The discovery of binary neutron star merger GW170817 from its gravitational wave signature, together with its accompanying electromagnetic emission spanning gamma-ray to radio, marked the birth of GW+EM multi-messenger astrophysics. The radioactively-powered thermal kilonova, which dominated the ultraviolet to infrared in the hours to weeks after the merger, indicates that such mergers are the site of heavy element nucleosynthesis, likely extending to the third r-process peak. The prompt gamma-ray flash, and late time non-thermal (X-ray to radio) emission, indicate that the merger also produced an ultra-relativistic jet, thus tying this event to the phenomena of short-duration gamma-ray bursts. In the future, observations of further mergers promise to establish their contribution to global nucleosynthesis, allow investigation of jet launching and structure, provide independent estimates of the cosmological parameters, constrain the neutron star equation of state, and address questions of fundamental physics.

KEYWORDS:
Gravitational waves, gamma rays:bursts, stars:neutron, nucleosynthesis

1 | INTRODUCTION

The primary targets of the current generation of ground gravitational wave (GW) detectors are tight binaries consisting of either neutron stars (NS) or stellar mass black holes (BH), which eventually in-spiral and coalesce. Systems in which at least one component is a neutron star represent a scientific “Rosetta Stone”, lying as they do at the intersection of several long-standing problems in astrophysics. Specifically: (1) they have been hypothesised to generate both bright electromagnetic signals and gravitational waves (e.g. Nissanke, Kasliwal, & Georgieva 2013); (2) they have long been favoured as the likely progenitors of so-called “short-duration” gamma-ray bursts (short-GRBs; Nakar 2007); (3) material ejected from the neutron star during the merger has been proposed as the dominant site for the production of heavy r-process elements in the universe (Freiburghaus, Rosswog, & Thielemann, 1999; Lattimer & Schramm, 1974, 1976), whose origin is otherwise hard to explain. Furthermore, detection of such systems can be used to determine the parameters of cosmology (Abbott et al., 2017b; Schutz, 1986), make tests of fundamental physics, such as the relative speed of gravity and light (Abbott et al., 2017a), and constrain the neutron star equation of state (e.g. Bauswein, Just, Janka, & Stergioulas, 2017; Coughlin et al., 2018).

The discovery of the binary neutron star merger GW170817 from its gravitational wave (GW) emission by Advanced-LIGO/Virgo, together with its accompanying electromagnetic (EM) signals, heralded the birth of GW+EM multi-messenger astrophysics (Abbott et al., 2017d). This is providing remarkable new opportunities for observers working across the EM spectrum.

Abbreviations: GW, gravitational waves; GRB, gamma-ray burst; HST, Hubble Space Telescope; NS, Neutron Star; BH, Black Hole; LIGO, Laser Interferometer Gravitational-Wave Observatory; EM, Electromagnetic; KN, kilonova/-macronova; GBM, Gamma-ray Burst Monitor; JWST, James Webb Space Telescope; SVOM, Stellar Variable Optical Monitor
GW170817 was the first system detected in gravitational waves which was unambiguously from a binary neutron star merger (Abbott et al., 2017c). The richness of the EM phenomena it exhibited, and the lessons learnt, illustrate much of the promise of the field moving forward. Arguably the most surprising aspect of the event was that the GW in-spiral was followed less than 2 s later by a short gamma-ray burst signal detected by Fermi/GBM and INTEGRAL (Abbott et al., 2017a). This was unexpected since, while binary neutron star mergers have long been suspected to be the progenitors of short-GRBs, the collimation of GRB jets should mean that our line of sight usually is not within the initial beam of the jet. However, the short-GRB 170817A was also peculiar in the sense of being several orders of magnitude fainter, intrinsically, than previously observed short-GRBs, and it seems likely that in this case it was not due to emission from the jet itself (i.e. primarily via an internal shock mechanism), but instead was from the breakout of a shock created by the interaction of the jet with an envelope of existing matter ejected during the process of the merger. This interaction should create a shocked cocoon of hot material expanding at mildly relativistic velocities, which then may create gamma-rays as the shock escapes the slower dynamical ejecta (see Fig. 1; Gottlieb, Nakar, Piran, & Hotokezaka 2018; Lazatti, Deich, Morsony, & Workman 2017). This origin is supported by a softening of the gamma-ray spectrum over the course of the emission, to a more thermal spectrum (Goldstein et al., 2017).

