters used to produce this signal, displayed by the dashed vertical lines, are $m_1 = 1.4 M_\odot$, $m_2 = 3.0 M_\odot$ ($m_c = 1.759 M_\odot$ and $\eta = 0.217$), $d_L = 41.47$ Mpc and $t_c = 49.9815$ s.

The MCMC parameter estimation code was applied to all of the LIGO–Virgo data from this study, and here we show examples from the analysis of the LIGO signals. Figure 6 shows an example of MCMC generated estimates for the posterior probability distribution functions for a simulated event in the H1 data detected with the LIGO pipeline. This signal had real parameter values of $m_1 = 1.4 M_\odot$, $m_2 = 3.0 M_\odot$ ($m_c = 1.759 M_\odot$ and $\eta = 0.217$), $d_L = 41.47$ Mpc and $t_c = 49.9815$ s. From the generated posterior PDFs estimates for the parameter values can be associated with the mean of the distribution. The error in the parameter estimate would correspond to a particular width. We will give the 5–95 percentile range of the posterior distribution, which gives a 90% posterior credibility interval. For this example the parameters' mean values and 90% credibility ranges are $m_c = 1.7597(1.7580\text{–}1.7623) M_\odot$, $\eta = 0.2215(0.2126\text{–}0.2407)$, $d_L = 46.669(36.936\text{–}60.023)$ Mpc and $t_c = 49.9823(49.9815\text{–}49.9837)$ s.
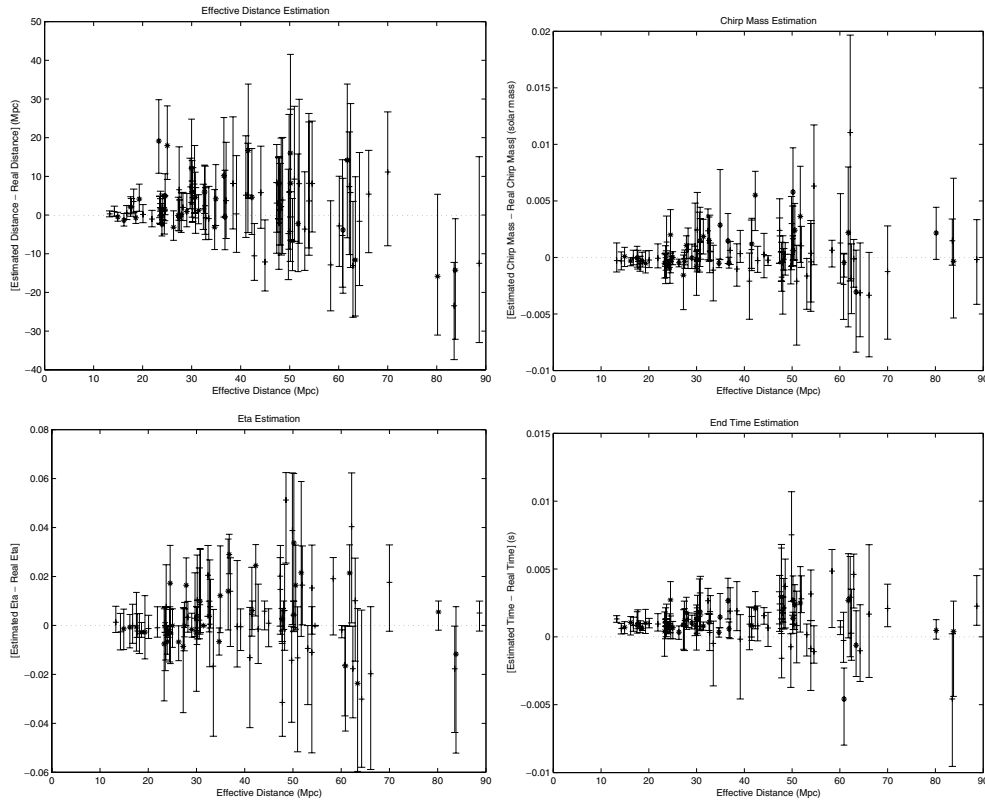
Figure 7. The parameter estimation accuracy for the MCMC chirp mass m_c, mass ratio η, effective distance d_L and the coalescence time t_c are shown. Displayed are the differences between the MCMC parameter estimated and the actual value, as a function of effective distance. These events were from the H1 and L1 data. The error bars correspond to the 90% credibility ranges.

