

THE FIRST TWO YEARS OF ELECTROMAGNETIC FOLLOW-UP WITH ADVANCED LIGO AND VIRGO

LEO P. SINGER¹, LARRY R. PRICE¹, BEN FARR^{2,3}, ALEX L. URBAN⁴, CHRIS PANKOW⁴, SALVATORE VITALE⁵, JOHN VEITCH^{6,3},
WILL M. FARR³, CHAD HANNA^{7,8}, KIPP CANNON⁹, TOM DOWNES⁴, PHILIP GRAFF¹⁰, CARL-JOHAN HASTER³, ILYA MANDEL³,
TREVOR SIDERY³, AND ALBERTO VECCHIO³

Revised draft submitted to ApJ

ABSTRACT

We anticipate the first direct detections of gravitational waves with Advanced LIGO and Virgo later this decade. Though this groundbreaking technical achievement will be its own reward, a still greater prize could be observations of compact binary mergers in both gravitational and electromagnetic channels simultaneously. During Advanced LIGO and Virgo's first two years of operation, 2015 through 2016, we expect the global gravitational-wave detector array to improve in sensitivity and livetime and expand from two to three detectors. We model the detection rate and the sky localization accuracy for binary neutron star mergers across this transition. We have analyzed a large, astrophysically motivated source population using real-time detection and sky localization codes and higher-latency parameter estimation codes that have been expressly built for operation in the Advanced LIGO/Virgo era. We show that for most binary neutron star events the rapid sky localization, available about a minute after a detection, is as accurate as the full parameter estimation. We demonstrate that Advanced Virgo will play an important role in sky localization, even though it is anticipated to come online with only 1/3 as much sensitivity as the Advanced LIGO detectors. We find that the median 90% confidence region shrinks from $\sim 500 \text{ deg}^2$ in 2015 to $\sim 200 \text{ deg}^2$ in 2016. A few distinct scenarios for the first LIGO/Virgo detections emerge from our simulations.

Keywords: gravitational waves — stars: neutron — surveys

1. INTRODUCTION

We expect this decade to bring the first direct detection of gravitational waves (GWs) from compact objects. The LIGO and Virgo detectors are being rebuilt with redesigned mirror suspensions, bigger optics, novel optical coatings, and higher laser power (Harry 2010; Acernese et al. 2013). In their final configuration, Advanced LIGO and Virgo are expected to reach ~ 10 times further into the local Universe than their initial configurations did. The best-understood sources for LIGO and Virgo are binary neutron star (BNS) mergers. They also offer a multitude of plausible electromagnetic (EM) counterparts (Metzger & Berger 2012) including collimated short-hard gamma-ray bursts (short GRBs; see for example Paczynski 1986; Eichler et al. 1989; Narayan et al. 1992; Nakar 2007; Rezzolla et al. 2011) and X-ray/optical afterglows, near-infrared kilonovae (viewable from all angles; Li & Paczyński 1998; Barnes & Kasen 2013, etc.), and late-time radio emission (Nakar & Piran 2011; Piran et al.

2013). Yet, typically poor GW localizations of $\gtrsim 100 \text{ deg}^2$ will present formidable challenges to observers hunting for their EM counterparts.

Several planned optical astronomy projects with a range of fields of view and apertures are preparing to pursue these elusive events. These include Zwicky Transient Facility (Kulkarni 2012), PanSTARRS¹¹, BlackGEM¹², and LSST (Ivezic et al. 2008), to name a few. Advanced LIGO is scheduled to start taking data in 2015 (Aasi et al. 2013b). Preparations for joint EM and GW observations require a complete understanding of when and how well localized the first GW detections will be. Plausible scenarios for the evolution of the configuration and sensitivity of the worldwide GW detector network as it evolves from 2015 through 2022, as well as rough estimates of sky localization area, are outlined in Aasi et al. (2013b).

To provide a more realistic and complete picture, we have conducted Monte Carlo simulations of the 2015 and 2016 detector network configurations, probing the transition from two to three detectors as Advanced Virgo is scheduled to begin science operation. Prior work has focused on various aspects of position reconstruction with advanced GW detectors (Fairhurst 2009; Wen & Chen 2010; Fairhurst 2011; Vitale & Zanolin 2011; Rodriguez et al. 2013; Nissanke et al. 2011; Nissanke et al. 2013; Kasliwal & Nissanke 2014; Grover et al. 2014; Sidery et al. 2014), but ours is the first to bring together a large astrophysically motivated population, an educated guess about the detector commissioning timetable, a realistic signal-to-noise ratio (SNR) distribution, and the Advanced LIGO/Virgo data analysis pipeline itself.

We have simulated hundreds of GW events, recovered them with a real-time detection pipeline, and generated sky maps using both real-time and thorough off-line parameter estima-

lsinger@caltech.edu

¹ LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

² Department of Physics and Astronomy & Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, Evanston, IL 60208, USA

³ School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, UK

⁴ Leonard E. Parker Center for Gravitation, Cosmology, and Astrophysics, University of Wisconsin–Milwaukee, Milwaukee, WI 53201, USA

⁵ Massachusetts Institute of Technology, 185 Albany St, Cambridge, Massachusetts 02139, USA

⁶ Nikhef, Science Park 105, Amsterdam 1098XG, Netherlands

⁷ Perimeter Institute for Theoretical Physics, Ontario, N2L 2Y5, Canada

⁸ The Pennsylvania State University, University Park, PA 16802, USA

⁹ Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, Ontario, M5S 3H8, Canada

¹⁰ NASA Goddard Space Flight Center, Greenbelt, MD, USA

¹¹ <http://pan-starrs.ifa.hawaii.edu/public/>

¹² <https://www.astro.ru.nl/wiki/research/blackgemarray>

tion codes that will be operating in 2015 and beyond. This study contains some of the first results with BAYESTAR, a rapid Bayesian position reconstruction code that will produce accurate sky maps less than a minute after any BNS merger detection. The LALINFERENCE_MCMC (van der Sluys et al. 2008b; Raymond et al. 2009), LALINFERENCE_NEST (Veitch & Vecchio 2010), and LALINFERENCE_BAMBI (Graff et al. 2012, 2013) stochastic samplers respectively were also used to follow up a subset of detected GW events. Though these analyses are significantly more computationally costly than BAYESTAR, taking hours to days, they can provide improved sky location estimates when the GW signal is very weak in one detector, and also yield not just sky localization but the full multidimensional probability distribution describing the parameters of a circularized compact binary merger. All four algorithms are part of the LALINFERENCE library (Aasi et al. 2013a), developed specifically for estimating the parameters of GW sources from ground-based detectors. Together, these analyses will be able to provide sky localizations on time scales that enable searching for all expected electromagnetic counterparts of compact binary mergers (except the GRB itself).

With the benefit of a much larger sample size, important features of the 2015 and 2016 configurations come into focus. First, we find that even in 2015 when only the two LIGO detectors are operating (or in 2016 during periods when the Virgo detector is not in science mode), there is at least a 60% chance of encountering the source upon searching an area of about 200 deg^2 . Second, many of these two-detector events will not be localized to a single simply connected region in the sky. We elucidate two nearly degenerate sky locations, separated by 180° , that arise when only the two LIGO detectors are operating. When a GW source falls within this degeneracy, its sky map will consist of two diametrically opposed islands of probability. Third, in our simulations, we add a third detector, Advanced Virgo, in 2016. Even though at that time Virgo is anticipated to be only 1/3 as sensitive as the other two detectors due to differing LIGO and Virgo commissioning timetables, we find that coherence with the signal in Virgo generally breaks the previously mentioned degeneracy and shrinks areas to a third of what they were with two detectors. Fourth and most importantly, a picture of a typical early Advanced LIGO event emerges, with most occurring in a limited range of Earth-fixed locations, and most sky maps broadly fitting a small number of specific morphologies.

2. SOURCES AND SENSITIVITY

BNS systems are the most promising and best understood targets for joint GW and EM detection. Though rate estimates remain uncertain, ranging from 0.01 to $10 \text{ Mpc}^{-3} \text{ Myr}^{-1}$, we choose to work with the “realistic” rate obtained from Abadie et al. (2010) of $1 \text{ Mpc}^{-3} \text{ Myr}^{-1}$. This rate leads to a GW detection rate of 40 yr^{-1} at final Advanced LIGO design sensitivity. Some mergers of neutron star–black hole binaries (NSBHs) are also promising sources of GW and EM emission. Two Galactic high-mass X-ray binaries (HMXBs) have been identified as possible NSBH progenitors (Belczynski et al. 2011, 2013). From these can be extrapolated a lower bound on the GW detection rate of at least 0.1 yr^{-1} at Advanced LIGO’s final design sensitivity, although rates comparable to BNS detections are empirically plausible. Black holes in binaries may possess large spins, causing precession during the inspiral. Precession-altered phase evolution can aid in parameter es-

timation (van der Sluys et al. 2008a,b; Harry et al. 2013; Nitz et al. 2013; Raymond et al. 2009), but models of waveforms suitable for rapid detection and parameter estimation are still under active development (Blackman et al. 2014; Hannam et al. 2013; Taracchini et al. 2013). As for the binary black hole mergers detectable by Advanced LIGO and Virgo, there are currently no compelling mechanisms for electromagnetic counterparts associated with them. We therefore restrict our attention to BNS mergers, because they have the best understood rates, GW signal models, and data analysis methods.

2.1. Measures of detector sensitivity

The sensitivity of a single GW detector is customarily described by the horizon distance, or the maximum distance at which a particular source would create a signal with a maximum fiducial single-detector SNR.¹³ It is given by

$$d_H \approx \frac{G^{5/6} M^{1/3} \mu^{1/2}}{c^{3/2} \pi^{2/3} \text{SNR}} \sqrt{\frac{5}{6} \int_{f_1}^{f_2} \frac{f^{-7/3}}{S(f)} df} \quad (1)$$

where G is Newton’s gravitational constant, c is the speed of light, M the sum of the component masses, μ the reduced mass, $f^{-7/3}$ the approximate power spectral density (PSD) of the inspiral signal, and $S(f)$ the PSD of the detector’s noise. The lower integration limit f_1 is the low-frequency extent of the detector’s sensitive band. For the Advanced LIGO and Virgo detectors, ultimately limited at low frequency by ground motion (Adhikari 2014), we take $f_1 = 10 \text{ Hz}$. Using a typical value of the detector sensitivity $S(100 \text{ Hz}) = 10^{-46} \text{ Hz}^{-1}$, we can write Equation 1 as a scaling law,

$$d_H \approx 72.5 \text{ Mpc} \left(\frac{M}{M_\odot} \right)^{1/3} \left(\frac{\mu}{M_\odot} \right)^{1/2} \left(\frac{1}{\text{SNR}} \right) \cdot \left[\int_{\frac{f_1}{\text{Hz}}}^{\frac{f_2}{\text{Hz}}} \left(\frac{f}{100 \text{ Hz}} \right)^{-7/3} \left(\frac{10^{-46} \text{ Hz}^{-1}}{S(f)} \right) d \left(\frac{f}{\text{Hz}} \right) \right]^{1/2}. \quad (2)$$

For BNS masses, the inspiral ends with a merger and black hole ring down well outside LIGO’s most sensitive band. A reasonable approximation is to simply truncate the SNR integration at the last stable orbit of a Schwarzschild black hole with the same total mass (Maggiore 2008),

$$f_2 \approx (4400 \text{ Hz}) \frac{M_\odot}{M}. \quad (3)$$

Usually, $\text{SNR}=8$ is assumed because $\text{SNR}=8$ signals in two detectors (for a root-sum-squared network SNR of $\text{SNR}_{\text{net}} = 8\sqrt{2} = 11.3$) is nearly adequate for a confident detection (see discussion of detection thresholds in Section 3). Another measure of sensitivity is the BNS range d_R , the volume-, direction-, and orientation-averaged distance of a source with $\text{SNR} \geq 8$, drawn from a homogeneous population. Due to the directional sensitivity or antenna pattern of interferometric detectors, the range is a factor of 2.26 smaller than the horizon distance for the same SNR threshold. See also Allen et al. (2012); Abadie et al. (2012).