GW170817 was the second binary merger to be observed with the 3 detector GW network, thanks to which it benefited from a relatively small GW positional error region of only ∼ 30 sq-deg on the sky. It was also detected with sufficient signal-to-noise to provide a comparatively precise estimate of the distance (to ∼ 25%), which further constrained the required volume to be searched for EM counterparts (note, the Fermi and INTEGRAL gamma-ray localisation in this case was much poorer, although if Swift had happened to be pointing in the right direction the localisation would have been immediately good to a few arc-min). Due to its position on the sky being southerly and unfortunately close to the Sun, the first deep searches of this error region for an EM counterpart at other wavelengths did not take place until more than 10 hr after the event. Remarkably, within a short space of time multiple observers had independently located a counterpart in the galaxy NGC 4993 at ∼40 Mpc, which was initially rising in brightness and spanned the UV to optical to near-IR (Abbott et al., 2017d).

This emission was monitored over the course of the next several weeks, and exhibited a rapid and marked evolution from blue to red as it began to fade (Fig. 2). Such behaviour is the predicted signature of a radioactively powered kilonova/macronova (KN) arising from the decay of elements synthesised in the ejecta from the neutron star (e.g. Kasen, Badnell, & Barnes, 2013; Kasen, Fernández, & Metzger 2015; B. D. Metzger & Fernández, 2014), but it is unlike any previously known extragalactic transient (Coulter et al., 2017; Cowperthwaite et al., 2017; Kasliwal et al., 2017; Pian et al., 2017; Smartt et al., 2017). In qualitative terms, the colour changes are understood as differing ejecta components coming to dominate the light. In detail, though, there are surprises, particularly the initial brightness in the UV and blue. This seems to require a significant component of fast-moving (∼ 0.3c) dynamical ejecta in the polar direction being produced with a relatively high electron fraction (e.g. Evans et al., 2017). That in turn has been widely interpreted as the signature of weak interactions, enhanced by shock heating and neutrino irradiation from a putative hyper-massive neutron star (HMNS), driving down the neutron fraction in the collisional component (e.g. Sekiguchi, Kiuchi, Kyutoku, Shibata, & Taniguchi, 2016). As the kilonova evolves, the emission comes to be dominated by lower electron fraction, likely lanthanide-rich, ejecta, for which strong line-blanketing of the optical results in the emission appearing primarily longward of 1.5 microns (e.g. Kasen, Metzger, Barnes, Quataert, & Ramirez-Ruiz, 2017; Rosswog et al., 2018). The energy here indicates a comparatively large amount of material (few% of M), presumed to indicate a significant contribution from an accretion disk wind given off after collapse of the HMNS, in addition to the mainly equatorial tidal ejecta (e.g. B. D. Metzger, 2017).

The proximity of the event allowed detailed studies of the host galaxy, finding it to be a disturbed early-type galaxy with some modest star formation, and structure indicative of a merger event in the past few hundred Myr (Levan et al., 2017). It also allowed independent electromagnetic distance estimates to be made for the host, thus testing and helping break degeneracies in the GW analysis (e.g. Cantiello et al., 2018; Hjorth et al., 2017; Im et al., 2017).

2.1 R-process nucleosynthesis

It was realised in the 1950s that many of the elements in the universe heavier than iron, including all the naturally occurring actinides and with notable peaks around xenon (Z=54, A=131) and platinum (Z=78, A=195), exist in the universe in too great abundances to have been formed by the relatively well understood process of slow neutron capture (the s-process, which is thought to occur predominantly towards the end of He and C burning in massive stars, Burbidge, Burbidge, Fowler, & Hoyle, 1957). These elements appear instead to require an environment where neutron captures occur significantly more rapidly than beta decays, leading initially to a distribution far
from the valley of beta stability, closer to the neutron drip-line. Nuclides tend to pile up at closed-shell neutron “magic numbers” (N=50, 82, 126) and subsequent beta decays then increase their final atomic numbers (at the same atomic mass). Given sufficient neutron density, this can naturally produce r-process peaks of stable (and long-lived) isotopes around the locations seen in the Solar System abundance distribution.

