Strategies for the follow-up of gravitational wave transients with the Cherenkov Telescope Array

I. Bartos,1,2★ T. Di Girolamo,2,3★ J. R. Gair,4 M. Hendry,5 I. S. Heng,5 T. B. Humensky,2 S. Márka,2 Z. Márka,2 C. Messenger,5 R. Mukherjee,6 D. Nieto,7 P. O’Brien8 and M. Santander6,9

1 Department of Physics, University of Florida, Gainesville, FL 32611, USA
2 Department of Physics, Columbia University, New York, NY 10027, USA
3 Dipartimento di Fisica ‘Ettore Pancini’ dell’Università ‘Federico II’ and Istituto Nazionale di Fisica Nucleare, I-80126 Napoli, Italy
4 School of Mathematics, University of Edinburgh, Edinburgh EH9 3FD, UK
5 SUPA, University of Glasgow, Glasgow G12 8QQ, UK
6 Department of Physics & Astronomy, Barnard College, Columbia University, New York, NY 10027, USA
7 Facultad de Ciencias Físicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain
8 Space Research Centre, Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK
9 Department of Physics & Astronomy, University of Alabama, Tuscaloosa, AL 35487, USA

Accepted 2018 February 28. Received 2018 February 27; in original form 2018 January 29

ABSTRACT

The observation of the electromagnetic counterpart of gravitational-wave (GW) transient GW170817 demonstrated the potential in extracting astrophysical information from multimessenger discoveries. The forthcoming deployment of the first telescopes of the Cherenkov Telescope Array (CTA) observatory will coincide with Advanced LIGO/Virgo’s next observing run, O3, enabling the monitoring of gamma-ray emission at $E > 20$ GeV, and thus particle acceleration, from GW sources. CTA will not be greatly limited by the precision of GW localization as it will be capable of rapidly covering the GW error region with sufficient sensitivity. We examine the current status of GW searches and their follow-up effort, as well as the status of CTA, in order to identify some of the general strategies that will enhance CTA’s contribution to multimessenger discoveries.

Key words: gravitational waves – instrumentation: detectors – gamma-ray burst: general.

1 INTRODUCTION

The Advanced LIGO observatories’ first two observing runs saw an extensive effort to search for multimessenger emission from gravitational-wave (GW) sources, covering both the electromagnetic and neutrino spectra (Abbott et al. 2016i; Adrián-Martínez et al. 2016; Abbott et al. 2017c; Albert et al. 2017a,b). This effort culminated, on 2017 August 17, with the observation across the electromagnetic spectrum of a gamma-ray burst (GRB) and a kilonova as a consequence of a binary neutron star merger which was detected in GWs (Abbott et al. 2017b,c,d). The Advanced Virgo interferometer (Acernese et al. 2015) was also operational at this time and aided the discovery, reducing the sky localization region. The success of the observational campaign for this event, which marked the start of multimessenger astronomy with GWs, shows the importance of coordinated follow-up observations and of the strategies to carry out them. In the next few years, the LIGO and Virgo detectors will continuously improve their sensitivity (Abbott et al. 2016b), while additional GW interferometers are envisaged to come online (Iyer et al. 2011; Aso et al. 2013), promising the regular detection of a variety of sources with potential multimessenger signatures.

The Cherenkov Telescope Array (CTA, Acharya et al. 2013) will soon expand the multimessenger observational horizon with an unprecedented sensitivity to sources producing very high-energy ($>10$ GeV) gamma-ray emission (Ong 1998). The first CTA telescopes are envisaged to come online in 2019, bringing the first joint CTA-GW searches during the LIGO/Virgo third observing period. CTA will continuously increase its sensitivity as more telescopes are installed, broadening its reach in parallel with increased GW capabilities.

Multimessenger sources of interest include the formation of black hole–accretion disc systems that drive relativistic outflows, giving rise to high-energy emission. Such a system can arise from the formation or merger of compact objects, such as neutron star–neutron star or black hole–neutron star mergers (Abadie et al. 2010), core collapse supernovae (CCSNe) with rapidly rotating cores (Bartos, Brady & Márka 2013), and plausibly from binary black hole (BBH) mergers (Loeb 2016; Murase et al. 2016; Perna, Lazzati &...
A binary system with total mass up to a few hundred solar masses will generate GWs detectable by Advanced LIGO/Virgo, if the event occurs within the detector’s horizon, during the final stages of binary inspiral and merger (Sathyaprakash & Schutz 2009). The requirement that the binary system is about to merge ensures that only binary systems containing neutron stars or black holes are potential sources of GW emission for Advanced LIGO. During this final phase leading through merger and into ringdown, the evolution of the binary system is expected to be gravitationally dominated and the GW emission can therefore be accurately predicted by solving the Einstein equations of general relativity. The inspiral can be characterized using post-Newtonian theory, which constructs GW emission by evaluating an expansion of the field equations in powers of the velocity (Blanchet 2014). The post-Newtonian expansion cannot be used to describe the last few cycles of inspiral and the subsequent merger, where the velocities of the binary components approach a significant fraction of the speed of light. Modelling this portion of the signal requires a full numerical solution of Einstein’s field equations on a computer using numerical relativity (Pretorius 2005). After merger, the two binary components form a single, highly perturbed, black hole, which then settles down to a stationary state through evaporation of GW radiation as a superposition of damped sinusoids, which is known as the ringdown. Two hybrid waveform families also now exist that smoothly combine the three phases into a single model that is tuned to match numerical relativity simulations (Khan et al. 2016; Bohé et al. 2017) and is better suited for use in GW data analysis. The fact that all three phases of the GW signal can be well modelled allows binary systems to be identified in the LIGO/Virgo data set using matched filtering, which significantly improves the distance to which such systems can be observed.