¹³ Even at its final design sensitivity, Advanced LIGO’s range for BNS mergers is only 200 Mpc or $z = 0.045$ (assuming the WMAP nine-year ΛCDM cosmology; Hinshaw et al. 2013). Because of the small distances considered in this study, we do not distinguish between different distance measures, nor do our gravitational waveforms contain any factors of $(1+z)$.

2.2. Observing scenarios

Aasi et al. (2013b) outline five observing scenarios representing the evolving configuration and capability of the Advanced GW detector array, from the first observing run in 2015, to achieving final design sensitivity in 2019, to adding a fourth detector at design sensitivity by 2022. In this study, we focus on the first two epochs. The first, in 2015, is envisioned as a three-month science run. LIGO Hanford (H) and LIGO Livingston (L) Observatories are operating with an averaged $(1.4, 1.4)M_{\odot}$ BNS range between 40 and 80 Mpc. The second, in 2016–17, is a six-month run with H and L operating between 80 and 120 Mpc and the addition of Advanced Virgo (V) with a range between 20 and 60 Mpc. For each configuration, we used model noise PSD curves in the middle of the ranges in Aasi et al. (2013b), plotted in Figure 1. For H and L, we used the “early” and “mid” noise curves from Barsotti & Fritschel (2012) for the 2015 and 2016 scenarios respectively. For V in 2016, we used the geometric mean of the high and low curves of Aasi et al. (2013b). Final LIGO and Virgo design sensitivity is several steps further in the commissioning schedule than we consider in this paper.

2.3. Simulated waveforms

For each of the two scenarios we made synthetic detector streams by placing post-Newtonian inspiral signals into two months of colored Gaussian noise. We used “SpinTaylorT4” waveforms, employing the TaylorT4 approximant and accurate to 3.5PN order in phase and 1.5PN order in amplitude (Buonanno et al. 2003, 2006, 2009).¹⁴ There was an average waiting time of ≈ 100 s between coalescences. At any given time, one BNS inspiral signal was entering LIGO’s sensitive band while another binary was merging, but both signals were cleanly separated due to their extreme narrowness in time-frequency space. The PSD estimation used enough averaging that it was unaffected by the overlapping signals. Component masses were distributed uniformly between 1.2 and $1.6 M_{\odot}$, bracketing measured masses for components of known BNS systems as well as the $1\text{-}\sigma$ intervals of the intrinsic mass distributions inferred for a variety of NS formation channels (Pejcha et al. 2012; Özel et al. 2012).

We gave each NS a randomly oriented spin with a maximum magnitude of $\chi = c |\mathbf{S}| / Gm^2 \leq 0.05$, where \mathbf{S} is the star’s spin angular momentum and m is its mass. This range includes the most rapidly rotating pulsar that has been found in a binary, PSR J0737-3039A (Burgay et al. 2003; Brown et al. 2012). However, the fastest-spinning millisecond pulsar, PSR J1748-2446ad (Hessels et al. 2006), has a dimensionless spin parameter of ~ 0.4 , and the theoretical evolutionary limits on NS spin-up in BNS systems are uncertain.

2.4. Sensitivity to assumptions

The total detection rate depends on some of these assumptions, and in particular is sensitive to the assumed NS mass distribution. As can be seen from Equation (1), binaries with the greatest and most symmetric component masses can be detected to the farthest distance. According to Equation (3), for BNS systems the merger always occurs at kHz frequencies, on the upward $S(f) \propto f^2$ slope of the noise noise curves in Figure 1 in the regime dominated by photon shot

noise (Buonanno & Chen 2001; Adhikari 2014). As a result, the integral in Equation (1) depends only weakly on masses. For equal component masses, the horizon distance scales as $d_H \propto m_{\text{NS}}^{5/6}$, so the detection rate scales rapidly with mass as $\dot{N} \propto d_H^3 \propto m_{\text{NS}}^{2.5}$.

The normalized distribution of sky localization areas depends only weakly on the distribution of NS masses. Fairhurst (2009) computes the approximate scaling of sky localization area by considering the Fisher information associated with time of arrival measurement. Valid for moderately high SNR, the RMS uncertainty in the time of arrival in a given detector is

$$\sigma_t = \frac{1}{2\pi \text{SNR} \sqrt{\bar{f}^2 - \bar{f}^k}} \quad (4)$$

where $\bar{f} = \bar{f}^1$, and \bar{f}^k is the k th moment of frequency, weighted by the signal to noise per unit frequency,

$$\bar{f}^k \approx \left[\int_{f_1}^{f_2} \frac{|h(f)|^2 f^k}{S(f)} df \right] \left[\int_{f_1}^{f_2} \frac{|h(f)|^2}{S(f)} df \right]^{-1}. \quad (5)$$

As in Equation (1), we can substitute the approximate inspiral signal spectrum $|h(f)|^2 \propto f^{-7/3}$. The areas then scale as the product of the timing uncertainty in individual detectors, or as simply the square of Equation (4) for a network of detectors with similar (up to proportionality) noise PSDs. As m_{NS} varies from 1 to $2 M_{\odot}$, the upper limit of integration f_2 given by Equation (3) changes somewhat, but areas change by a factor of less than 1.5. (See also Grover et al. 2014 for scaling of sky localization area with mass).

Introducing faster NS spins would result in smaller sky localization areas, since orbital precession can aid in breaking GW parameter estimation degeneracies (Raymond et al. 2009). However, rapid spins could require more exotic BNS formation channels, and certainly would require using more sophisticated and more computationally expensive GW waveforms for parameter estimation.

2.5. Source locations

Source locations were random and isotropic, and uniform in distance cubed. The source distribution was cut off at the $\text{SNR}=5$, $(1.6, 1.6)M_{\odot}$ horizon distance, far enough away that the selection of detected binaries was determined primarily by the sensitivity of the instruments and the detection pipeline, not by the artificial distance boundary.

2.6. Duty cycle

Following Aasi et al. (2013b), we assumed that each detector had an independent and random $d = 80\%$ duty cycle. In the 2015, HL configuration, this implies that both detectors are in operation $d^2 = 64\%$ of the time. In 2016, there are three detectors operating $d^3 = 51.2\%$ of the time and each of three pairs operating $d^2(1-d) = 12.8\%$ of the time. We do not simulate any of the time when there are one or zero detectors operating, but instead fold this into conversion factors from our Monte Carlo counts to detection rates.

3. DETECTION AND POSITION RECONSTRUCTION

Searches for GWs from compact binaries (Allen et al. 2012; Babak et al. 2013) employ banks of matched filters, in the data from all of the detectors are convolved with an array of

¹⁴ There is a C language implementation as the function `XLALSimInspiralSpinTaylorT4` in `LALSIMULATION`. See acknowledgements and Appendix.

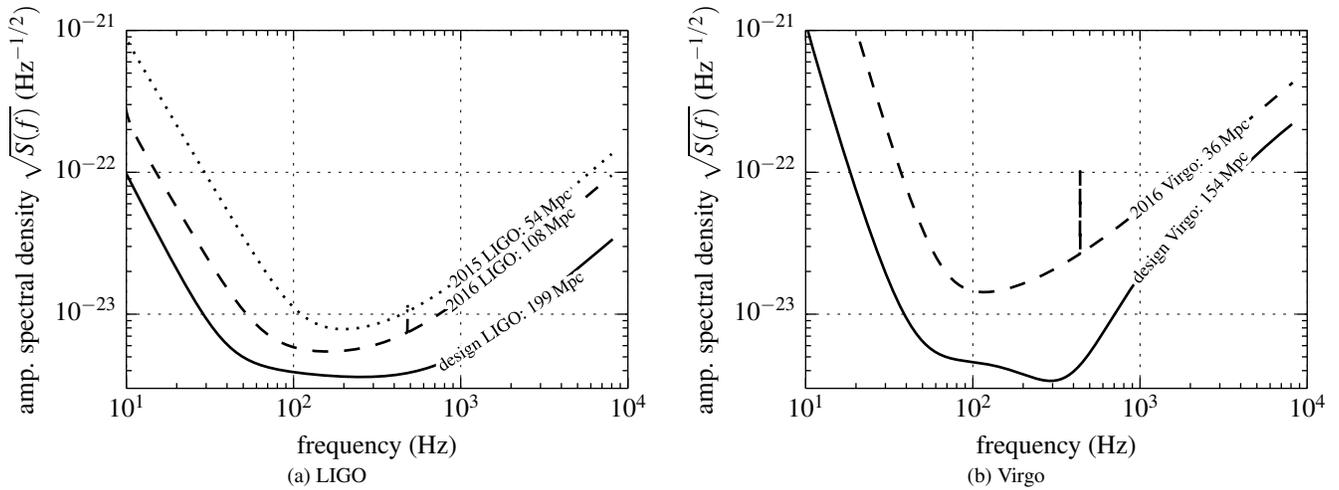


Figure 1. Model detector noise amplitude spectral density curves. The LIGO 2015, 2016, and final design noise curves are shown in the left panel and the Virgo 2016 and final design noise curves in the right panel. The averaged SNR=8 range d_R for $(1.4, 1.4) M_\odot$ BNS mergers is given for each detector.

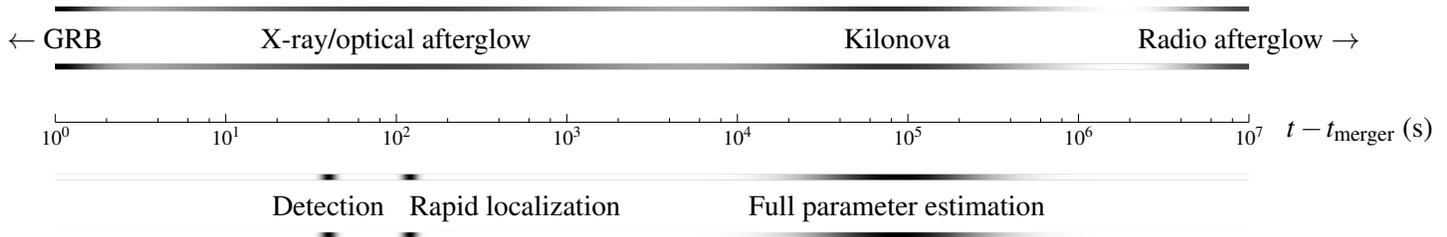


Figure 2. Rough timeline of compact binary merger electromagnetic emissions in relation to the timescale of the Advanced LIGO/Virgo analysis described in this paper. The time axis measures seconds after the merger.

template waveforms. The output of each filter is the instantaneous SNR with respect to that template in that detector. An excursion above a threshold SNR in two or more detectors with exactly the same binary parameters and within approximately one light-travel time between detectors is considered a coincidence. Coincidences may be accidental, due to chance noise fluctuations or, in real GW data streams, environmental disturbances and instrument glitches. Coincidences with sufficiently high SNR_{net} (root-sum-square of the SNRs in the individual detectors) are considered detection candidates. A χ^2 statistic is used to aid in separating the true, astrophysical signals from accidental coincidences or false positives (Allen 2005; Hanna 2008; Cannon et al. 2013).

Offline inspiral searches used in past LIGO/Virgo science runs will be computationally strained in Advanced LIGO/Virgo due to denser template banks and BNS signals that remain in band for up to $\sim 10^3$ s. To address these issues and achieve latencies of $\lesssim 1$ min, a rapid matched-filter detection pipeline called GSTLAL_INSPIRAL (Cannon et al. 2012) has been developed. To mimic Advanced LIGO/Virgo observations as closely as possible, we used GSTLAL_INSPIRAL to extract simulated detection candidates from our two-month data streams.