As an example of the parameter estimation accuracy, figure 7 shows the difference between the MCMC estimated parameter values and the real values, as a function of effective distance. The errors represent the 90% credibility ranges. These events were from the simulated inspiral signals found by the LIGO pipeline in both the L1 and H1 data sets. All mass combinations are represented in this plot.

6. LIGO and Virgo inspiral detection pipeline comparison

The Merlino, MBTA and LIGO binary inspiral pipelines all operated in a comparable way, and produced good single interferometer detection statistics. Triggers were recorded, and those with the largest SNR had the parameter values for the templates noted. The ability to resolve the parameters was equally positive (within statistical errors and other uncertainties) for all of the pipelines. A major conclusion from our study is that the three inspiral search pipelines performed equally well, and there is mutual confidence between the groups in the others' inspiral search abilities. The comparisons presented in this section were all conducted using the Virgo data.
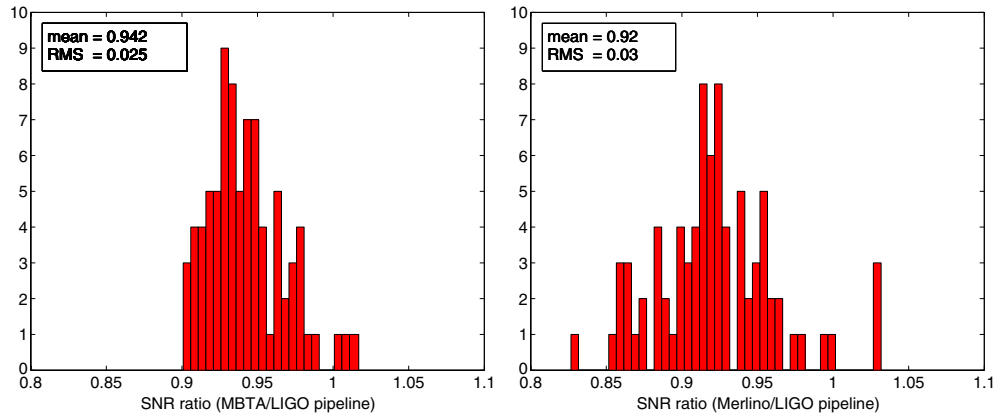
Figure 8. On the left is a histogram of the ratios of the SNRs for MBTA and LIGO events detected by both pipelines in the V1 data. The mean and RMS values for the distribution are given in the figure, showing a 6% excess in the size of the LIGO SNR. On the right is a histogram of the ratios of the SNRs for Merlino and LIGO events detected by both pipelines in the V1 data; there was a 8% excess in the size of the LIGO SNR.

In order to display the output of the pipelines, we present a direct comparison of the Merlino, MBTA and LIGO results from the V1 data. Figure 8 (left) shows a histogram of the ratio of the MBTA SNR to the LIGO SNR for the events detected. Similarly, figure 8 (right) shows a histogram of the ratio of the Merlino SNR to the LIGO SNR for the events detected. There were some slight differences in the results between the pipelines. For example, the SNR of detected events was about 6% larger from the LIGO pipeline versus those from the MBTA pipeline, and about 8% larger than those from the Merlino pipeline.