The presumed site for the r-process was in core-collapse supernovae (CCSN), although much subsequent work has cast doubt on whether the necessary conditions exist in CCSN to create the heavier r-process elements in particular (Arcones & Thielemann 2013; Martinez-Pinedo, Fischer, & Huther 2014). However, a generic prediction of binary neutron star (BNS) and neutron-star black-hole (NSBH) merger models is that some neutron star material should be expelled (from tidal disruption, debris from the collision itself, and material unbound or blown off from a subsequent accretion disk), and as early as 1974 it was pointed out that as it rapidly decompressed such material would naturally have an initially high density of neutrons and very low electron fraction, $Y_e$, and so provide an alternative site for r-process production (Lattimer & Schramm 1974). Recent detailed calculations have found that a good match to the heavier r-process abundance peaks is a robust prediction of nucleosynthesis within such ejecta (e.g. Just, Bauswein, Ardevol Pulpillo, Goriely, & Janka 2015; Korobkin, Rosswog, Arcones, & Winteler 2012; Mendoza-Temis et al. 2015).

On the observational side, there is some evidence pointing to a rare origin for the heavy r-process elements in particular (Thielemann et al. 2011). For example, 244Pu, a radioactive actinide nuclide with half-life ~80 Myr, has been found in low relative abundance in deep-sea floor samples, suggesting its abundance in the interstellar medium is well below the quasi-equilibrium value expected if it had been regularly replenished by core-collapse supernova enrichment (Hotokezaka, Piran, & Paul 2015; Wallner et al. 2015). The abundance of europium in very low metallicity stars seems also incompatible with
supernova enrichment (Macias & Ramirez-Ruiz 2018; Van- gioni, Goriely, Daigne, François, & Belczynski 2016). The obvious candidate for a rare enrichment source is the merger of compact binaries.

A crucial consequence of such a nucleosynthetic pathway is that the rapid radioactive decay of the initial highly neutron rich, unstable isotopes should itself be a source of heating that would power a short-lived transient event, similar to a faint supernova (Li & Paczyński 1998; B. D. Metzger et al. 2010). Searching for such emission accompanying low redshift short-GRBs is challenging due to contamination by afterglow light, and early attempts, which concentrated on optical wavebands, produced no confident detections (although see e.g. Jin et al. 2016 and refs therein for instances of excess I-band light in re-analyses of two old datasets, which could be kilonova components). A major development in the field came with our observation of nIR excess at ∼1 week in the rest-frame following GRB 130603B, consistent with emission from an accompanying kilonova. Thus, this observation tied the two phenomena together, giving a window on the nucleosynthesis of the heavy (r-process) elements in the Universe and offering an isotropic electromagnetic signature of systems for which direct detections of gravitational wave (GW) radiation were anticipated (Berger, Fong, & Chornock 2013; Tanvir et al. 2013).

The situation changed in 2013 with the realisation that if any ejecta do produce elements up to the 3rd r-process peak, then the enormous number of electronic transitions available for the newly synthesized lanthanide and actinide species, in particular, should result in very high line-blanketing. This would render such kilonova sources very faint in the optical band, with the light largely appearing instead in the near-infrared on time-scales of several days to weeks (e.g. Barnes & Kasen 2013; Grossman, Korobkin, Rosswog, & Piran 2014; Kasen et al. 2013; Tanaka & Hotokezaka 2013) consistent with the observations of short-GRB 130603B (see below).

On the other hand, it was also hypothesised (and now confirmed with the advent of GW170817) that some ejecta components may acquire a higher electron fraction due to shock heating of collisional ejecta combined with neutrino irradiation from a remnant neutron star (e.g. B. D. Metzger & Fernández 2014). This would lead to nucleosynthesis of only light r-process elements, and consequently produce a bluer component of kilonova light in the first few hours and days, at least if observed along a sight-line with no obscuring lanthanide-rich material (see B. D. Metzger 2017 for review). Distinguishing and quantifying such differing contributions is therefore of key importance for understanding the overall contribution to chemical enrichment, with kilonovae accompanying low redshift short-GRBs, with precise modelling of the afterglow, providing further targets.