3.1. Template waveforms

The templates were constructed from a frequency domain, post-Newtonian model describing the inspiral of two compact objects, accurate to 3.5 post-Newtonian order in phase

and Newtonian order in amplitude (Buonanno et al. 2009).¹⁵ These waveforms neglect spins entirely. This is known to have a minimal impact on detection efficiency for BNS sources with low spins (Brown et al. 2012). These waveforms are adequate for recovering the weakly spinning simulated signals that we placed into the data stream.

3.2. Detection threshold

In our study, we imposed a single-detector threshold SNR of 4. A simulated signal is then considered to be detected by GSTLAL_INSPIRAL if it gives rise to a coincidence with sufficiently low false alarm probability as estimated from the SNR and χ^2 values. We follow the lead of Aasi et al. (2013b) in adopting a false alarm rate (FAR) threshold of $\text{FAR} \leq 10^{-2} \text{ yr}^{-1}$. Aasi et al. (2013b) claim that in data of similar quality to previous LIGO/Virgo science runs, this FAR threshold corresponds to a network SNR threshold of $\text{SNR}_{\text{net}} \geq 12$. Since our data is Gaussian and perfectly free of glitches, to obtain easily reproducible results we imposed a second explicit detection cut of $\text{SNR}_{\text{net}} \geq 12$. We find that our joint threshold on FAR and SNR differs negligibly from a threshold on SNR alone. Because any given simulated signal will cause multiple coincidences at slightly different masses and arrival times, for each simulated signal we keep only the matching candidate with the lowest SNR.

¹⁵ These are in LALSIMULATION as the function XLLSimInspiralTaylorF2. See acknowledgements and Appendix.

3.3. Sky localization and parameter estimation

All detection candidates are followed up with rapid sky localization by BAYESTAR and a subset were followed up with full parameter estimation by the LALINFERENCE_MCMC/NEST/BAMBI stochastic samplers. The three different stochastic samplers all use the same likelihood, but serve as useful cross-verification. Both BAYESTAR and the three stochastic samplers are coherent (exploiting the phase consistency across all detectors) and Bayesian (making use of both the GW observations and prior distributions over the source parameters). They differ primarily in their input data.

BAYESTAR’s likelihood function depends on only the information associated with the triggers comprising a coincidence: the times, phases, and amplitudes on arrival at each of the detectors. BAYESTAR exploits the leading-order independence of errors in the extrinsic and intrinsic parameters by holding the masses fixed at the values estimated by the detection pipeline. Marginalized posterior distributions for the sky positions of sources are produced by numerically integrating the posterior in each pixel of the sky map. Because BAYESTAR’s analysis explores only a small sector of the full parameter space, never performs costly evaluations of the post-Newtonian GW waveforms, and uses highly tuned standard numerical quadrature techniques, it takes well under a minute (see Figure 2).

On the other hand, the likelihood function used for the stochastic samplers depends on the full GW data, and is the combination of independent Gaussian distributions for the residual in each frequency bin after model subtraction. Posterior distributions for the sky position are produced by sampling the full parameter space of the signal model, then marginalizing over all parameters but the sky location. This method requires the generation of a model waveform for each sample in parameter space, making it far more expensive than the BAYESTAR approach, but independent of the methods and models used for detection. Most importantly, intrinsic parameters (including spins) can be estimated using these higher-latency methods. For the purposes of this study, parameter estimation used the same frequency-domain, non-spinning waveform approximant as the detection pipeline. Analyses that account for the spin of the compact objects are more costly, currently taking weeks instead of days to complete, and will be the subject of a future study.

4. RESULTS

Of $\sim 100,000$ simulated sources spread across the 2015 and 2016 scenarios, ≈ 1000 events survived as confident GW detections.¹⁶ No false alarms due to chance noise excursions survived our detection threshold; all events which should have been detectable were detected. We constructed probability sky maps using BAYESTAR for all events and using LALINFERENCE_NEST/MCMC for a randomly selected subsample of 250 events from each scenario. Results from LALINFERENCE_BAMBI are not shown because this sampler was run for only 30 events, and the sampling error bars would overwhelm

¹⁶ There were slightly fewer surviving events in the 2016 configuration than in the 2015 configuration. This is because adding a third detector required us to apportion the two months of Gaussian noise to different combinations of detectors. In the 2015 simulation, all two months of data were allocated to the HL network. In 2016 about 43 days were devoted to the HLV and HL configurations, with the remaining 17 days of HV and LV mode contributing few detections.

the plots.¹⁷ The top four panels (a, b, c, d) of Figure 3 show cumulative histograms of the areas in deg^2 inside of the smallest 50% and 90% confidence regions for each event, for both sky localization methods. These contours were constructed using a ‘water-filling’ algorithm: we sampled the sky maps using equal-area HEALPix (Hierarchical Equal Area isoLatitude Pixelization; Górski et al. 2005) pixels, ranked the pixels from most probable to least, and finally counted how many pixels summed to a given total probability. In the bottom two panels (e, f) of Figure 3, we also show a histogram of the smallest such constructed region that happened to contain the true location of each simulated source. We call this the searched area.

Panels (a, b, c, d) and (e, f) may be thought of as measuring precision and accuracy respectively. The former measure how dispersed or concentrated each individual sky map is, while the latter describe how far the localization is from the true sky position. The 90% area histograms and the searched area histograms also answer different but complementary questions that relate to two different strategies for following up LIGO/Virgo events. One might decide in 2015 to search for optical counterparts of all GW events whose 90% areas are smaller than, for example, 200 deg^2 . By finding 200 deg^2 on the horizontal axis of 90% area histogram, one would find that this corresponds to following up 10% of all GW detections. On the other hand, one might decide to always search the most probable 200 deg^2 area for every GW event, corresponding to a different confidence level for every event. In this case, one would find 200 deg^2 on the horizontal axis of the searched area histogram, and find that this strategy would enclose the true location of the GW source 64% of the time.¹⁸

The left-hand axes of all four panels of Figure 3 show the expected cumulative number of detections, assuming the ‘realistic’ BNS merger rates from Abadie et al. (2010). We stress that the absolute detection rate might be two orders of magnitude smaller or one order of magnitude higher due to the large systematic uncertainty in the volumetric rate of BNS mergers, estimated from population synthesis and the small sample of Galactic binary pulsars (Abadie et al. 2010). An additional source of uncertainty in the detection rates is the Advanced LIGO/Virgo commissioning schedule given in Aasi et al. (2013b). The proposed sensitivity in the 2016 scenario may be considered a plausible upper bound on the performance of the GW detector network in 2015, if commissioning occurs faster than anticipated. Likewise, the quoted sensitivity in the 2015 scenario is a plausible lower bound on the performance in 2016. The right-hand axes show the cumulative percentage of all detected sources. These percentages depend only on the gross features of the detector configuration and not on the astrophysical rates, so are relatively immune to the systematics described above.

Table 1 summarizes these results.

4.1. 2015

¹⁷ The three stochastic samplers LALINFERENCE_NEST/MCMC/BAMBI were interchangeable to the extent that they used the same likelihood and produced sky maps that agreed with each other.

¹⁸ One might naively expect that self-consistency would require the 90% confidence area and searched area histograms to intersect at 90% of detections, but this is not generally required because the posteriors of different events have widely different dimensions. However, it is true that 90% of sources should be found within their respective 90% confidence contours. This can be formalized into a graphical self-consistency test; see Sidery et al. (2014) for an example of application to GW parameter estimation.

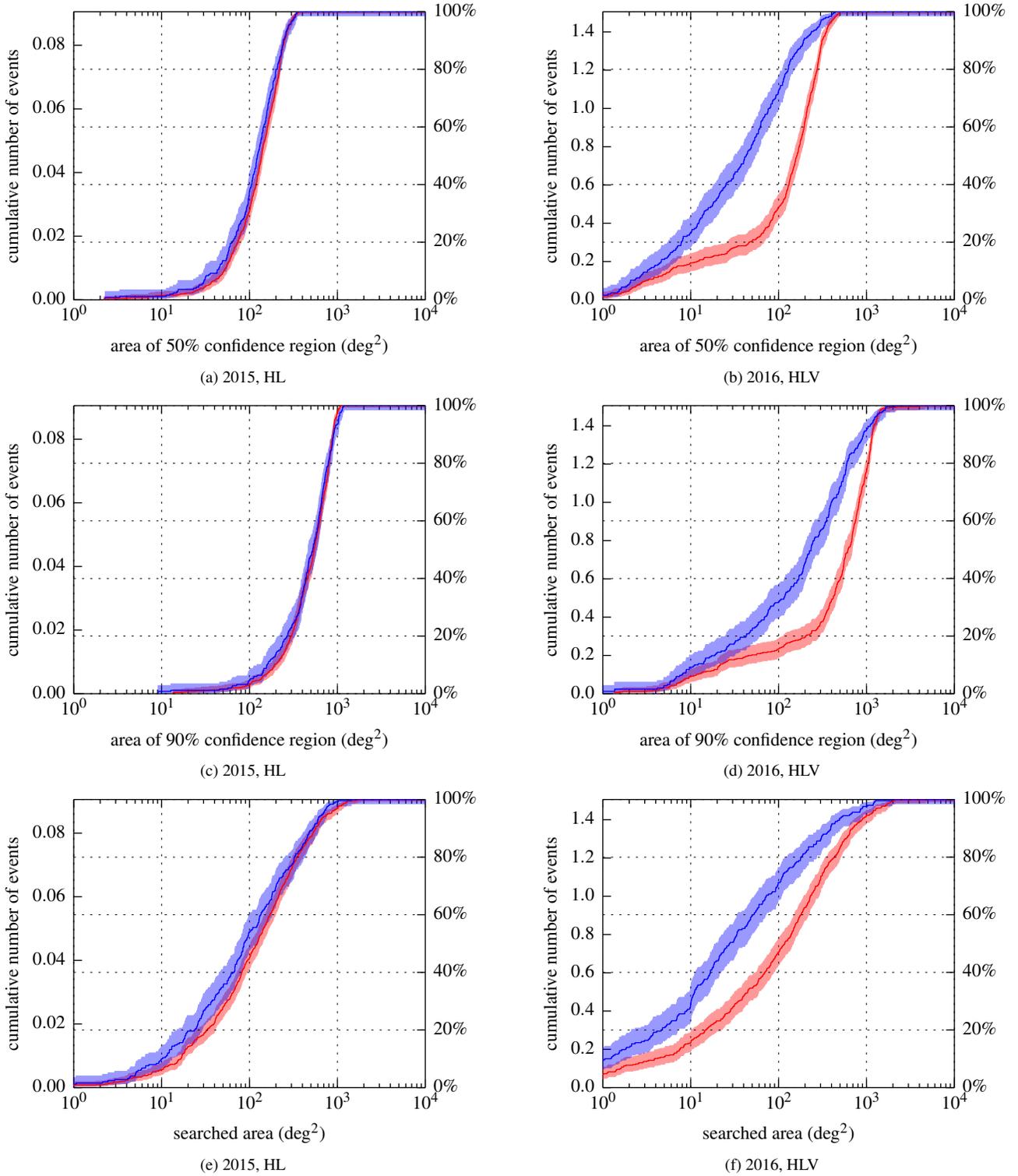


Figure 3. Cumulative histogram of sky localization areas in the 2015 (HL) and 2016 (HLV) scenarios. Plots in the left column (panels a, c, and e) refer to the 2015 configuration and in the right column (panels b, d, and f) to the 2016 configuration. The first row (panels a and b) shows the area of the 50% confidence region, the second row (panels c and d) shows the 90% confidence region, and the third row (panels e and f) shows the “searched area,” the area of the smallest confidence region containing the true location of the source. The red lines comprise all detections and their sky maps produced with BAYESTAR, and the blue lines represent sky maps for the random subsample of 250 detections analyzed with LALINFERENCE_NEST/MCMC. The light shaded region encloses a 95% confidence interval accounting for sampling errors (computed from the quantiles of a beta distribution; Cameron 2011). The left axes show the number of detections localized within a given area assuming the “realistic” BNS merger rates from (Abadie et al. 2010). The right axes show the percentage out of all detected events. (A color version of this figure is available in the online journal.)