The difference in the SNRs was predominantly affected by slightly differing methods for calculating the noise power spectral density (PSD). Specifically, the LIGO pipeline calculated its PSD via the median power spectrum (the median is calculated by frequency-bin) of 15 overlapping segments, where each segment was 256 s in length. The median was chosen for the LIGO pipeline PSD generation so that it would not be overly biased by a single loud glitch (as could happen when using the mean), at the cost of a larger bias in the case of Gaussian noise. The PSD calculated by the Virgo pipeline was computed from a larger number of averages, as the mean from 1800 s of data, each section 16.38 s long, with no overlap. We conducted a direct comparison of the PSDs generated by the LIGO and Virgo pipelines. The estimate of the SNR^2 scales with frequency like $f^{-7/3}/S(f)$, where $S(f)$ is the noise PSD. When we summed $f^{-7/3}/S(f)$ over the frequencies from 40 Hz to 2048 Hz the LIGO result exceeded that of Virgo by 10%, giving an overestimation in the LIGO SNR of 5%. Numerical experiments have subsequently verified that increasing the number of segments used in calculating the noise PSD will reduce the bias. It should be noted that the bias from the noise PSD strictly cancels out in the estimation of the effective distance.

We were also concerned that the frequency domain stationary phase approximation used by LIGO could affect the SNR estimation. Specifically, the signals for this study were generated to 2.0 PN order in the time domain. For both Virgo pipelines the frequency domain detection templates were the Fourier transforms of the time domain 2.0 PN signals. On the other hand, the LIGO pipeline used 2.0 PN frequency domain templates generated via a stationary phase approximation. We studied the overlap between Fourier transforms of 2.0 PN time domain

signals and the stationary phase frequency domain signals. Our results indicate that the SNR for the LIGO would be reduced by at most 1% (and only for signals from the highest mass pairs). In addition we verified the *effectualness* and *faithfulness* of the signals used in this study, similar to what was done in [22]. Specifically for the mass pairs used in this study, the faithfulness varies between 96% and 98%. Lower values correspond to the larger mass ratios. The effectualness is always above 99%, except for the $3M_\odot$–$3M_\odot$ mass pair, for which the effectualness is 98.8%. The accuracy in the estimation of the chirp mass using the stationary phase templates ranges from a few parts in 10^{-5} to a few parts in 10^{-4}.

We also examined how template placement in the m_1 versus m_2 plane grid would affect the recovered SNR, and we found that for our studies this had a small effect on the SNR difference. The density of the template grid was higher for the LIGO pipeline than the Virgo pipelines. Specifics of the template placement for the inspiral pipelines used in our study are presented in [7]. When examining the SNR ratio from the MBTA and LIGO pipelines we found the excess value from the LIGO pipeline to be consistently present for the cases when the two masses were equal, or when they had a relatively large difference. The increased density of the LIGO grid would give an excess in SNR, but for the study presented here it appears that the effect would not produce a SNR difference exceeding 1%.

From these studies, we believe that we understand how differences were created in the SNR of the inspiral triggers by the LIGO and Virgo pipelines. The slightly higher detection efficiency produced by the LIGO pipeline was a direct consequence of its elevated value of the produced SNR ratio, which in turn was due to the difference in the methods used to estimate the noise PSD. All three pipelines in this study used a SNR > 6 threshold. Those events near this cutoff would have a slight preference of being seen by the LIGO pipeline. When one accounts for the SNR artifact the detection efficiencies of all of the pipelines were seen to be the same.

Another small difference in the LIGO and Virgo pipelines concerns the estimate of the end time. The LIGO pipeline typically estimated an end time that was 1 ms off from the actual value. This can be seen in the output of the LIGO pipeline in figure 3 and the MCMC generated parameter estimates in figure 7. The 1 ms offset in the LIGO end-time estimation is also apparent in figure 9 (left), where the end-time accuracy for MBTA and LIGO events detected by both pipelines in the V1 data is presented. Similarly, a comparison of the LIGO and Merlino estimates of the end time also display an offset of about 1 ms for the LIGO results (see figure 9 (right)). The small difference in the nature of the detection templates (frequency domain stationary phase for LIGO, and Fourier transforms of the time domain templates for Virgo) was responsible for the small time shift.