Binary neutron star mergers sweep through the whole of the LIGO/Virgo observation band of ∼10–1000 Hz. The final stages of merger occur at frequencies away from the most sensitive part of the LIGO noise curve and the majority of the signal-to-noise ratio and information comes from the inspiral portion of the signal.

BBH systems have higher masses and hence reach merger at lower frequencies, which for systems with total mass of a few tens of solar masses can be in the most sensitive portion of the LIGO noise spectrum. The majority of the signal-to-noise for higher mass BBH systems comes from the final stages of inspiral, the merger, and subsequent ringdown. The higher mass typical of BBH systems means they can be observed to greater distances, with systems like GW150914 (component masses of ∼30 and ∼35 M⊙) being detectable to distances of ∼9 Gpc by Advanced LIGO at its design configuration (Abbott et al. 2016h). Systems with masses much higher than GW150914 are still potentially detectable by Advanced LIGO, but only during the merger and ringdown phases. These systems will therefore tend to be less well characterized than lighter BBHs (The LIGO Scientific Collaboration 2017). For systems with total mass above ∼500 M⊙ only the ringdown signal is in the LIGO frequency band and LIGO’s sensitivity to such systems is significantly poorer (Aasi et al. 2014).

CCSNe are also potential sources of GW emission (Abbott et al. 2016d). Significant GW emission occurs only if there is substantial asymmetric acceleration of the stellar material during the supernova. A number of mechanisms generating the required asymmetries have been discussed. The most extensively studied is the presence of significant rotation during core collapse. Rotation generates an asymmetric oblate deformation of the collapsing core. The extreme acceleration of the material at core-bounce generates a burst of GWs that is linearly polarized. In slowly rotating stars, the rotating core collapse model does not apply, but significant GW emission can be generated by neutrino-driven convection or the standing accretion shock instability. Simulations of CCSNe in 2D and 3D have been...
used to demonstrate the generation of GWs, but in general the core collapse mechanism is complex such that the GW emission cannot be sufficiently well modelled to generate templates for data analysis. However, the GW emission from a CCSN is expected to be short in duration and broad in spectrum, making these good candidates for LIGO/Virgo burst detection algorithms (Abbott et al. 2016d).

Short GRBs are thought to be powered by neutron star–neutron star (NS–NS) or neutron star–black hole (NS–BH) mergers (e.g. see Berger 2014). The unambiguous association of GRB170817A with GW170817 recently confirmed this hypothesis at least for some short GRBs, however with a short prompt emission extending only to ∼1 MeV (Abbott et al. 2017d; Goldstein et al. 2017; Savchenko et al. 2017), while at very high energies the H.E.S.S. IACTs set some upper limits at later times (Abdalla et al. 2017). Fermi-LAT has detected emission above 100 MeV from several short GRBs, most notably GRB090510 (Abdo et al. 2009a; Ackermann et al. 2010). The GeV component is likely produced by inverse Compton scattering, though the nature of the seed photon population is still unsettled and may depend on the environment the GRB is expanding into; possibilities include the synchrotron photons produced by electrons accelerated at the external shock generated when the relativistic jet is decelerated by the external medium (Meszaros & Rees 1994; Meszaros, Rees & Papathanassiou 1994), or prompt radiation emitted at smaller radii. In either case, the thermal plasma behind the external shock provides the energetic leptons (Beloborodov, Hascoët & Vurm 2014). Because Fermi-LAT is fluence limited, it is clear that CTA will have the raw sensitivity required to detect a significant population of short GRBs. A time delay may be possible between a GW trigger and any short GRB emission in the case of an NS–NS merger, if the merger yields a supramassive NS [though Margalit, Metzger & Beloborodov (2015) argues against the viability of this scenario]; in this case the GRB may be delayed by O(10^3 s) with respect to the GW trigger (Vietri & Stella 1998; Cioffi & Siegel 2015), which will need to be taken into consideration in designing the electromagnetic follow-up strategy. While photon–photon absorption can suppress the emission of gamma rays above tens of GeV, a sufficiently high bulk Lorentz factor at late times can allow a significant flux of gamma rays to escape. Photon–photon absorption due to the EBL can further suppress these gamma rays for GRBs occurring at redshifts beyond the anticipated LIGO/Virgo horizon. An estimate of the rate of detections by CTA of short GRBs associated with GWs gives ~0.03 yr^{-1} (Bartos et al. 2014); however, considering off-axis events like the recent GRB170817A, this rate should increase (Abbott et al. 2017d; Lazzati et al. 2017).

Some long GRBs are associated with the core collapse of massive stars (Woosley & Bloom 2006). If the collapse is asymmetric enough to produce a detectable GW emission, its signal should precede the burst (Kobayashi & Mészáros 2003; van Putten et al. 2004). Fermi-LAT observations showed that very high-energy gamma-ray emission is a relatively common feature in long GRBs [e.g. GRB130427A (Ackermann et al. 2014) and GRB080916C (Atwood et al. 2013)]. Some very high-energy photons were detected after the prompt emission, such as the 33 GeV photon from GRB090902B, which arrived 82 s after the trigger time and about 50 s after the end of the prompt-phase emission (Abdo et al. 2009b). This burst also exhibited a power-law component at GeV energies, distinct from the usual band model emission at MeV energies (Band et al. 1993), and this component showed significant spectral hardening towards the end of the prompt phase. The theoretical model proposed in Beloborodov et al. (2014) can explain the emission of GeV photons from long GRBs in connection with their massive progenitors (Hascoët, Vurm & Beloborodov 2015), also predicting TeV emission up to hours after the explosion (Vurm & Beloborodov 2017).