Table 1

Summary of the 2015 and 2016 scenarios, listing the participating detectors, BNS horizon distance, run duration, and fractions of events localized within 5, 20, 100, 200, or 500 deg². A dash (—) represents less than 1% of detections.

		2015		2016	
Detectors		HL	HLV		
LIGO (HL) BNS range		54 Mpc	108 Mpc		
Run duration		3 months	6 months		
No. detections		0.091	1.5		
		rapid	full PE	rapid	full PE
fraction 50% CR smaller than	5 deg ²	—	—	9%	14%
	20 deg ²	2%	3%	15%	35%
	100 deg ²	30%	37%	32%	72%
	200 deg ²	74%	80%	62%	90%
	500 deg ²	100%	100%	100%	100%
fraction 90% CR smaller than	5 deg ²	—	—	2%	2%
	20 deg ²	—	—	8%	14%
	100 deg ²	3%	4%	15%	31%
	200 deg ²	10%	13%	19%	45%
	500 deg ²	44%	48%	39%	71%
fraction searched area smaller than	5 deg ²	3%	4%	11%	20%
	20 deg ²	14%	19%	23%	44%
	100 deg ²	45%	54%	47%	71%
	200 deg ²	64%	70%	62%	81%
	500 deg ²	87%	89%	83%	93%
median area	50% CR	138 deg ²	124 deg ²	162 deg ²	43 deg ²
	90% CR	545 deg ²	529 deg ²	621 deg ²	235 deg ²
	searched	123 deg ²	88 deg ²	118 deg ²	29 deg ²

Our 2015 scenario assumes two detectors (HL) operating at an anticipated range of 54 Mpc. About 0.1 detectable BNS mergers are expected, though there are nearly two orders of magnitude systematic uncertainty in this number due to the uncertain astrophysical rates. A detection in 2015 is possible, but likely only if the BNS merger rates or the detectors’ sensitivity are on the modestly optimistic side. A typical or median event (with a localization area in the 50th percentile of all detectable events) would be localized to a 90% confidence area of ~ 500 deg².

We find that the area histograms arising from the BAYESTAR rapid sky localization and the full parameter estimation agree within sampling errors, and that the sky maps resulting from the two analyses are comparable for any individual event. Put differently, the rapid sky localization contains nearly all of the information about sky localization for these events, with the full probability distributions over masses and spins becoming available when the stochastic samplers finish on a timescale of a day.

Figure 4(a) shows a histogram of the cosine of the angular separation between the true location of the simulated GW source and the maximum *a posteriori* estimate (the mode of the sky map, or the most probable location). The main feature is a peak at low separation. However, there is a second peak at the polar opposite of the true location, 180° away; about 15% of events are recovered between 100 and 180° away from the true location.

Correspondingly, for any one event, it is common to find the probability distributed across two antipodal islands on opposite sides of the mean detector plane. We define this plane by finding the average of the two vectors that are normal to the planes of the two detectors’ arms, and then taking the plane to which this vector is normal. This plane partitions the sky into two hemispheres. We find that one hemisphere is favored over the other by less than 20% (which is to say that the odds favor-

ing one hemisphere over the other are as even as 60%/40%) for 20% of events.

The second peak admits a simple explanation as an unavoidable degeneracy due to the relative positions of the H and L interferometers. Before the Hanford and Livingston sites were selected, it was decided that the detectors’ arms would be as closely aligned as possible (Vogt 1989, section V-C-2). Significant misalignment would have created patches of the sky that were accessible to one detector but in a null of the other detector’s antenna pattern, useless for a coincidence test.

The near alignment maximized the range of the detectors in coincidence, though at certain expenses for parameter estimation. Observe that the sensitivity of an interferometric GW detector is identical at antipodal points (i.e., symmetric under all rotations by 180°). Therefore, any source that lies in the plane of zero time delay between the detectors is always localized to two opposite patches. Because the HL detectors were placed nearby (at continental rather than intercontinental distances) on the surface of the Earth so as to keep their arms nearly coplanar, their combined network antenna pattern has two maxima that lie on opposite sides of that great circle. As a consequence, a large fraction of sources are localized to two islands of probability that cannot be distinguished based on time or amplitude on arrival. See Figure 5 for an illustration of these two degenerate patches.

A second undesirable side effect of the aligned antenna patterns is that GW polarization, observed as the phase difference on arrival at these two detectors, is of limited help for parameter estimation.

A fairly typical sky map morphology, even at modestly high SNR_{net}, will consist of two extended arc-shaped modes, one over North America and a mirror image on the opposite side of the Earth. See Figure 6 for a typical event exhibiting this degeneracy. In this example, it is also possible to discern two narrow stripes resembling the forked tongue of a snake. This is a reflection of the HL network’s limited polarization sensitivity. It occurs when the GW phases on arrival support two different binary inclination angles, with the orbital plane nearly facing the observer but with opposite handedness (usually peaked at $\iota \approx 30^\circ$ and $\iota \approx 150^\circ$; see Schutz 2011). The two forks cross at a sky location where the GW data cannot distinguish between a clockwise or counterclockwise orbit.

The HL degeneracy is even apparent in earlier works on localization of GW bursts with networks of four or more detectors: Klimenko et al. (2011) drew a connection between accurate position reconstruction and sensitivity to both the ‘+’ and ‘x’ GW polarizations, and noted that the close alignment of the HL detector network adversely affects position reconstruction. (They did not, however, point out the common occurrence of nearly 180° errors, or note that the worst GW localizations paradoxically occur where the HL network’s sensitivity is the greatest.)

The HL degeneracy affects most events that occur $\lesssim 30^\circ$ from one of the antenna pattern maxima. Most events that are $\gtrsim 50^\circ$ away have localizations that consist of a single long, thin arc or ring. See Figure 7 for an example.

In Figure 8, we have plotted a histogram of the number of disconnected modes comprising the 50% and 90% confidence regions and the searched area, for the rapid localizations in the 2015 configuration. The ratios of events having one, two, or three or more modes depends weakly on the selected confidence level. In 2015, using either the 50% contour or the searched area, we find that about half of events are unimodal

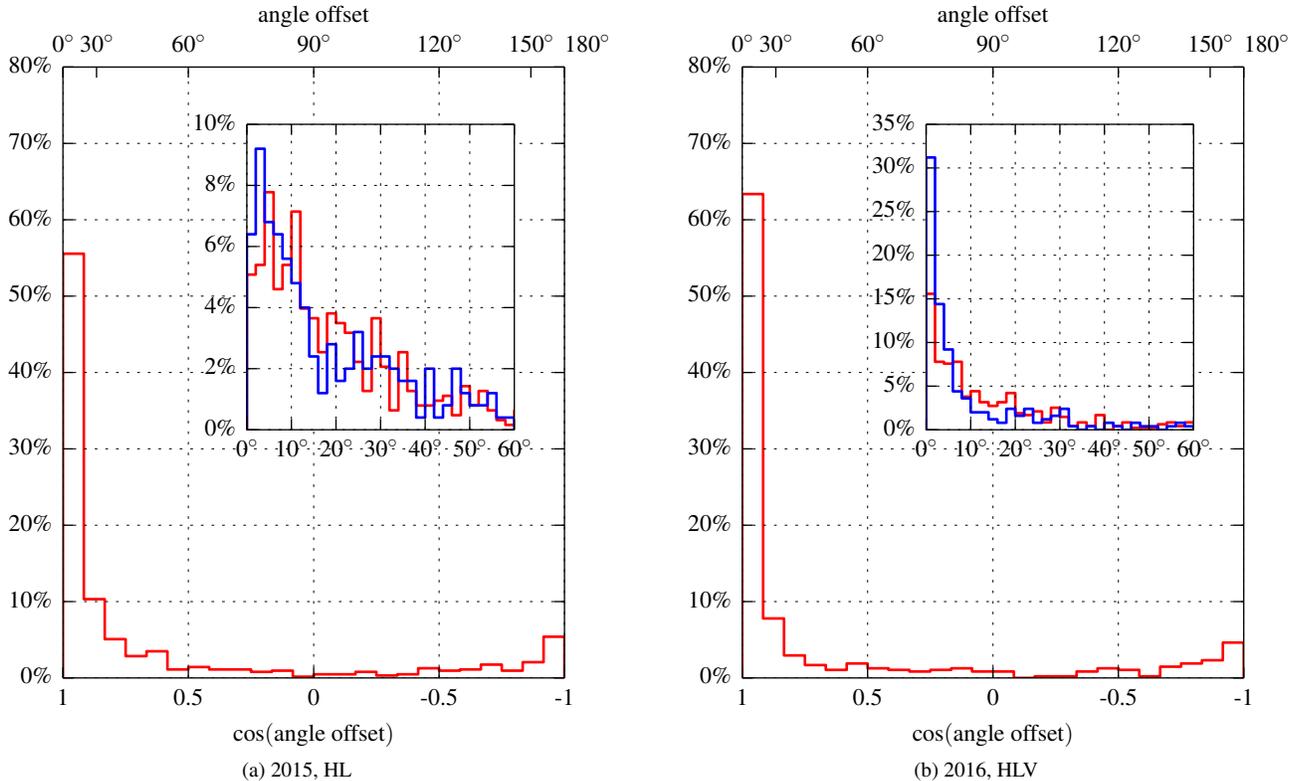


Figure 4. Normalized histogram of the cosine angular separation between the location of the simulated GW source and the maximum *a posteriori* location estimate, for (a) the 2015 configuration and (b) the 2016 configuration. The red line encompasses all detections and their BAYESTAR localizations, and the blue line the subsample of 250 events analyzed by LALINFERENCE_NEST/MCMC. The inset shows the distribution of angle offsets for angles less than 60° . (A color version of this figure is available in the online journal.)

and about a third are bimodal, the rest comprising three or more modes. Using the 90% contour, we find that about a third of the events are unimodal and about half are bimodal.

4.2. 2016

In our 2016 scenario, the HL detectors double in range to 108 Mpc and the V detector begins observations with a range of 36 Mpc. Over this six-month science run we expect ~ 1.5 detections, assuming a BNS merger rate of $1 \text{ Mpc}^{-3} \text{ Myr}^{-1}$. Figure 9 shows how livetime and duty cycle breaks down according to detector network (HLV, HL, LV, or HV). About half of the time all three detectors are online, with the remaining time divided in four almost equal parts among the three pairs of detectors or ≤ 1 detector. However, the HLV network accounts for about three quarters of detections and the HL network for most of the rest.

When all three detectors (HLV) are operating, most detections are comprised of H and L triggers, lacking a trigger from V because the signal is below the single-detector threshold of $\text{SNR}=4$. Slightly more than half (57%) of all detections have a signal below threshold in one operating detector (almost always V) while slightly less than half (43%) consist of triggers from all operating detectors.