Figure 10 (left) displays a comparison of the accuracy of the determination of the chirp mass for events detected by both the MBTA and LIGO pipelines from within the V1 data set, while a similar chirp mass comparison from the LIGO and Merlino pipelines is displayed in figure 10 (right). In figure 11 (left), we see the recovered effective distance divided by the actual injected effective distance for the signals detected by MBTA and LIGO in the V1 data, while figure 11 (right) shows a similar effective distance comparison between the Merlino and LIGO pipelines. The chirp mass and effective distance estimates by LIGO, Merlino and MBTA were essentially the same.

7. Benefits from a combined LIGO–Virgo inspiral search

This close examination of the LIGO and Virgo binary inspiral search pipelines has also provided additional information on the benefits one can get in conducting a mutual search for signals. In the effort to detect binary inspiral gravitational waves the inclusion of Virgo
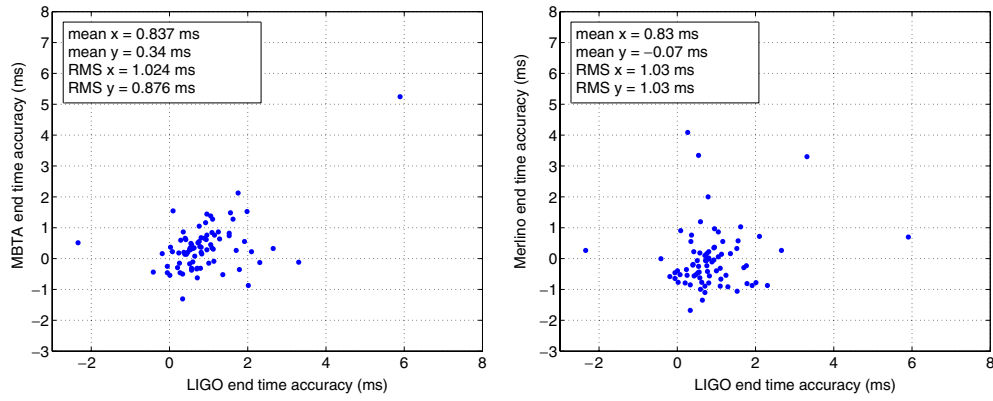
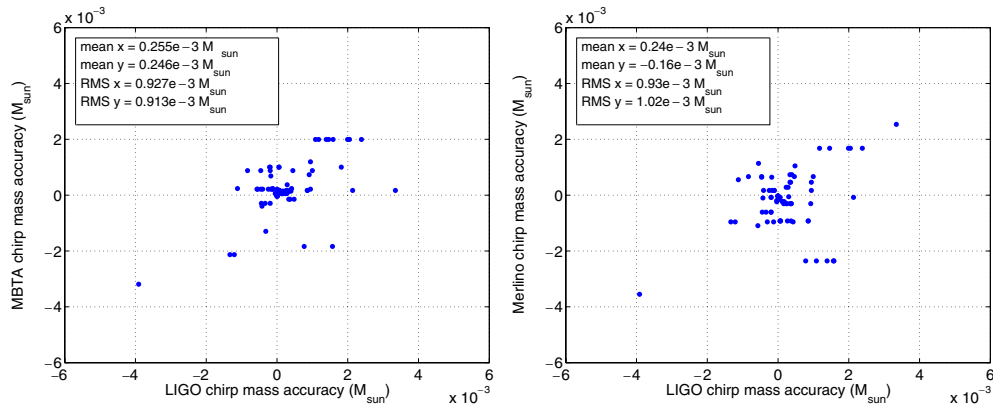**Figure 9.** On the left is a scatter plot of the end-time accuracy for MBTA and LIGO events detected by both pipelines in the V1 data. The accuracy is defined as the actual end time subtracted from the recovered end time. The mean and RMS values are given in the figure, and show that the LIGO inspiral pipeline tends to overestimate the end time by about 1 ms. On the right is a scatter plot of the end-time accuracy for Merlino and LIGO events detected by both pipelines in the V1 data; again, the LIGO inspiral pipeline tends to overestimate the end time by about 1 ms.