The Gamma-ray Burst Monitor (GBM) on the Fermi satellite detected a weak transient above 50 keV, 0.4 s after the event GW150914, with a false-alarm probability of 2.9σ (Connaughton et al. 2016). If this transient, lasting 1 s and with duration and spectrum consistent with a weak short GRB at a large angle to the detector pointing direction, is indeed associated with GW150914 and not a chance coincidence, it is an unexpected electromagnetic emission from a BBH (Lyutikov 2016). However, particular environmental conditions for the BBH merger may give sufficient local material to also produce transient high-energy gamma emission (Janiuk, Charzyński & Bejger 2013; Loeb 2016; Murase et al. 2016; Perna, Lazzati & Giacomazzo 2016; Yamazaki, Asano & Ohira 2016; Zhu & Wang 2016; Bartos et al. 2017b). None the less, the GRB origin of Fermi-GBM’s detection has been debated (Greiner et al. 2016; Xiong 2016), and no corresponding signal was found by the SPI instrument on board the INTEGRAL (Savchenko et al. 2016) satellite in the same energy region. The AGILE satellite, while it did not cover the GW localization region during the time of GW150914, provided limits on gamma-ray emission in the minutes prior, and following, the prompt event (Tavani et al. 2016). Recently, a refined analysis of the data of the MiniCALorimeter (MCAL) on the AGILE satellite, operating in the energy band 0.4–100 MeV, found a weak event lasting about 32 ms and occurring 0.46 s before GW170104, with a post-trial significance of 2.4–2.7σ (Verrecchia et al. 2017), which was also produced by the coalescence of a BBH (Abbott et al. 2017a). The characteristics of this event are similar to those of the weak precursor of short GRB090510, also in its timing, being detected about 0.46 s before its brightest emission by both AGILE-MCAL and Fermi-GBM (Abdo et al. 2009a; Giuliani et al. 2010). If confirmed by different space instruments, this association would prove that a BBH coalescence may be preceded by electromagnetic emission.

3 Gravitational-wave detectors and observation strategies

3.1 Gravitational-wave detectors in the CTA era

In 2018, we expect Advanced LIGO and Virgo to start taking data in their observing run, O3, at almost their design sensitivities, with sensitive ranges to binary neutron stars between 120–170 and 65–85 Mpc, respectively (see table I in Abbott et al. 2016b). Given the discovery of GW170817, there can potentially be multiple BNS merger detections during this run. By the end of the decade and beyond, advanced detectors will reach their design sensitivities, with ranges to BNS systems of ∼190 and ∼125 Mpc in Advanced LIGO and Virgo, respectively (Abbott et al. 2016b). Here, we defined range as the volume and orientation-averaged distance at which the source can be detected.

Furthermore, there are plans to implement technology upgrades to Advanced LIGO to further improve its sensitivity (LIGO Scientific Collaboration 2015). This upgraded Advanced LIGO detector, often referred to as “A+”, will likely come into operation sometime after 2020 and lead to sensitivity corresponding to a range of 320 Mpc for BNS signals.

About 10–20 per cent of detected BNS events will have sky localization with uncertainties of 20 deg^2 or less during O3 when a three-detector network consisting of the LIGO and Virgo interferometers will be in operation. With the addition of LIGO India to
the global network, approximately half of all observed BNS events will have sky location uncertainties of 20 deg$^2$ or less.

The Japanese KAGRA detector is being constructed underground near the Kamioka mines to reduce seismic noise (Aso et al. 2013). It will have cryogenically cooled test masses to reduce thermal noise. KAGRA will operate with a simple Michelson interferometer configuration from 2018 onwards before upgrading to the full interferometric configuration, with technologies including Fabry–Perot cavities and signal recycling, in 2019 (Abbott et al. 2016b).

3.2 Sensitivity to potential gravitational-wave sources

The detection of GW170817 gives an estimated BNS rate of \(1.5\times10^2 \text{ Gpc}^{-3} \text{ yr}^{-1}\) in the local Universe (Abbott et al. 2017b). This rate is remarkably consistent with previous expectations (Abadie et al. 2010). We can use the expected average distance of GW detectors out to which a BNS merger could be detected to calculate the expected detection rate. Assuming (i) an average distance of 120–170 Mpc for the O3 observing run and 190 Mpc at design sensitivity, (ii) 75 per cent single-detector duty cycle for LIGO and requiring that 2 LIGO detectors are operational for a detection, and (iii) a full year of operation for O3 (Abbott et al. 2016b), we obtain an expected 1–54 detections for O3, and an expected rate of 5–76 yr$^{-1}$ at design sensitivity.

The lack of NSBH binary detections by Advanced LIGO to date has allowed an upper limit on their rate of 3600 Gpc$^{-3}$ yr$^{-1}$ (Abbott et al. 2016b) to be determined.

The expected number of NSBH coalescences is 0.01–100 in the O3 observing run and rises in a similar way to the expected BNS coalescences for future observing runs (Abbott et al. 2016b). This range has some sensitivity on the assumptions about the NSBH population, in particular the mass of the black hole and assumptions about the alignment of the black hole spin with the orbital angular momentum. However, going from a conservative population (5 M$_\odot$ black holes and randomly oriented spins) to an optimistic population (30 M$_\odot$ black holes with aligned spins) changes the expected rate by only a factor of 3. For BNS systems we expect spins to be small and the two components to have masses close to 1.4 M$_\odot$ so there is little dependence on the system parameters.