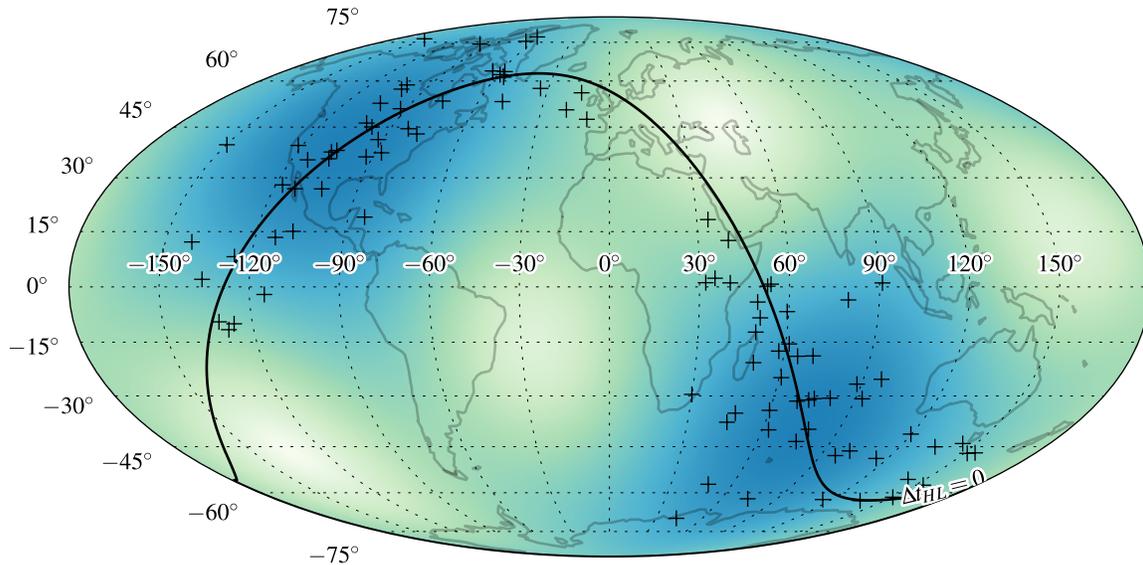
For the first half, mostly HLV events that are detected by HL but not Virgo, the stochastic samplers provide a marked improvement in sky localization and 90% confidence regions that have about $1/3$ as much area as the rapid localization. This is because the rapid localization makes use of only the triggers provided by the detection pipeline, lacking information about the signal in Virgo if its SNR is < 4 . The stochastic samplers, on the other hand, can use data from all oper-

ating detectors, regardless of SNR. Therefore, in the present analysis, an improved sky localization would be available for half of the detections on a timescale of a day. Fortunately, for BNS sources, it is immediately known whether an improved localization is possible, since this statement only depends on what detectors were online and which contributed triggers. On the other hand, it may be possible to provide prompt sky localizations for all events by simply lowering the single-detector threshold. If the single-detector threshold was dropped to unity, essentially no event would lack a Virgo trigger. There are also efforts to do away with the single-detector threshold entirely (Keppel 2012, 2013). Simultaneously, there is promising work under way in speeding up the full parameter estimation using reduced order quadratures (Canizares et al. 2013), interpolation (Smith et al. 2013), jump proposals that harness knowledge of the multimodal structure of the posterior (Farr et al. 2013), hierarchical techniques (Farr & Kalogera 2013), and machine learning techniques to accelerate likelihood evaluation (Graff et al. 2012, 2013). It seems possible that the delayed improvement in sky localization may be a temporary limitation that can be overcome well before 2016.

The second half consists of HLV events with triggers from all three detectors and events that occur when only HL, HV, or LV are operating. For these, the BAYESTAR analysis and the full stochastic samplers produce comparable sky maps.

For nearby, loud sources ($\text{SNR}_{\text{net}} \gtrsim 20$), the HLV network frequently produces compact sky localizations concentrated in a single island of probability. However at low SNR ($\text{SNR}_{\text{net}} \lesssim 20$), and especially for the events that are detected as only double coincidence (HL), the refined localizations

Figure 5. The HL degeneracy. This, like all sky plots in this paper, is a Mollweide projection in geographic coordinates to emphasize spatial relationships with respect to the Earth-fixed GW detector network as well as possible ground-based telescope sites. Pluses denote the locations of signals whose best-estimate locations are offset by $\geq 100^\circ$, comprising the large-offset peak that is evident in Figure 4(a). The locations of zero time delay (simultaneous arrival at the H and L detectors) is shown as a thick black line. Shading indicates the RMS network antenna pattern, with darker areas corresponding to high sensitivity and white corresponding to null sensitivity. (A color version of this figure is available in the online journal.)



from the full stochastic samplers often identify many smaller modes. An $\text{SNR}_{\text{net}} = 13.4$ example is shown in Figure 10. In this event, the rapid sky localization has two modes and has a morphology that is well described by the HL degeneracy explained in Section 4.1. However, the refined, full parameter estimation breaks this into at least four smaller modes.

Of the remaining events, most occur when only the two HL detectors are operating. These look qualitatively the same as the 2015 case; their sky maps generally exhibit one or two slender islands of probability. However, percentage-wise, two-detector events are localized worse in the 2016 scenario than in the 2015 scenario. This unusual result is easily explained. Though the LIGO detectors improve in sensitivity at every frequency, with the particular noise curves that we assumed the signal bandwidth is actually slightly lower with the 2016 sensitivity compared to 2015. This is because of improved sensitivity at low frequency. Applying Equation (4), we find that for a $(1.4, 1.4)M_{\odot}$ binary at $\text{SNR} = 10$, one of the 2015 LIGO detectors has an RMS timing uncertainty of $131 \mu\text{s}$, whereas one of the 2016 LIGO detectors has a timing uncertainty of $158 \mu\text{s}$. Clearly, the 2016 detectors will produce more constraining parameter estimates for sources at any fixed distance as the SNR improves. However, for constant SNR the 2016 LIGO detectors should find areas that are $(158/131)^2 = 1.45$ times larger than events at the same SNR in 2015. This is indeed what we find.

Two-detector events involving Virgo (HV and LV) are rare, accounting for only about 6% of detections. Sky maps for these events sometimes exhibit multiple fringes spread over a quadrant of the sky. These are in part due to the increased

importance of phase-on-arrival due to the oblique alignment of the LIGO and Virgo antenna patterns, which gives the network a limited ability to measure GW polarization. Occasionally there are also diffuse clouds of probability near the participating LIGO detector's two antenna pattern maxima, which may be a vestige of the antenna pattern. A typical HV event that exhibits both features is shown in Figure 11.

5. DISCUSSION

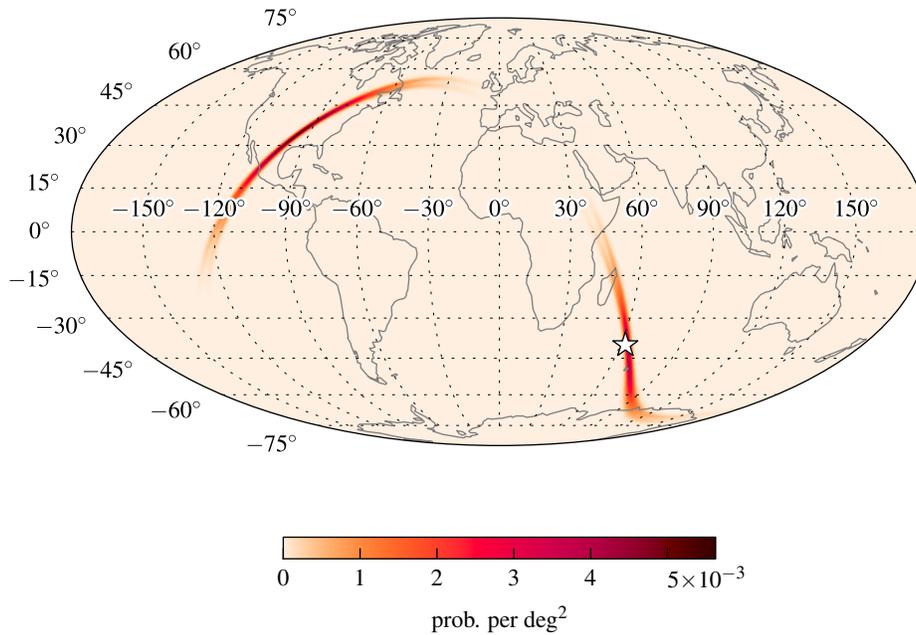
5.1. Caveats

We reiterate that the scenarios we have described make assumptions about the astrophysical rate of BNS mergers and the Advanced LIGO/Virgo sensitivity as a function of time. The former is subject to orders of magnitude uncertainty due to the small sample of known galactic binary pulsars as well as model dependence in population synthesis (Abadie et al. 2010). The latter could deviate from Aasi et al. (2013b) depending on actual Advanced LIGO/Virgo commissioning progress. However, the fractions of events localized within a given area are robust with respect to both of these effects.

We have dealt only with BNS mergers. NSBH mergers are also promising sources for closely related GW signals and EM transients. A similar, comprehensive investigation of GW sky localization accuracy for NSBH signals is warranted.

In this simulation, we have used ideal Gaussian noise, but selected a detection threshold that is designed to reproduce expected performance in detectors with realistically wide tails due to instrumental and environmental glitches. If Advanced LIGO and Virgo's improved seismic isolation and control systems are even more effective at suppressing such glitches than

Figure 6. Localization of a typical circa 2015 GW detection. This is a Mollweide projection in geographic coordinates. Shading is proportional to posterior probability per deg^2 . This is a moderately loud event with $\text{SNR}_{\text{net}} = 15.0$, but its 90% confidence area of 630 deg^2 is fairly typical, in the 60th percentile of all detections. The sky map is bimodal with two long, thin islands of probability over the north and southern antenna pattern maxima. Neither mode is strongly favored over the other. Each island is forked like a snake’s tongue, with one fork corresponding to the binary having face-on inclination ($\iota \approx 0^\circ$) and the other fork corresponding to face-off ($\iota \approx 180^\circ$). This is event ID 18951 from Tables 2 and 3 and the online supplement (see Appendix for more details). (A color version of this figure is available in the online journal.)



their initial counterparts were, then the SNR_{net} threshold for confident detection would decrease, yielding discoveries earlier but with larger typical sky localization areas.

We remind the reader that the events comprising this study would be regarded as confident detections, with $\text{FAR} \lesssim 10^{-2} \text{ yr}^{-1}$, based on GW observations alone. In practice, some observers may choose to follow up more marginal detection candidates. For instance, a group with enough resources and telescope time to follow up one candidate per month might filter events with $\text{FAR} \leq 12 \text{ yr}^{-1}$. A high false alarm rate threshold will admit correspondingly lower SNR_{net} candidates with coarser localizations than what we have presented here.

Last, on a positive note, the number of detections is expected to increase considerably as commissioning proceeds toward final design sensitivity. Furthermore, sky localization will improve radically as the HLV detectors approach comparable sensitivity. The addition of two more planned ground-based GW detectors, LIGO–India and KAGRA, would likewise increase rates and improve sky localizations dramatically (Schutz 2011; Veitch et al. 2012; Fairhurst 2012; Nissanke et al. 2013; Aasi et al. 2013b).

5.2. Detection scenarios

From our representative sample of hundreds of early Advanced LIGO/Virgo events emerge a few common morphologies and several possible scenarios for the early detections of GWs from a BNS merger.

We find that in both 2015 and 2016, the detection rate is highly anisotropic, proportional to the cube of the network

antenna pattern with a strong excess above North America and the Indian Ocean and deficits in four spots over the south Pacific, south Atlantic, Caucasus, and north Pacific.

1. *HL event, single arc* — This scenario is relevant for the HL network configuration and applies to both 2015 and 2016. Figure 7 shows a typical sky map for a near-threshold detection with a $\text{SNR}_{\text{net}} = 12.7$, exhibiting a single long, extended arc spanning $\sim 500 \text{ deg}^2$.
2. *HL event, two degenerate arcs* — This scenario also applies to 2015 or to HL livetime in 2016. Figure 6 shows a typical sky map with a moderately high $\text{SNR}_{\text{net}} = 15.0$, localized to $\sim 600 \text{ deg}^2$. Its localization embodies the HL degeneracy, with two strong, long, thin modes over North America and the Indian Ocean, separated by nearly 180° and therefore 12 hours apart in hour angle. Inevitably, one of these two modes will be nearly sun-facing and inaccessible to optical facilities. Because of the bimodality, these sky maps can span slightly larger areas than case 1. After taking an inevitable 50% hit in visibility, such events resemble the single arc scenario.