Figure 10. On the left is a scatter plot of the chirp mass accuracy for MBTA and LIGO events detected by both pipelines in the V1 data. On the right is a scatter plot of the chirp mass accuracy for Merlino and LIGO events detected by both pipelines in the V1 data. The accuracy is defined as the actual chirp mass subtracted from the recovered chirp mass. The mean and RMS values are given in the figures, and show that the Virgo MBTA, Virgo Merlino and LIGO inspiral pipelines all estimated the chirp mass accurately.

data significantly increases the probability for observing a coincident signal in at least two interferometers. The results presented above show that each pipeline was able to effectively detect events with sufficiently low effective distance, and recover the chirp mass, coalescence time and effective distance. Since the effective distance value is influenced by detector orientation and source location, it is not a parameter that can be used as a test for coincident detection. However, the chirp mass and coalescence time can be used to require consistency. We set coincident window sizes of $0.2M_\odot$ for the chirp mass and ± 8 ms about the light travel time between interferometers for the coalescence time (for reference, the light travel
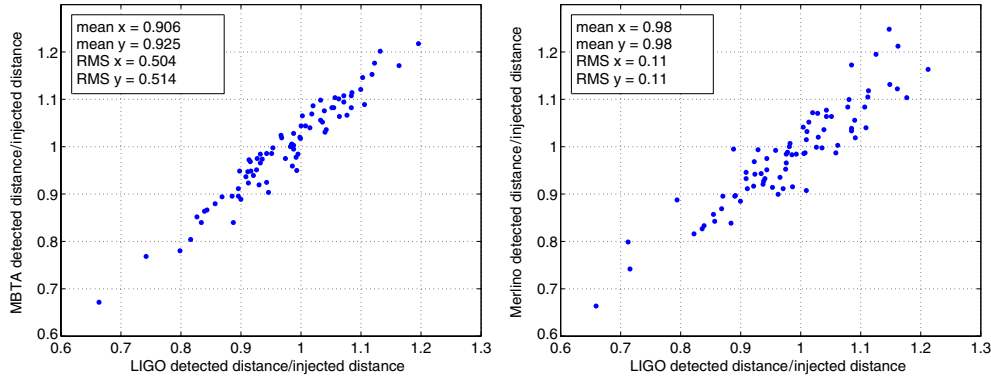
Figure 11. On the left is a scatter plot of the effective distance determination ratio for MBTA and LIGO events detected by both pipelines in the V1 data; on the right is the same but for Merlino and LIGO events. The parameter determination ratio is defined as the detected effective distance divided by the actual injected effective distance.

Table 5. The efficiency of detecting inspiral injections from NGC 6744 (at a distance of 10 Mpc) and M87 in the Virgo cluster (at a distance of 16 Mpc) using different combinations of the LIGO and Virgo detectors and an SNR threshold of 6 in all detectors. These results are from the MBTA pipeline.

	HLV	HL	HV	LV	HL $\cup$ HV $\cup$ LV
NGC 6744 efficiency	48%	65%	54%	49%	72%
M87 efficiency	24%	42%	32%	30%	56%

times between the detectors are 27.2 ms for Hanford–Virgo, 26.4 ms for Livingston–Virgo and 10 ms for Hanford–Livingston), and a SNR > 6 threshold. With these settings we found no triple coincident false events, and only one double coincident false alarm.

The double coincidence detection ability for the injected signals increased with the inclusion of V1 over just the H1–L1 coincidence alone. This is summarized in table 5, where the triple coincidence results and various two detector coincidence results are presented. The efficiency for detection of injections from NGC 6744 is larger since it is closer than M87.