Advanced LIGO observations of GW150914, LVT151012, and GW151226 have constrained the rate of BBH coalescences to be in the range 12–213 Gpc$^{-3}$ yr$^{-1}$ (Abbott et al. 2017a). This range includes both statistical uncertainties and uncertainties arising from the astrophysical population of BBH systems. The higher end of the rate range comes from assuming that the mass distribution in BBH binaries follows a power law with slope \(\alpha = -2.35\), and the lower end comes from assuming the mass distribution is flat in logarithmic scale. The power-law distribution intrinsically predicts fewer heavier black hole systems, to which Advanced LIGO is more sensitive, and hence the allowed rate of mergers is higher since it is dominated by systems that Advanced LIGO can only detect at moderate distances (Abbott et al. 2016k). During the O3 science run, Advanced LIGO should be able to detect equal-mass, non-spinning BBH systems with total mass of 20 M$_\odot$ out to distances of \(\sim 1.5 \text{ Gpc}\). The range increases as the total mass of the system increases, to \(\sim 3 \text{ Gpc}\) for systems of total mass 40 M$_\odot$, to \(\sim 4.5 \text{ Gpc}\) for systems with total mass of 60 M$_\odot$ and then \(\sim 8 \text{ Gpc}\) for systems with total mass of 100 M$_\odot$ (Abbott et al. 2016h). All of these ranges approximately double by the time Advanced LIGO achieves its design sensitivity. The range is increased if the components in the BBH have significant spins, and reduced (at fixed total mass) if the two components have unequal masses. The events observed by Advanced LIGO/Virgo to date are all nearly equal mass (mass ratios between 0.5 and 1) and only GW151226 shows evidence for spin, and that spin is moderate (effective spin \(\sim 0.2\)) (Abbott et al. 2016a). If these events are representative of the true astrophysical population, the non-spinning equal-mass ranges provide an accurate indication of LIGO’s likely sensitivity to BBH mergers.

Based on current rate estimates, there is between a 90 per cent and 99 per cent probability that Advanced LIGO and Virgo will observe more than 10 BBH mergers during the O3 science run in 2018–2019. The probability range arises from uncertainties in the sensitivity that will be achieved during that science run. There is also a probability of between 20 per cent and 80 per cent that Advanced LIGO and Virgo will observe more than 40 BBH events during O3 (Abbott et al. 2016a). At design sensitivity, Advanced LIGO’s range will be double that during O3 and so it will be probing a comoving volume between 3 and 5 times larger (depending on the mass of the BBH system). We would therefore expect Advanced LIGO to be observing 100 events per year at design sensitivity with high probability, with a plausible range for the expected number of detections per year of \(\sim 50–1000\) (Abbott et al. 2016k,g,a).

Numerical CCSN simulations suggest that Advanced LIGO/Virgo may be able to detect GWs from a CCSN event if it will occur within the Milky Way, but probably not at greater distances. However, if GW emission will be generated at much higher amplitudes than numerical simulations suggest, as predicted by some extreme analytic models of exotic emission scenarios, then Advanced LIGO could have a distance reach of as much as \(\sim 10 \text{ Mpc}\) (Abbott et al. 2016d). The Galactic supernova rate is 0.02–0.03/yr, which makes a detection during Advanced LIGO/Virgo operations plausible but improbable, although a Milky Way supernova is overdue. The CCSN rate within 15 Mpc is \(\sim 1/\text{yr}\), so the prospects for Advanced LIGO detection are significantly improved if CCSN\text{e} do generate much more GW emission than current numerical models suggest (Abbott et al. 2016d).

3.3 Low-latency electromagnetic follow-up observations

The utility of GWs in studying astrophysical processes is greatly increased by the simultaneous observation of electromagnetic and/or neutrino emission from the same sources. Consequently, there is a significant effort to enable Earth-based GW detectors to rapidly identify and localize GW source candidates, and share this information with partner observatories (Piscicincere et al. 2007; Kanner et al. 2008; LIGO Scientific Collaboration 2012; Smith et al. 2013; Abbott et al. 2016i; Adrián-Martínez et al. 2016). GW candidates were rapidly shared with a large number of partner observatories already during the operation of Initial LIGO-Virgo (Abadie et al. 2012), which was further expanded during LIGO’s first observing run (O1; Abbott et al. 2016i). Additional improvements and an increased quality of communication were implemented for the second observing run (O2; Abbott et al. 2017c).

Going forward, GW candidates will be shared at increasing rates and decreasing latency. The false alarm rate of shared triggers was set at 1/month during the O2 observing run both for compact binary merger candidates and separately for GW transient candidates identified with minimal assumptions on the source (Abbott et al. 2016c). The trigger rate will likely be higher than this, given the expected large and growing rate of GW detections (The LIGO Scientific Collaboration 2016).

Follow-up observatories will be able to selectively investigate GW triggers. For example, they can constrain follow-up searches...
based on the reconstructed progenitor type, source distance, direction, significance, and other parameters. This sub-selection can optimize the use of telescope time and the follow-up strategy.