Whether a given source falls into scenario 1 or 2 is largely determined by its sky location relative to the network antenna pattern. The transition occurs between $\sim 30^\circ$ and $\sim 50^\circ$ away from the two points of maximum sensitivity.

3. *HLV event, degeneracy broken by Virgo* — This sce-

Figure 7. Localization of a typical circa 2015 GW detection. This is a Mollweide projection in geographic coordinates. Shading is proportional to posterior probability per deg^2 . This event’s $\text{SNR}_{\text{net}} = 12.7$ is near threshold, but its 90% confidence area of 530 deg^2 near the median. The sky map consists of a single, long, thin island exhibiting the forked-tongue morphology. This is event ID 20342 from Tables 2 and 3 and the online supplement (see Appendix for more details). (A color version of this figure is available in the online journal.)

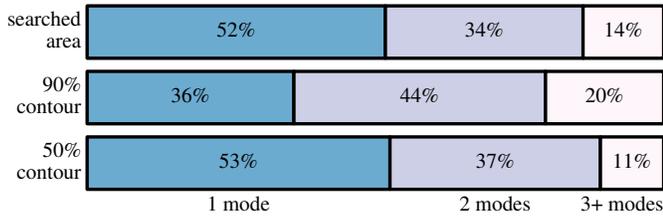
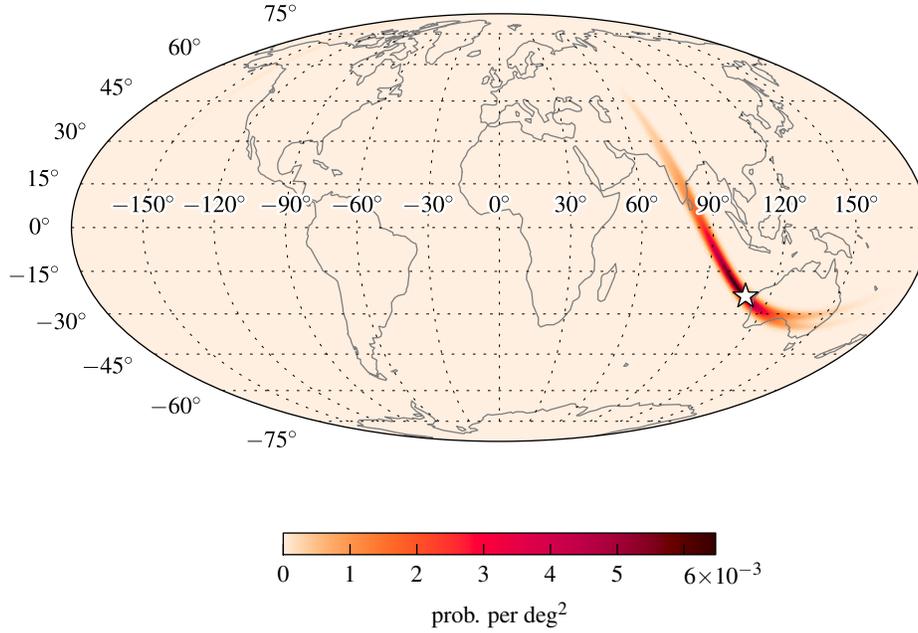


Figure 8. Frequency with which GW sky maps have one, two, or more disconnected modes during 2015. From top to bottom are the number of modes contained within the smallest confidence contour containing each simulated signal, the smallest 90% contour, and the smallest 50% contour. In 2015, roughly half of the sky maps will be unimodal, with most of the remainder being bimodal. (A color version of this figure is available in the online journal.)

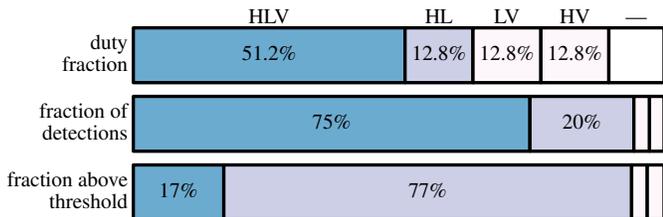


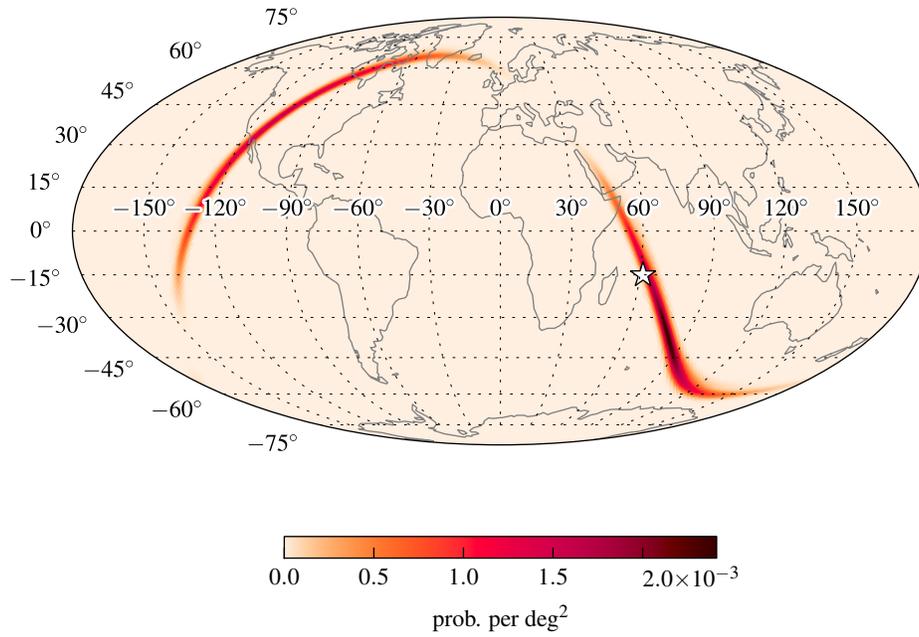
Figure 9. Breakdown of 2016 scenario by detector network. The top row shows the duty fraction of each subset of the detector network, the fraction of time when all three detectors (HLV) are observing, when any pair of detectors are observing (HL, LV, or HV), or when zero or one detectors are observing (—). The second row shows the fraction of coincident detections that occur under any given network (HLV, HL, LV, or HV). The last row shows the fraction of coincident detections for which the given detectors have signals above the single-detector threshold of $\text{SNR}=4$. (A color version of this figure is available in the online journal.)

nario applies only to the 2016 configuration, while all three instruments are online. The rapid sky localization looks similar to the previous scenario, a pair of long, thin rings over the northern and southern hemispheres, but the full parameter estimation cuts this down to a handful of islands of probability covering as little as a half to a third of the area, $\sim 200 \text{ deg}^2$. For such an event, the refined localization could be used to guide \sim day cadence kilonova-hunting observations or to re-target the vetting of afterglow candidates arising from early-time observations. Several wide-field facilities could be employed to monitor modes that lie in different hemispheres.

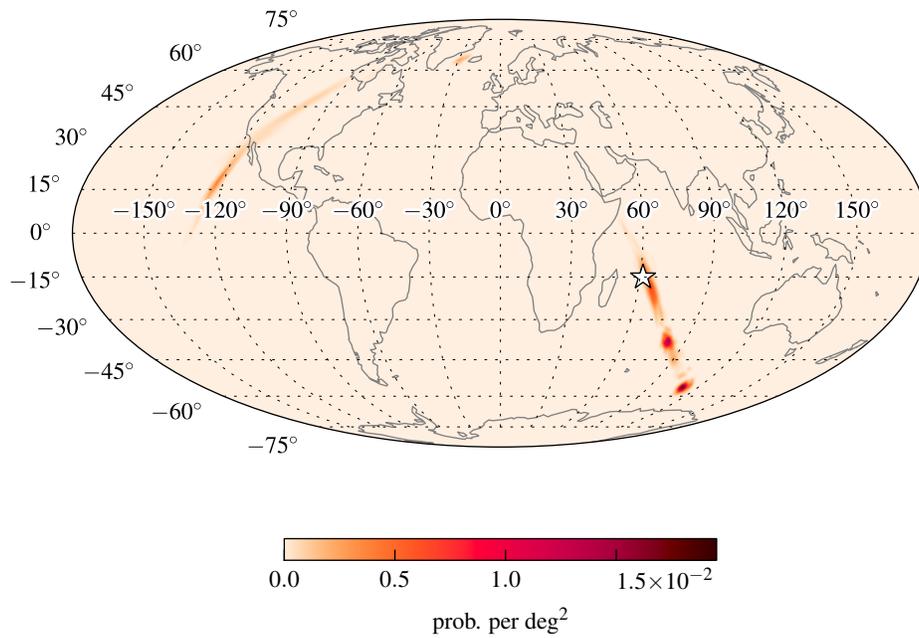
4. *HLV event, compactly localized* — This is another 2016, three-detector scenario. It describes many of the events that are detected with triggers in all three instruments. These are many of the best-localized events, with 90% confidence regions only a few times 10 deg^2 in area. These events are generally localized to one simply connected region and exhibit a less pronounced preference for particular sky locations. In this scenario, it is most likely that the rapid sky localization and the full parameter estimation will be similar. This is observationally the simplest scenario: just one of the several wide-field optical searches (for instance, ZTF or BlackGEM) would be able to scan the whole error region at a daily cadence.

5.3. Comparison with other studies

Figure 10. Localization of a typical circa 2016 GW detection in the HLV network configuration. This is a Mollweide projection in geographic coordinates. This event consists of triggers in H and L and has a $\text{SNR}_{\text{net}} = 13.4$. The rapid sky localization gives a 90% confidence region with an area of 1100 deg^2 and the full stochastic sampler gives 515 deg^2 . This is event ID 821759 from Tables 4 and 5 and the online supplement (see Appendix for more details). (A color version of this figure is available in the online journal.)

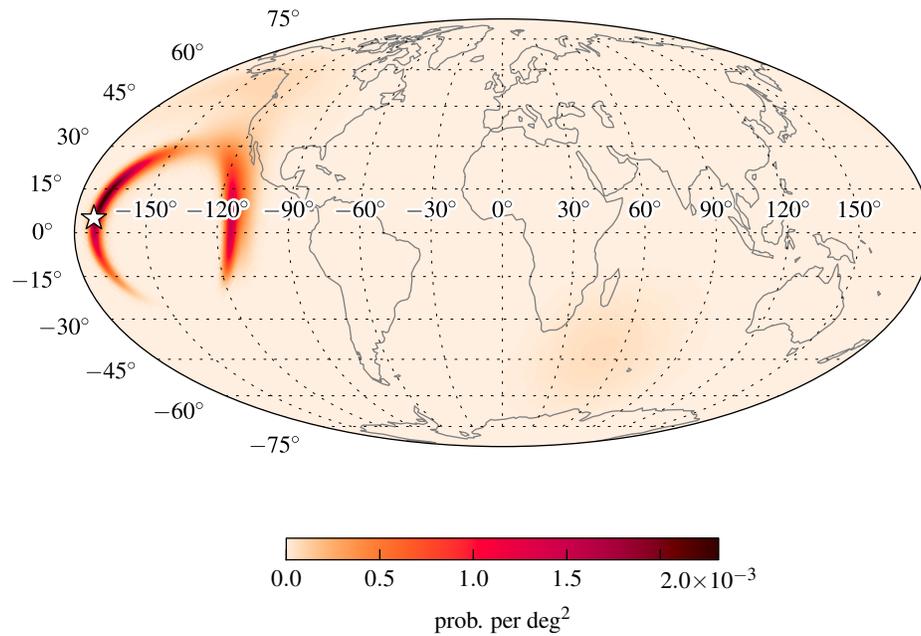


(a) BAYESTAR



(b) LALINFERENCE_MCMC

Figure 11. Rapid localization of a typical circa 2016 GW detection in the HV network configuration. This is a Mollweide projection in geographic coordinates. This event's $\text{SNR}_{\text{net}} = 12.2$ is near the detection threshold. Its 90% confidence area is 4600 deg^2 , but the true position of the source (marked with the white pentagram) is found after searching 65 deg^2 . This is event ID 655803 from Tables 4 and 5 and the online supplement (see Appendix for more details). (A color version of this figure is available in the online journal.)