The coincidence results show the benefits of performing a search including all three detectors. The highest efficiency was obtained by requiring a signal to be observed in any two of the three detectors. For the closer NGC 6744 galaxy, the main advantage of adding the Virgo detector to a LIGO only search was the good triple coincident efficiency. Not only was the triple coincident false alarm rate very low, but also with a trigger in three detectors we can reconstruct the sky location of the source.

For signals from both M87 and NGC 6744, the two detector LIGO efficiency was greater than either the H1–V1 or L1–V1 efficiency. This was to be expected due to the similar orientations of the two LIGO detectors. However, by including Virgo and requiring a coincident trigger in two of the three detectors, we do obtain a 25% relative increase in efficiency. The M87 galaxy is in the Virgo cluster, which contains a significant fraction of potential binary neutron star inspiral sources for the initial interferometric detectors. A 25% increase in efficiency to these sources significantly increases the chance of making a detection.

The reason for the increase in two detector efficiency can be seen in figure 12. Displayed are the detected and missed signals for the LIGO pipeline on the H1 data, the Virgo Merlino
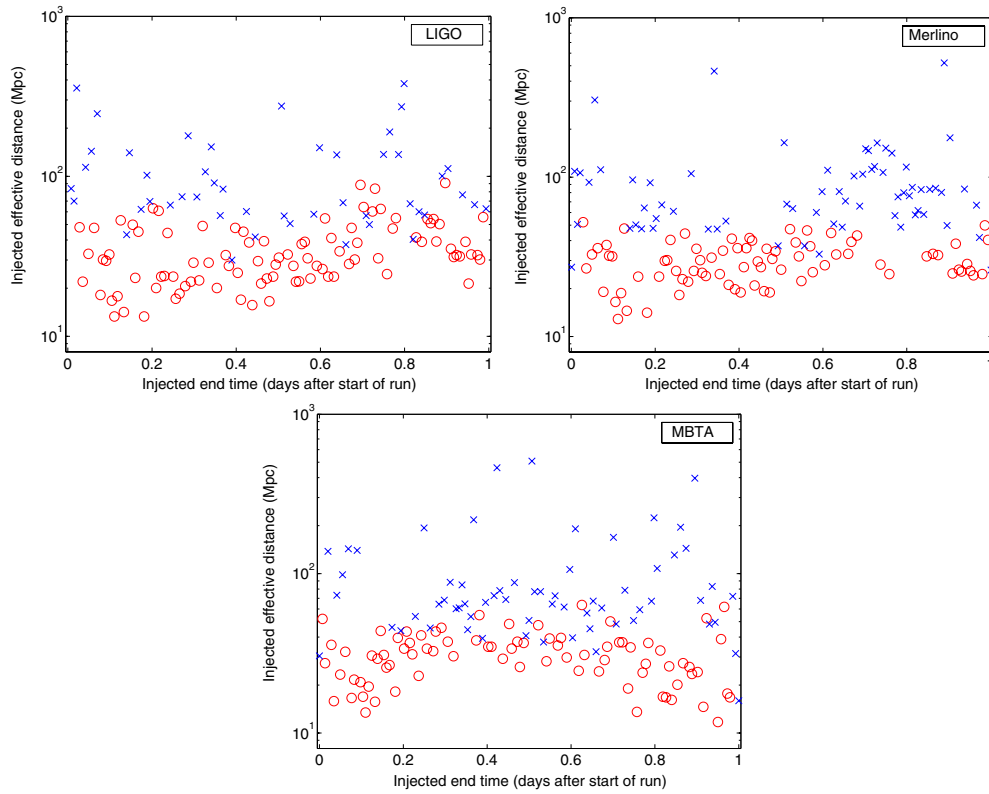
Figure 12. The detected (o) and missed (x) events for the LIGO pipeline on the H1 data, the Virgo Merlino pipeline on the L1 data, and the Virgo MBTA pipeline on the V1 signals. Note too that the dependence of detection efficiency versus effective distance can be seen in these plots; events with an effective distance in excess of 50 Mpc are difficult to detect.

pipeline on the L1 data, and the Virgo MBTA for the V1 signals. One can see in these plots that the reason for signals being undetected, in all three pipelines, was due to large effective distance values. It can be seen that over the course of 24 sidereal hours there was a variation in the ability to detect signals from the two source galaxies due to changes in interferometer alignment as the earth rotates. Virgo's inclusion provides detections at times when the LIGO interferometer network's alignment may be sub-optimal.