A significant amount of information will be rapidly available from GW reconstruction algorithms that can facilitate optimizing follow-up strategies. Information on GW candidates include their time of arrival, localization (skymap), the type of GW search pipeline that detected the candidate (template-based compact binary search, generic transient search with minimal assumptions, or both), and false alarm rate. Additionally, for compact binary mergers, a source distance estimate will be available, along with whether at least one of the objects in the binary is a neutron star that could be disrupted. This latter case differentiates between likely electromagnetically bright sources, such as (i) binary neutron stars or (ii) black hole–neutron star pairs with relatively small black hole mass such that the neutron star may be disrupted, from (i) BBHs or (ii) black hole–neutron star pairs with larger black hole mass such that the neutron star plunges into the black hole without disruption. For GW transient candidates other than binary mergers, the observed duration, characteristic frequency, and total GW fluence will be available and can help constrain the possible source types and distances.

GW candidates are currently identified with a latency of ~1 min. Following this initial detection, candidates undergo manual checks (by humans) to ensure data quality, which introduces an additional ~30 min delay. For some follow-up observatories, such as CTA, humans in the loop introduce too long a delay, given the expected short duration of high-energy emission, and these observatories will therefore need to rely on the earliest available reconstruction.

4 THE CHERENKOV TELESCOPE ARRAY AND OBSERVATION STRATEGIES

In terms of raw sensitivity, CTA can detect GRBs to very high redshifts as long as the GRB spectrum continues out to very high energies when extrapolating from the observed Fermi-LAT spectrum (e.g. Inoue et al. 2013a). However, the observed spectra at very high energies will be attenuated by the interaction of source photons with the EBL (Hauser & Dwek 2001; Domínguez et al. very high energies will be attenuated by the interaction of source photons with the EBL (Hauser & Dwek 2001; Domínguez et al.

4.1 CTA telescopes

The CTA observatory is being designed by an international consortium, which is currently building prototypes and characterizing them. To provide all-sky access, CTA will comprise two arrays, with one deployed in the Northern hemisphere, on La Palma (Spain), while Paranal (Chile) is the site in the Southern hemisphere. Meeting the ambitious CTA design goals, including an overall increase in sensitivity of about an order of magnitude compared with the current generation of IACTs (Acharya et al. 2017), requires a large number of telescopes of different sizes in order to cover the energy range from 20 GeV up to above 100 TeV. The telescopes are grouped in three sizes; the large-sized (23 m diameter LSTs), medium-sized (12 m MSTs), and small-sized (4 m SSTs). A prototype of a dual-mirror version of the MST [Schwarzschild-Couder Telescope (SCT) with a 9.7-m primary mirror] is also being built. The LSTs provide access to the low-energy range (≤0.1 TeV), the SSTs to the high-energy range (>10 TeV) while the MSTs ensure enhanced sensitivity in the core energy range of CTA (0.1–10 TeV). The telescopes will be arranged on the ground such that the LSTs are grouped together, aiming to sample a substantial fraction (≈10 per cent) of the Cherenkov light pool, surrounded by an array of MSTs, that will ensure an excellent shower reconstruction due to a large stereoscopic multiplicity, and a more numerous collection of SSTs that will extend the array footprint, thus increasing the effective area of the instrument in a domain where event statistics are the main limiting factor. Going from the lowest to the highest energies, CTA will provide an angular resolution from ∼0.25′ to ∼0.03′ and a field of view from ∼5° to ∼10° (Acharya et al. 2017). However, the sensitivity across the field of view is not uniform, particularly below 100 GeV, where the sensitivity is dominated by the LSTs and is ∼50 percent lower at 1.5′ off-axis than it is on axis (Maier et al. 2017). At the lowest energies, where there is some overlap of CTA and Fermi-LAT, the former gains the most in sensitivity on short time-scales compared to the latter (Funk et al. 2013), which on the other hand has a greater sky coverage. Thus, slew speed is important, and the LSTs are designed to slew to any point in the sky within 20 s, while the MSTs can slew to any point in 90 s.

4.2 Array deployment

The array baseline design proposed to provide the required sensitivity and energy range is for 4 LSTs and 15 MSTs in the Northern hemisphere and 4 LSTs, 25 MSTs, and 70 SSTs in the Southern hemisphere. The SSTs will be deployed only in the South to enhance observations of Galactic plane sources. Concerning the deployment schedule, the pre-production phase calls for the first CTA telescopes to be installed in 2019, which will include a prototype LST on La Palma, at the site where the two Major Atmospheric Gamma-ray Imaging Cherenkov (MAGIC) telescopes are currently hosted (Aleksić et al. 2016). The production phase follows until full commissioning and science verification is completed. The field of view of a given size telescope does not vary significantly, but sensitivity and angular resolution do improve as more telescopes are added into the array. Transient observations will start early on during construction (in principle with only a single telescope in operation), but initially, due to the current LST deployment schedule, the lowest energy range of CTA will be available only from the North.

4.3 Transient Follow-up

GW transients are proposed as the highest priority target to be investigated by CTA with a rapid and coordinated response within CTA’s Key Science Project on transients. During its early phase, before array completion, a maximum of 20 h yr−1 of observation time is foreseen for each CTA site to follow up GW transients, reduced to 5–10 h yr−1 in the subsequent years with the full array (Acharya et al. 2017). CTA has some unique capabilities for GW follow-up, including a large instantaneous field of view and access to both hemispheres (Bartos et al. 2014). As GW alerts can have large associated localization regions, both the northern and southern CTA arrays may have to be triggered to ensure coverage. The field of view of the CTA telescopes is sufficiently large that some galaxies within the Advanced LIGO/VIRGO horizon should always be within view and hence targeting could be based only on the localization probability map via a set of pointing directions (filing).