This is the first study so far to combine an astrophysically motivated source population, realistic sensitivity and detector network configurations, event selection effects arising from a genuine detection pipeline instead of an ad-hoc threshold, and parameter estimation algorithms that will be deployed for GW data analysis. This study also has a much larger sample size and lower statistical uncertainty than most of the prior work. It is therefore somewhat difficult to compare results to other studies which each have some but not all of these virtues.

To the best of the authors’ knowledge, Raymond et al. (2009) were the first to point out the power of Bayesian priors for breaking sky degeneracies in two-detector networks, challenging a prevailing assumption at the time that two detectors could only constrain the sky location of a compact binary signal to a degenerate annulus. Aasi et al. (2013b) speculated that two-detector, 2015, HL configuration sky maps would be rings spanning “hundreds to thousands” of deg^2 , that coherence and amplitude consistency would “sometimes” resolve the localizations to shorter arcs. With our simulations, we would only revise that statement to read “hundreds to a thousand” deg^2 and change “sometimes” to “always.” Kasliwal & Nissanke (2014) recently argued for the feasibility of optical transient searches (in the context of kilonovae) with two-detector GW localizations.

Aasi et al. (2013b) used time-of-arrival triangulation (Fairhurst 2009) to estimate the fraction of sources with 90% confidence regions smaller than 20 deg^2 , finding a range of 5–12% for 2016. We find 14%. Our values are more optimistic, but perhaps also more realistic for the assumed detector sensitivity. Our sky localization takes into account phase and amplitude information, which Grover et al. (2014) points out can produce ≈ 0.4 times smaller areas compared to timing alone. However, it is clear from both Aasi et al. (2013b) and our study that such well-localized events will comprise an exceedingly small fraction of GW detections until the end of the decade. We therefore echo Kasliwal & Nissanke (2014) in stressing the importance of preparing to deal with areas of hundreds of deg^2 in the early years of Advanced LIGO and Virgo.

Nissanke et al. (2013) used an astrophysical distance distribution, drawing source positions uniformly in comoving volume for distances $d_L > 200 \text{ Mpc}$ and from a B -band luminosity-weighted galaxy catalog for distances $d_L \leq 200 \text{ Mpc}$. They generated sky maps using their own MCMC code. Similar to this study, they imposed a threshold of network $\text{SNR}_{\text{net}} > 12$. They explore several different GW detector network configurations. The most similar to our 2016 scenario is an HLV network at final design sensitivity. They find a median 95% confidence region area of $\sim 20 \text{ deg}^2$. For comparison, we find a 95 deg^2 confidence area of 374 deg^2 . Our much larger number is explained by several factors. First, we did not draw nearby sources from a galaxy catalog, so we have fewer loud, nearby sources. Second, since we accounted for duty cycle, poorly localized two-detector events account for a quarter of our sample. Third, and most important, we assumed Advanced Virgo’s anticipated initial sensitivity rather than its final design sensitivity.

Rodriguez et al. (2013) also studied an HLV network at final design sensitivity. Their simulated signals had identically zero noise content, the average noise contribution among all realizations of zero-mean Gaussian noise. All of their simulated events had a relatively high $\text{SNR}_{\text{net}} = 20$. They find a median 95% confidence area of 11.2 deg^2 . If we consider all

of our 2016 scenario HLV events with $19.5 \leq \text{SNR}_{\text{net}} \leq 20.5$, we find a median area of 126 deg^2 . Our significantly larger number is once again partly explained by our less sensitive Virgo detector, which introduces several multimodal events even at this high SNR_{net} .

Similarly, Grover et al. (2014) and Sidery et al. (2014) studied a 3-detector network, but at initial LIGO design sensitivity. These studies were primarily concerned with evaluating Bayesian parameter estimation techniques with respect to triangulation methods. They found much smaller areas, with a median of about 3 deg^2 . Both papers used a source population that consisted mainly of very high-SNR signals with binary black hole masses, with distances distributed logarithmically. All of these effects contribute to unrealistically small areas.

Finally, Kasliwal & Nissanke (2014) made the first small-scale systematic study of localizability with 2 LIGO detectors, albeit at final Advanced LIGO design sensitivity. For these noise curves and a $(1.4, 1.4) M_{\odot}$ binary with single-detector $\text{SNR} = 10$, Equation (4) gives a timing uncertainty of $142 \mu\text{s}$. Their different choice of noise curves should result in areas that are $(131/142)^2 \approx 0.85$ times smaller than ours, at a given SNR_{net} . As we have, they imposed a network SNR threshold of $\text{SNR}_{\text{net}} \geq 12$ on all detections¹⁹. They find a median 95% confidence area of $\sim 250 \text{ deg}^2$ from a catalog of 17 events. From our 2015 scenario, we find a median 90% confidence area that is almost twice as large, $\sim 500 \text{ deg}^2$. Though we cannot directly compare our 90% area to their 95% area, our 95% area would be still larger. Several factors could account for this difference, including the smaller sample size in Kasliwal & Nissanke (2014). Also Kasliwal & Nissanke (2014), like Nissanke et al. (2013), drew nearby sources from a galaxy catalog to account for clustering, so their population may contain more nearby, well-localized events than ours. Another difference is that Kasliwal & Nissanke (2014) do not report any multimodal localizations or the 180° degeneracy that we described in Section 4.1.

5.4. Conclusion

Many previous sky localization studies have found that networks of three or more advanced GW detectors will localize BNS mergers to tens of deg^2 . However, given realistic commissioning schedules, areas of hundreds of deg^2 will be typical in the early years of Advanced LIGO and Virgo.

We caution that multimodality and long, extended arcs will be a common and persistent feature of Advanced LIGO/Virgo detections. We caution that existing robotic follow-up infrastructure designed for GRBs, whose localizations are typically nearly Gaussian and unimodal, will need to be adapted to cope with more complicated geometry. In particular, we advise optical facilities to evaluate the whole GW sky map when determining if and when a given event is visible from a particular site.

We have elucidated a degeneracy caused by the relative orientations of the two LIGO detectors, such that position reconstructions will often consist of two islands of probability sep-

¹⁹ Nissanke et al. (2013); Kasliwal & Nissanke (2014) present a parallel set of results for a threshold $\text{SNR}_{\text{net}} > 8.5$, relevant for a coherent GW search described by Harry & Fairhurst (2011). However, the coherent detection statistic described by Harry & Fairhurst (2011) is designed for targeted searches at a known sky location (for instance, in response to a GRB). Thus the $\text{SNR}_{\text{net}} > 8.5$ threshold is not relevant for optical follow-up triggered by a detection from an all-sky GW search. Furthermore, this reduced threshold is not relevant to the HL configuration because the coherent detection statistic reduces to the network SNR for networks of two detectors or fewer.

arated by 180° . We have shown that this degeneracy is largely broken by adding Virgo as a third detector, even with its significantly lower sensitivity. We have shown that sub-threshold GW observations are important for sky localization and parameter estimation.

We have demonstrated a real-time detection, sky localization, and parameter estimation infrastructure that is ready to deliver Advanced LIGO/Virgo science. The current analysis has some limitations for the three-detector network, an undesirable trade-off of sky localization accuracy and timescale. Work is ongoing to lift these limitations by providing the rapid sky localization with information below the present single-detector threshold and by speeding up the full parameter estimation by a variety of methods (Canizares et al. 2013; Smith et al. 2013; Farr et al. 2013; Farr & Kalogera 2013; Graff et al. 2012).

We have exhibited an approach that involves three tiers of analysis, which will likely map onto a sequence of three automated alerts with progressively more information on longer timescales, much as how observers in the GRB community are used to receiving a sequence of Gamma-ray Coordinates Network (GCN) notices about a high-energy event.

The maximum timescale of the online GW analysis, about a day, is appropriate for searching for kilonova emission. However, the availability of BAYESTAR’s rapid localizations within minutes of merger makes it possible to search for X-ray and optical emission. Due to jet collimation, these early-time signatures are expected to accompany only a small fraction of LIGO/Virgo events. However, the comparative brightness and distinctively short timescale of the optical afterglow makes it an attractive target. PTF has recently proved the practicality of wide-field afterglow searches through the blind discovery of afterglow-like optical transients (Cenko et al. 2013; Cenko et al. 2014) and the detection of optical afterglows of *Fermi* GBM bursts (Singer et al. 2013). We encourage optical transient experiments such as ZTF and BlackGEM to begin searching for EM counterparts promptly based on the rapid GW localization. In the most common situation of no afterglow detection, the early observations may be used as reference images for longer-cadence kilonova searches.

A catalog of the sky maps used in this study is available from <http://www.ligo.org/scientists/first2years>. See the Appendix for more details.

L.P.S. and B.F. thank generous support from the National Science Foundation (NSF) in the form of Graduate Research Fellowships. B.F. acknowledges support through NSF grants DGE-0824162 and PHY-0969820. A.L.U. and C.P. gratefully acknowledge NSF support under grants PHY-0970074 and PHY-1307429 at the University of Wisconsin–Milwaukee (UWM) Research Growth Initiative. J.V. was supported by the research programme of the Foundation for Fundamental Research on Matter (FOM), which is partially supported by the Netherlands Organisation for Scientific Research (NWO), and by Science and Technology Facilities Council (STFC) grant ST/K005014/1. P.G. is supported by a NASA Postdoctoral Fellowship administered by the Oak Ridge Associated Universities.

GSTLAL analyses were produced on the NEMO computing cluster operated by the Center for Gravitation and Cosmology at UWM under NSF Grants PHY-0923409 and PHY-0600953. The BAYESTAR analyses were performed on the LIGO–Caltech computing cluster. The MCMC

computations were performed on Northwestern’s CIERA High-Performance Computing cluster GRAIL²⁰.

We thank Patrick Brady, Vicky Kalogera, Erik Katsavounidis, Richard O’Shaughnessy, Ruslan Vaulin, and Alan Weinstein for helpful discussions.

This research made use of Astropy²¹ (Robitaille et al. 2013), a community-developed core Python package for Astronomy. Some of the results in this paper have been derived using HEALPix (Górski et al. 2005). Public-domain cartographic data is courtesy of Natural Earth²² and processed with MapShaper²³.

BAYESTAR, LALINFERENCE_NEST, and LALINFERENCE_MCMC are part of the LIGO Algorithm Library Suite²⁴ and the LIGO parameter estimation package, LALINFERENCE. Source code for GSTLAL²⁵ and LALINFERENCE²⁶ are available online under the terms of the GNU General Public License.

LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the NSF and operates under cooperative agreement PHY-0757058.

This is LIGO document number LIGO-P1300187-v24.

APPENDIX

We describe a catalog of all simulated events, detections, and sky maps that were generated in this study.

For the 2015 scenario, parameters of simulated signals are given in Table 2. In the same order, parameters of the detection including the operating detector network, false alarm rate, SNR_{net} , SNR in each detector, recovered masses, and sky localization areas are given in Table 3. For the 2016 scenario, the simulated signals are recorded in Table 4 and the detections in Table 5. In the printed edition, parameters are given for just the four sample events that appear earlier in the text (see Figures 6, 7, 10, and 11). In the machine readable tables in the online edition, parameters are given for all detected signals.

The tables give two integer IDs. The ‘event ID’ column corresponds to a field in that scenario’s full GSTLAL output that identifies one coincident detection candidate. The ‘simulation ID’ likewise identifies one simulated signal. In the full GSTLAL output, there may be zero or many event candidates that match any given simulated signal. However, in our catalog there is one-to-one correspondence between simulation and event IDs because we have retained only simulated signals that are detected above threshold, and only the highest SNR detection candidate for each signal.