7.1. Source directional information

The use of three widely spaced interferometers provides the opportunity for identifying the location on the sky of a binary inspiral event. This is another positive outcome of a combined LIGO and Virgo search. Figure 13 shows the recovered sky position for events that were successfully detected in H1, L1 and V1. The MBTA pipeline was used to identify the triple coincidences using clustered triggers; for this demonstration we selected the highest SNR trigger within ± 10 ms of injection, in each interferometer.

The determination of the sky position is affected by the estimates for the other parameters, as the reconstructed values of the parameters are not independent. For example, a higher mass binary inspiral will traverse the sensitive band of the detectors more rapidly than one of lower
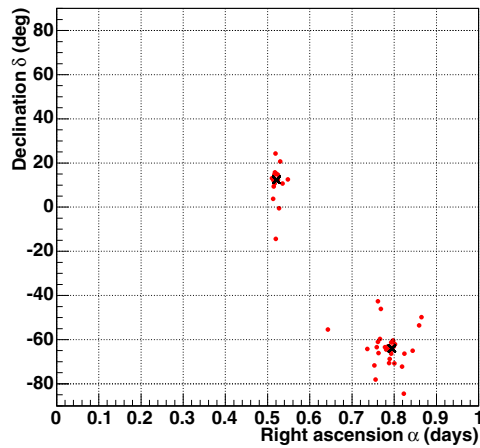
Figure 13. The recovered (dots) and injected (crosses) sky locations of the inspiral injections seen in all three detectors. For reference, the galaxy NGC is located at $\alpha = 19$ h 9.8 min, $\delta = -64°$ and M87 is located at $\alpha = 12$ hr 30.8 min, $\delta = 12°$.

mass. The reconstructed coalescence time and masses of the system will be correlated. These correlations make it difficult to determine the coalescence time, and then sky position, with good accuracy. It was possible, however, to improve on the sky position determination. Again using the MBTA pipeline, the search for a signal in H1 and L1 data was restricted such that the identified clusters for individual triggers were only those issued by the same template as the one leading the identified cluster in the V1 data (choosing the highest SNR if more than one is found). Specifically, the analysis of the H1 and L1 data was redone with the same template grid as was used to analyze the V1 data. However, if no trigger with the same templates as in V1 can be found in H1 or L1, then L1 and V1 are examined for triggers with the same masses as in H1, and if not successful do the same with L1 as a reference. By conducting the search accordingly the ability to estimate the mass parameters (and hence the end time and sky position) improves, and we end up with 39 injections (out of 49 triple coincident detections) found with the same template in $m_1 m_2$ space. The order in which each interferometer is used in turn as a reference is somewhat arbitrary. However for the present study the reason for starting with Virgo is due to the better mass resolution of using the V1 data (due to the better low frequency sensitivity). Figure 14 shows the sky position accuracy obtained after enforcing this mass correction technique. The position accuracy distribution (recovered position minus true position) for the 16 events from galaxy M87 had a mean of 4.1° and an RMS of 3.5°, while for the 23 events from galaxy NGC 6744 the position accuracy distribution had a mean of 2.3° and an RMS of 1.2°.

8. Summary

LIGO and Virgo both have efficient data analysis pipelines for detecting gravitational waves from the inspiral of binary neutron star systems. In the near future LIGO and Virgo will undertake a mutual search for binary inspiral signals. Both groups now have confidence in each other's ability to accurately detect these signals. The results presented in this paper validate the performance of all the LIGO and Virgo inspiral search pipelines.