---

1 https://www.cta-observatory.org
In Bartos et al. (2014) it was demonstrated that CTA will be capable of following up GW transients associated with short GRBs inside the horizon distance of Advanced LIGO/Virgo over a sky area as large as $\sim 1000$ deg$^2$. Considering here the detectability of a transient gamma-ray source with unknown direction (survey mode) over an area of $\sim 100$ deg$^2$, we have to take into account the scaling factor $f_{\text{SE}} \simeq [(\theta_{\text{CTA}}/2)^2/\Omega_{\text{GW}}]^{1/2}$, where $\theta_{\text{CTA}}$ is the diameter of the field of view of CTA and $\Omega_{\text{GW}}$ is the area of the error region of the GW transient. Thus, the fluence threshold is reduced by a factor $\approx 3$ compared to Bartos et al. (2014), increasing the detectability of intrinsically weaker sources, as well as sources with a cut-off at lower energies and/or with a larger delay in the start of observations after the trigger (see table 1 in Bartos et al. 2014) and/or larger zenith angles.

For a GeV gamma-ray source with a temporal decay $\propto t^{-1.38}$, like that observed in the short GRB090510 (De Pasquale et al. 2010), an emission lasting 100 s in an observation starting 100 s after the trigger gives a fluence lower by a factor $\approx 3$ with respect to a duration of 1000 s, as considered in Bartos et al. (2014). Therefore, the improvement in the sky localization of the GW transients from $\sim 1000$ to $\sim 100$ deg$^2$ increases also the detectability of sources with shorter duration, of the order of the observations given in De Pasquale et al. (2010).

Alternative observation modes with CTA, like the so-called ‘divergent’ pointing mode, are under study (Szanoecki et al. 2015). Exploratory simulations of single-dish MST arrays indicate that effective fields of view of $20^\circ$ can be achieved from divergent pointing, offering a better sensitivity than the conventional parallel pointing when scanning large portions of the sky (at the expense of energy and angular resolutions). If proved efficient, this observation mode could be employed for the follow-up of GW transients, reducing the time required to tile a large localization region at a given sensitivity. If useful for surveys, this operational mode would also increase the probability of a joint prompt detection for short-duration transients. For a given target sensitivity, a further reduction to the time required to survey the localization region may come from the better off-axis performance provided by the SCTs as compared to single-dished MSTs (Hassan et al. 2015), at the price of a slightly increased energy threshold.

The real-time analysis (RTA) pipeline will automatically determine if a new source has been detected and issue an alert within 30 s from the triggering event collection, ensuring fast communication with the astrophysics community (e.g. with Virtual Observatory Events; Fioretti et al. 2015). On short time-scales (hundreds of seconds), CTA is unlikely to detect a new steady source, but in any case CTA follow-up observations will occur for any new source detected. The CTA design requires that the sensitivity of the RTA search for transients (on multiple time-scales) should not be worse than three times the nominal CTA performance. Assuming a factor three, the RTA with the southern full array will achieve in a few hours of observations and with 30 s of reconstruction and processing the same sensitivity of 50 h of integration of the current Cherenkov telescopes. As a reference value, the Southern CTA array will be able to detect 10 percent of the Crab nebula integral flux with 1000 s of pointed observation, for an energy threshold less than 10 TeV (Fioretti et al. 2016). However, there is still space for improvements in the algorithms and hardware, and thus in the decrease of the minimum detectable flux with the RTA.

The CTA Consortium will receive GW alerts from the Advanced LIGO/Virgo interferometers as stipulated in a memorandum of understanding already signed, and will follow-up those during dark time with zenith angles less than 70$^\circ$ for 2 h each, adding exposure time in case of positive detections (Acharya et al. 2017).

The duty cycle of current-generation IACTs is affected, among other factors, by the lunar phase, which prevents observations during full Moon due to the elevated brightness of the sky, thus potentially reducing the overlap between Advanced LIGO/Virgo and present IACT uptimes. However, observations are routinely performed under moderate moonlight, representing a $\sim 30$ per cent increase over an average, dark-sky only, observing year (e.g. Archambault et al. 2017). Silicon photo-multiplier-based IACT cameras have proven to be effective in the detection of cosmic showers under bright moonlight conditions without risking the photodetectors’ integrity or accelerating their aging (Biland et al. 2014), allowing for an increased duty cycle, although with reduced sensitivity and larger energy threshold. This technological advancement will be utilized in the SCT camera (Otte et al. 2015), as well as in the SST cameras (Montaruli et al. 2015), therefore opening the possibility of following up GW alerts even during bright moonlight conditions.

The closeness of the MAGIC telescopes and the prototype LST may give the opportunity, if they will be operated in coincidence, to start carrying out a follow-up of GW transients at the CTA Northern site with a system of three large and fast slewing Cherenkov telescopes in stereoscopic mode.

5 JOINT SEARCH METHODOLOGY

5.1 Previous search strategies with Cherenkov telescopes

The current-generation IACTs – H.E.S.S., MAGIC, and VERITAS – have been used to perform searches of very high-energy gamma-ray emission associated with GW triggers. The cameras of these IACTs cover, with radially dependent sensitivity, an area in the sky with a size between $\sim 8$ and $\sim 20$ deg$^2$, making them suitable to survey a fraction of the localization uncertainty regions of LIGO and LIGO/Virgo events. We briefly discuss results from these searches as they provide a learning experience for future follow-up observations using CTA.