Note that the arbitrary dates of the simulated signals range from 21 August through 19 October 2010. This reflects the two-month duration of the simulated data stream, not the dates or durations of the anticipated Advanced LIGO/Virgo observing runs.

For convenience, we also provide a browsable sky map catalog²⁷. This web page provides a searchable version of Tables 2, 3, 4, and 5, with posterior sky map images from both the rapid parameter estimation and the stochastic samplers.

²⁰ http://ciera.northwestern.edu/Research/GraIl_Cluster.php

²¹ <http://www.astropy.org>

²² <http://www.naturalearthdata.com>

²³ <http://www.mapshaper.org>

²⁴ <http://www.lsc-group.phys.uwm.edu/cgit/lalsuite/tree>

²⁵ <http://www.lsc-group.phys.uwm.edu/cgit/gstlal/tree/>

²⁶ <http://www.lsc-group.phys.uwm.edu/cgit/lalsuite/tree/lalinference>

²⁷ <http://www.ligo.org/scientists/first2years>

The web page also provides, for each localization, a FITS file representing the posterior in the HEALPix projection (Górski et al. 2005) using the NESTED indexing scheme. For reading these files, the authors recommend the Python package Healpy²⁸ or the HEALPix C/C++/IDL/Java/Fortran library²⁹. They can also be displayed by many standard imaging programs such as DS9³⁰ and Aladin³¹.

Synthetic gravitational-wave time series data and posterior sample chains are available upon request.

REFERENCES

- Aasi, J., et al. 2013a, *Phys. Rev. D*, 88, 062001
 —. 2013b, arXiv:1304.0670
 Abadie, J., et al. 2012, arXiv:1203.2674
 —. 2010, *Class. Quantum Grav.*, 27, 173001
 Acernese, F., et al. 2013, *Tech. Rep. VIR-0027A-09*
 Adhikari, R. X. 2014, *Reviews of Modern Physics*, 86, 121
 Allen, B. 2005, *Phys. Rev. D*, 71, 062001
 Allen, B., et al. 2012, *Phys. Rev. D*, 85, 122006
 Anderson, W. G., et al. 2001, *Phys. Rev. D*, 63, 042003
 Babak, S., et al. 2013, *Phys. Rev. D*, 87, 024033
 Barnes, J., & Kasen, D. 2013, *ApJ*, 775, 18
 Barsotti, L., & Fritschel, P. 2012, *Early aLIGO Configurations: example scenarios toward design sensitivity*, *Tech. Rep. LIGO-T1200307-v4*
 Belczynski, K., Bulik, T., & Bailyn, C. 2011, *ApJ*, 742, L2
 Belczynski, K., et al. 2013, *ApJ*, 764, 96
 Blackman, J., et al. 2014
 Brown, D. A., et al. 2012, *Phys. Rev. D*, 86, 084017
 Buonanno, A., & Chen, Y. 2001, *Phys. Rev. D*, 64, 042006
 Buonanno, A., Chen, Y., & Vallisneri, M. 2003, *Phys. Rev. D*, 67, 104025
 —. 2006, *Phys. Rev. D*, 74, 029904
 Buonanno, A., et al. 2009, *Phys. Rev. D*, 80, 084043
 Burgay, M., et al. 2003, *Nature*, 426, 531
 Cameron, E. 2011, *Proc. Astron. Soc. Aust.*, 28, 128
 Canizares, P., et al. 2013, *Phys. Rev. D*, 87, 124005
 Cannon, K., et al. 2012, *ApJ*, 748, 136
 Cannon, K., Hanna, C., & Keppel, D. 2013, *Phys. Rev. D*, 88, 024025
 Cenko, S. B., et al. 2014, *GCN*, 15883, 1
 Cenko, S. B., et al. 2013, *ApJ*, 769, 130
 Eichler, D., et al. 1989, *Nature*, 340, 126
 Fairhurst, S. 2009, *New Journal of Physics*, 11, 123006
 Fairhurst, S. 2011, *Classical and Quantum Gravity*, 28, 105021
 Fairhurst, S. 2012, arXiv:1205.6611
 Farr, B., & Kalogera, V. 2013, in *APS Meeting Abstracts*, 10003
 Farr, B., Kalogera, V., & Lijjten, E. 2013, arXiv:1309.7709
 Górski, K. M., et al. 2005, *ApJ*, 622, 759
 Graff, P., et al. 2012, *MNRAS*, 421, 169
 —. 2013, arXiv:1309.0790
 Grover, K., et al. 2014, *Phys. Rev. D*, 89, 042004
 Hanna, C. 2008, PhD thesis, Louisiana State University
 Hannam, M., et al. 2013, arXiv:1308.3271
 Harry, G. M. 2010, *Class. Quantum Grav.*, 27, 084006
 Harry, I., et al. 2013, arXiv:1307.3562
 Harry, I. W., & Fairhurst, S. 2011, *Class. Quantum Grav.*, 28, 134008
 Hessels, J. W. T., et al. 2006, *Science*, 311, 1901
 Hinshaw, G., et al. 2013, *ApJS*, 208, 19
 Ivezic, Z., et al. 2008, arXiv:0805.2366
 Kasliwal, M. M., & Nissanke, S. 2014, *ApJ*, 789, L5
 Keppel, D. 2012, *Phys. Rev. D*, 86, 123010
 —. 2013, arXiv:1307.4158
 Klimentenko, S., et al. 2011, *Phys. Rev. D*, 83, 102001
 Kulkarni, S. R. 2012, arXiv:1202.2381
 Li, L.-X., & Paczyński, B. 1998, *ApJ*, 507, L59
 Maggiore, M. 2008, *Gravitational Waves: Volume 1: Theory and Experiments*, *Gravitational Waves (OUP Oxford)*
 Metzger, B. D., & Berger, E. 2012, *ApJ*, 746, 48
 Nakar, E. 2007, *Phys. Rep.*, 442, 166
 Nakar, E., & Piran, T. 2011, *Nature*, 478, 82
 Narayan, R., Paczynski, B., & Piran, T. 1992, *ApJ*, 395, L83
 Nissanke, S., Kasliwal, M., & Georgieva, A. 2013, *ApJ*, 767, 124
 Nissanke, S., et al. 2011, *ApJ*, 739, 99
 Nitz, A. H., et al. 2013, arXiv:1307.1757
 Özel, F., et al. 2012, *ApJ*, 757, 55
 Paczynski, B. 1986, *ApJ*, 308, L43
 Pejcha, O., Thompson, T. A., & Kochanek, C. S. 2012, *MNRAS*, 424, 1570
 Piran, T., Nakar, E., & Rosswog, S. 2013, *MNRAS*, 430, 2121
 Raymond, V., et al. 2009, *Classical and Quantum Gravity*, 26, 114007
 Rezzolla, L., et al. 2011, *ApJ*, 732, L6
 Robitaille, T. P., et al. 2013, *A&A*, 558, A33
 Rodriguez, C. L., et al. 2013, arXiv:1309.3273
 Schutz, B. F. 2011, *Classical and Quantum Gravity*, 28, 125023
 Sidery, T., et al. 2014, *Phys. Rev. D*, 89, 084060
 Singer, L. P., et al. 2013, *ApJ*, 776, L34
 Smith, R., et al. 2013, arXiv:1305.3798
 Taracchini, A., et al. 2013, arXiv:1311.2544
 van der Sluys, M., et al. 2008a, *Classical and Quantum Gravity*, 25, 184011
 van der Sluys, M. V., et al. 2008b, *ApJ*, 688, L61
 Veitch, J., & Vecchio, A. 2010, *Phys. Rev. D*, 81, 062003
 Veitch, J., et al. 2012, *Phys. Rev. D*, 85, 104045
 Vitale, S., & Zanolin, M. 2011, *Phys. Rev. D*, 84, 104020
 Vogt, R. E. 1989, *A Laser Interferometer Gravitational-Wave Observatory*, *Tech. Rep. PHY-8803557*, Proposal to the National Science Foundation
 Wen, L., & Chen, Y. 2010, *Phys. Rev. D*, 81, 082001

²⁸ <http://healpy.readthedocs.org>

²⁹ <http://healpix.sourceforge.net>

³⁰ <http://ds9.si.edu>

³¹ <http://aladin.u-strasbg.fr>

Table 2
Simulated BNS signals for 2015 scenario.

event ID ^a	sim ID ^b	MJD ^c	α	orientation ^d				d (Mpc)	masses (M_{\odot})		spin 1				spin 2		
				δ	ι	ψ	ϕ_c		m_1	m_2	S_1^x	S_1^y	S_1^z	S_2^x	S_2^y	S_2^z	
18951	10807	55442.25864	137.8	-39.9	139	43	42	75	1.40	1.51	-0.01	-0.01	-0.04	-0.05	-0.01	+0.01	
20342	21002	55454.76654	19.8	-23.7	145	197	145	75	1.34	1.48	-0.03	+0.01	-0.03	-0.01	+0.02	-0.01	
...	

Note. — Table 2 is published in its entirety in the electronic edition of the Astrophysical Journal. A portion is shown here for guidance regarding its form and content.

^a Identifier for detection candidate. This is the same value as the `coinc_event_id` column in the GSTLAL output database and the OBJECT cards in sky map FITS headers, with the `coinc_event:coinc_event_id:` prefix stripped.

^b Identifier for simulated signal. This is the same value as the `simulation_id` column in the GSTLAL output database, with the `sim_inspiral:simulation_id:` prefix stripped.

^c Time of arrival at geocenter of GWs from last stable orbit.

^d α : RA, δ : Dec (J2000), ι : binary orbital inclination angle, ψ : polarization angle (Anderson et al. 2001, Appendix B), ϕ_c : orbital phase at coalescence.

Table 3
Detections and sky localization areas for 2015 scenario.

event ID	sim ID	network	net ^a	SNR			masses ^b		BAYESTAR			LALINFERENCE_NEST		
				H	L	V	m_1	m_2	50%	90%	searched	50%	90%	searched
18951	10807	HL	15.0	10.3	10.9	1.67	1.27	160	630	130	160	680	81	
20342	21002	HL	12.7	7.3	10.3	1.59	1.25	130	530	17	170	620	12	
...	

Note. — Table 3 is published in its entirety in the electronic edition of the Astrophysical Journal. A portion is shown here for guidance regarding its form and content.

^a Network SNR, or root-sum-squared SNR over all detectors.

^b Maximum likelihood estimate of masses as reported by GSTLAL.

Table 4
Simulated BNS signals for 2016 scenario.

event ID	sim ID	MJD	α	orientation				d (Mpc)	masses (M_{\odot})		spin 1				spin 2		
				δ	ι	ψ	ϕ_c		m_1	m_2	S_1^x	S_1^y	S_1^z	S_2^x	S_2^y	S_2^z	
655803	45345	55484.63177	79.2	+5.0	121	321	69	66	1.60	1.29	+0.00	+0.00	-0.00	+0.00	+0.00	-0.00	
821759	8914	55439.93634	18.3	-15.1	158	257	230	187	1.60	1.45	-0.00	+0.02	-0.01	+0.04	+0.03	-0.02	
...	

Note. — Table 4 is published in its entirety in the electronic edition of the Astrophysical Journal. A portion is shown here for guidance regarding its form and content.

Table 5
Detections and sky localization areas for 2016 scenario.

event ID	sim ID	network	net	SNR			masses		BAYESTAR			LALINFERENCE_MCMC		
				H ^a	L ^a	V ^a	m_1	m_2	50%	90%	searched	50%	90%	searched
655803	45345	HV	12.2	11.5	4.2	1.52	1.35	480	4600	65	300	4000	21	
821759	8914	HLV	13.4	8.5	10.4	1.57	1.47	340	1100	470	91	510	94	
...	

Note. — Table 5 is published in its entirety in the electronic edition of the Astrophysical Journal. A portion is shown here for guidance regarding its form and content.

^a Blank if SNR < 4 or detector is not online.