A combined search for inspiral signals by LIGO and Virgo will also produce significant advantages. The probability for an event to be seen simultaneously by two detectors increases
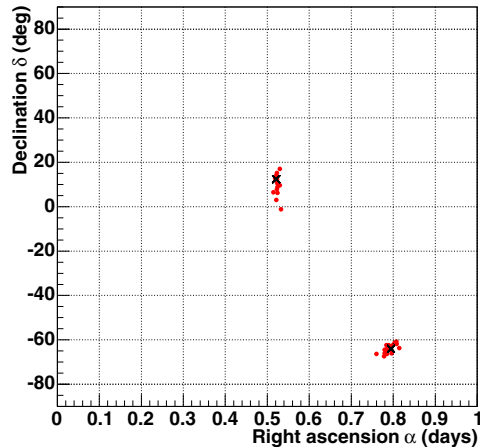
Figure 14. The recovered (dots) and injected (crosses) sky locations of the subset of inspiral injections which are seen in all three detectors using exactly the same template in $m_1 m_2$ space. The 39 detected events in this figure were the ones found to have the same mass parameters for the triggers, as described in the text.

significantly when Virgo is included with LIGO in an inspiral search. Also, it will be possible to determine the sky position of an inspiral event when observed by Virgo and interferometers at the two LIGO locations. Our future data analysis goals involve moving to the analysis of real data from the LIGO and Virgo interferometers. This will create additional issues that will need to be studied, such as data quality criteria and the use of veto channels.

Acknowledgments

References

[1] Abbott B *et al* 2004 *Nucl. Instrum. Methods Phys. Res.* A **517** 154
[2] Amens F *et al* 2005 *Class. Quantum Grav.* **22** S869
[3] Shapiro S L and Teukolsky S A 1983 *Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects* (New York: Wiley-Interscience)
[4] Abbott B *et al* 2004 *Phys. Rev.* D **69** 122001
[5] Abbott B *et al* 2005 *Phys. Rev.* D **72** 082001
[6] Beauville F *et al* 2008 A comparison of methods for gravitational wave burst searches from LIGO and Virgo *Class. Quantum Grav.* **25** 045002
[7] Beauville F *et al* 2005 *Class. Quantum Grav.* **22** S1149
[8] Beauville F *et al* 2005 *Class. Quantum Grav.* **22** S1293
[9] Finn L S and Chernoff D F 1993 *Phys. Rev.* D **47** 2198
[10] Blanchet L, Iyer B R, Will C M and Wiseman A G 1996 *Class. Quantum Grav.* **13** 575
[11] Blanchet L, Damour T and Iyer B R 1995 *Phys. Rev.* D **51** 5360
Blanchet L, Damour T and Iyer B R 1996 *Phys. Rev.* D **54** 1860 (erratum)

[12] Marion F *et al* 2004 Gravitational waves and experimental gravity *Proc. Rencontres de Moriond 2003* (Gif-sur-Yvette: Éditions Frontières)

[13] Amico P, Bosi L, Cattuto C, Gammaitoni L, Marchesoni F, Punturo M, Travasso F and Vocca H 2003 *Comput. Phys. Commun.* **53** 179

[14] Owen B J and Sathyaprakash B S 1999 *Phys. Rev.* D **60** 022002

[15] Beauville F *et al* 2003 *Class. Quantum Grav.* **20** S789

[16] Allen B 2005 *Phys. Rev.* D **71** 062001

[17] Christensen N, Libson A and Meyer R 2004 *Class. Quantum Grav.* **21** 317

[18] Gelman A, Carlin J B, Stern H and Rubin D B 1997 *Bayesian Data Analysis* (London: Chapman and Hall)

[19] Gelman A and Meng X-L 1998 *Stat. Sci.* **13** 163

[20] Gilks W R, Richardson S and Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall)

[21] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087

[22] Damour T, Iyer B R and Sathyaprakash B S 2000 *Phys. Rev.* D **62** 084036