During the O1 run, MAGIC performed follow-up observations (Carosi et al. 2017) of the event GW151226, later identified as due to a BBH merger (Abbott et al. 2016f). The event was detected on 2015 December 26 UT and a localization map was circulated by LIGO the following day, based on which four MAGIC pointings were manually selected maximizing the probability coverage and taking into account visibility, overlap with existing catalogues, and observations of other telescopes. The four positions were observed starting on December 28 UT with an average exposure of 42 min per pointing. No source was detected during these observations.

The first VHE follow-up during the O2 run was performed by VERITAS (Santander 2016) for the event GW170104 (Abbott et al. 2017a) detected on 2017 January 4 UT which was due to the coalescence of a 50-solar-mass BBH system at a redshift of 0.2. VERITAS opted for tiling the Northern fraction of the localization map above a 50$^\circ$ elevation using 39 consecutive pointings each observed for approximately 5 min. The survey started on January 5 UT and covered 27 per cent of the event containment probability. Although the presence of clouds affected observations, VERITAS reports that these observations were sensitive to sources with a flux greater than 50 per cent of the Crab nebula above 100 GeV.$^3$

$^2$ https://gcn.gsfc.nasa.gov/gcn3/18776.gcn3
$^3$ https://gcn.gsfc.nasa.gov/gcn3/21153.gcn3
The first detection of GWs from binary neutron stars (GW170817) took place during O2. IACT observations were started by H.E.S.S. 5.3 h after the detection of the event using an observational strategy that identified regions of high probability to find a GW counterpart. The first of these observed regions included the location of SSS17a, the EM counterpart for GW170817 identified later in the optical range. Two algorithms developed by H.E.S.S. optimized for real-time GW follow-up and offline scheduling are detailed in Seglar-Arroyo et al. (2017) which include folding the localization maps for the GW events with a galaxy catalogue, and the prioritization of different targets according to their distribution in the sky and observational constraints. For this follow-up, observations were started on August 17–18 UT and continued over several days, setting upper limits in the energy band between 0.28 and 8.55 TeV as no gamma-ray excess was identified in the observed region (Abdalla et al. 2017).

5.2 LIGO/Virgo alerts and follow-up

During the O2 observing period, LIGO and Virgo sent out alerts typically within about 30 min of the identification of an interesting event; most of the time taken to generate alerts was due to human vetting of identified events; as a case study of the alert process, we briefly summarize the steps leading up to the publication of the GCN alert that notified other observatories of the detection and likely source direction. We use the binary neutron star event GW170817 as our case.

LIGO recorded the GW event in data from its Hanford detector with a low-latency search 6 min after the merger (Abbott et al. 2017b,c). LIGO-Livingston data were initially not used due to a noise artefact overlapping with the signal, which was later removed. Virgo also recorded the signal, but its low-latency data transfer was delayed. LIGO and Virgo sent a GCN alert about 35 min after the BNS event was registered, corresponding to about 40 min after the BNS merger. A GW skymap using information from both LIGO detectors and Virgo was distributed five hours after the merger. Compared to the Fermi-GBM skymap for GRB170817A, this skymap identified a smaller region of the sky and led to the discovery of the optical transient and host galaxy identification. For comparison, the Fermi-GBM detection and initial skymap were distributed 14 s after detection (Goldstein et al. 2017).

We see that, while the analysis and the dissemination of information from GW170817 was too long for practical use by high-energy observatories like CTA, most delay occurred due to human involvement and technical difficulties that can be overcome for future detections.

Looking at the end result, the reconstructed 90 per cent CL skymap of GW170817 was only ∼30 deg². This localization and the source distance are much more favourable than the more conservative case discussed by Bartos et al. (2014), making the detection prospects of CTA promising, given that the delay due to GW data analysis is comparable to the 6 min achieved for GW170817. Note that, for the particular case of GW170817 in which detection was established early but the GW skymap was significantly delayed, it can also be feasible to take the fact of GW observation and the localization from the corresponding GRB, and scan the corresponding sky area (in fact this is what happened with GW170817 – observers scanned the Fermi localization region until an improved GW localization area became available).

More generally, GW detector networks can typically give larger sky localization regions (see Singer et al. 2014 for details). In these cases, CTA will need to cover as much of the GW sky localization region as possible. There have been multiple studies discussing optimal strategies for covering the broad GW sky localization regions and maximizing the probability of detecting transient counterparts (e.g. Coughlin & Stubbs 2016; Ghosh et al. 2016; Chan et al. 2017; Salafia et al. 2017). All of these studies highlight that the probability of detecting a counterpart transient can be boosted by factors of a few if the sky location is tiled efficiently and the time allocation per sky location is optimized for each telescope’s sensitivity and the skymap probability.

Additionally, the probability of counterpart detection can be boosted if observing strategies target galaxies within the GW sky localization region (Gehrels et al. 2016; Singer et al. 2016; Chan et al. 2017). A good example for this strategy is that followed by the Swope Telescope which was the first to discover the optical counterpart of GW170817 despite its small telescope size (Coulter et al. 2017).

It is relevant to note here that rapid skymaps generated for interesting GW events do not incorporate calibration uncertainties in GW detectors. Such calibrations are generated only after several days. We also note that the probability assigned to different regions in GW skymaps can change as the skymap is refined (see e.g. Abbott et al. 2016a).

An interesting future possibility is the first detection of GWs from a phenomenon other than compact binary mergers. For such a case, due to limited signal models, there could be additional uncertainties associated with the skymap. At the same time, such sources are expected to be detectable only at smaller distances than compact binary sources, making their potential observable very high-energy emission brighter and more detectable. Despite the large size of GW sky localization regions, CTA will typically have the capacity to cover the GW sky region even for these cases.

Since for most foreseeable cases, CTA will be able to cover the reconstructed GW skymap, and since these skymaps are typically generated at 90 per cent confidence level, upon non-detection it will be beneficial to allocate remaining observation resources to observe regions around the estimated GW sky localization, as contingency against variations in the GW sky localization due to factors laid out above.

6 CONCLUSION AND OUTLOOK

We reviewed the current status of GW-multimessenger observations and that of CTA, in order to examine the possible strategies of searching for very high-energy emission from GW transients with CTA. Our main goals were to present a summary to the GW and CTA communities of progress on the other side, as well as to identify what needs to be done before joint observations commence. This work was a continuation of Bartos et al. (2014), where we explored the utility of CTA for GW follow-up observations, and found that CTA is well-suited to work with even large localization uncertainties, often expected for GW signals.

Based on the present status of observatories and multimessenger observations, we consider the following directions to be important to maximize the GW-follow-up potential of CTA:

(i) Low-latency alerts: With the rapid fading of high-energy emission in the aftermath of a binary merger, it is critical that GW detections are shared with low latency with partner observatories. This requires full automation on the GW side. For CTA, a higher false alarm rate is tolerable as the follow-up of GW triggers only requires O(1000) s of observation time. Additionally, automation in the reception and response to the alert by CTA, and in the
execution of the follow-up observation, is also necessary to facilitate an optimal outcome.

(ii) **Start with a single CTA telescope:** If gamma-ray emission from short GRBs extends to $E > 20$ GeV, CTA may be able to detect such emission even with a single LST on ~100 Mpc distance scales relevant for GW observations. The commissioning of an automated alert system and the corresponding CTA follow-up observation execution should begin as soon as a single CTA telescope is deployed.

(iii) **No need for galaxy catalogues:** While galaxy catalogues played an important role in the follow-up of GW170817, their utility for CTA follow-ups will be limited. Since CTA will have a large multi-deg$^2$ field of view, albeit with some sensitivity degradation off-axis, a non-uniform galaxy distribution will rarely impact the prioritization of pointing directions. Additionally, the very high-energy sky has few transients, therefore galaxy catalogues are not needed to reduce the false-alarm rate.

(iv) **Most GW candidates can be followed up:** With CTA’s dedicated GW-follow-up observing time of ~10 h yr$^{-1}$, considering an observation time of 1000 s per event (Bartos et al. 2014), we expect that CTA will be able to follow-up all GW candidates other than BBH mergers that fall within the region of the sky accessible to CTA, given the expected rate of non-BBH GW candidates of one per month, and a one-per-month false alarm rate. This means that no prioritization is needed based on the properties of the non-BBH GW candidates, allowing the potential for the discovery of unusual sources. For BBH mergers, the detection rate could be several hundred per year once LIGO/Virgo reach their design sensitivity (Abbott et al. 2016a,b). Such a high rate will be unfeasible to comprehensively follow up and some prioritization will be needed.

(v) **Deeper observation of promising events:** Some of the available observing time should be utilized for a deeper observation of a promising GW event, preferably extending CTA’s original observation, in so far as this is technically viable, rather than observing the region of interest in successive nights. Such an extension can be motivated by insight in the nature of the event. For instance, such an event can be a binary neutron star merger whose reconstructed parameters indicate that it is nearby and its orbital axis is roughly pointing towards Earth, or even with an observed GRB counterpart.

For such an event, if an initial scan with CTA finds no very high-energy emission, it is worth integrating for longer in order to enable probing the high-energy cut-off of gamma-ray emission from the event. A complementary motivation for a deeper observation is the prompt finding of a hint of signal in CTA’s RTA.

(vi) **Multi-messenger follow-up:** Cosmic messengers that are promptly emitted from GW sources and that are monitored by ‘all-sky’ detectors can be rapidly available along with GWs. Such messengers include gamma-rays and high-energy neutrinos. It will be useful to plan CTA observations such that the low-latency detection of a GRB counterpart, or high-energy neutrinos, from a GW source can be incorporated in the follow-up. In particular, these other messengers can significantly improve the localization of the source. For instance, high-energy neutrino track events can be reconstructed to sub-degree precision (Albert et al. 2017b).

(vii) **Multi-messenger alert:** Once CTA identifies very high-energy emission from a GW source, its precise direction reconstruction ($\lesssim 0.1^\circ$; Bernlöhr et al. 2013; Acharya et al. 2017) can be used to point other follow-up observatories in the right source direction. It is important that such identification is communicated to partner observatories as soon as possible. For example, X-ray emission may rapidly fade similarly to very high-energy emission, and with the narrow fields of view of the most sensitive current instruments it will be beneficial to learn the true source direction quickly.

**ACKNOWLEDGEMENTS**

The authors thank Giulia Stratta for useful discussions. This paper has gone through internal review and has been approved for publication by the LIGO Scientific Collaboration, the Virgo Collaboration, and the CTA Consortium. IB, SM, ZM, and TBH are thankful for the generous support of Columbia University in the City of New York. IB, ZM, and SM are thankful for the generous support of the National Science Foundation under cooperative agreement PHY-1708028. IB is thankful for the generous support of the University of Florida. TBH is thankful for the generous support of the National Science Foundation under cooperative agreement PHY-1352567. DN is thankful for the generous support of the Spanish Ministry of Economy, Industry and Competitiveness, and the European Regional Development Fund through the grant FPA2015-73913-JIN. TDG is thankful for the kind hospitality of Columbia University in the City of New York.

**REFERENCES**

Aasi J. et al., 2014, Phys. Rev. D, 89, 102006

[Other references...]