3 CONNECTION TO SHORT-DURATION GAMMA-RAY BURSTS

The nature of gamma-ray bursts has been one of the outstanding astrophysical problems of the past several decades. As long as they were observed only in gamma-rays even their distance scale remained highly uncertain, but observations of GRB afterglows at lower energies (X-ray, optical, IR, radio) finally established that they are at cosmological distances and, at peak, the brightest GRBs are by far the most luminous sources known (e.g. M. R. Metzger et al. 1997; van Paradijs et al. 1997). It is widely thought that they are produced by ultrarelativistic, collimated jets with bulk Lorentz factors ∼300 (e.g. Piran 2004).

The population of gamma-ray bursts is known to split into (at least) two sub-classes, distinguished primarily by the duration of the initial prompt gamma-ray flash, with the dividing line being around 2 s. There is a large body of evidence that the so-called “long-duration” class is produced during the core collapse of some rare massive stars, as material accretes onto a newly formed black-hole or proto-neutron-star (e.g. Woosley & Bloom 2006). The “short-duration” class are rarer and typically have much fainter afterglows, the first not being identified until 2005 (e.g. Gehrels et al. 2005), meaning that progress in their study has been much slower. It has long been supposed that they may be the result of jets produced in the final moments before the merger of BNS or NSBH systems (e.g. Nakar 2007). This hypothesis is supported by the finding that short-GRBs are associated with a broader range of host galaxies than the long class, including some with only old stellar populations (e.g. Fong et al. 2013), and also that some short-GRBs appear to be located away from the optical light of their hosts, consistent with their progenitor having a high natal kick from the supernovae that produced the compact remnants (e.g. Tunnicliiffe et al. 2014).

At first sight, the coincidence of a gamma-ray flash with the GW170817 merger event would seem to cement the association of short-GRBs with compact object mergers. However, the fact that it was intrinsically so faint in the gamma-rays led to the suggestion that the gamma-rays here were more likely produced by shock break-out from a mildly relativistic cocoon rather than a jet itself, and indeed that, whilst a jet must have been launched in order to create a cocoon, it may not have escaped the ejecta.

Then, even after the kilonova light faded, GW170817 performed an encore, rising in brightness again in radio, optical and X-ray (see Fig. 3), explained as emission from the afterglow of a relativistic component of the ejecta, most likely some combination of emission from the mildly relativistic cocoon and a higher velocity jet seen at some ∼20–30 degrees off-axis (e.g. Alexander et al. 2018; D’Avanzo et al. 2018; Granot,
4 CONCLUSIONS AND FUTURE OUTLOOK

The “Advanced” generation LIGO detectors made the historic discovery of the first gravitational wave source, a binary black-hole coalescence event, just before the official start of their first science run in September 2015 (Abbott et al. 2016), and to date have reported six events in total. In the coming years ongoing upgrades of both LIGO and Virgo will increase the sensitivity, and hence horizon, further. At the same time, additional detectors, beginning with the Japanese KAGRA detector, and ultimately LIGO-India, will add significant extra capability. Thus we can expect many more opportunities to study the electromagnetic counterparts of binary neutron star mergers, and likely also black-hole neutron star mergers in the coming years. This will provide considerable insight into the diversity of heavy element nucleosynthetic yields of such events from UV/optical/IR, as well as the nature of the ultra-relativistic jets that they launch making use of observations from X-ray to radio. They will also provide further tests of fundamental physics and determinations of cosmological parameters. Another long-term goal is the detection of gravitational waves from so-called “burst” sources, which may arise, for example, from asymmetric supernova explosions in relatively nearby galaxies (e.g. Andersson et al. 2013).

Studies of short-GRBs will continue to be important for tying together the local and cosmological populations of mergers (e.g. Fong et al. 2017, Gompertz et al. 2018). In particular, this will help establish the r-process budget over cosmic time. From 2021, the launch of the new Chinese-French-led satellite, SVOM, increasing the rate of well-localised short-GRBs, together with the advent of JWST, which will allow extremely sensitive imaging and spectroscopy of associated kilonovae, potentially to $z \sim 1$.

ACKNOWLEDGMENTS

I acknowledge funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement no 725246).

Financial disclosure

None reported.

Conflict of interest

The author declares no potential conflict of interests.

REFERENCES

Hotokezaka, K., Piran, T., & Paul, M. 2015, Nature Physics, 11, 1042.
FIGURE 3 Late time photometry (triangular symbols are upper limits) of the non-thermal emission due to the relativistic ejecta component(s), as reported in Lyman et al. (2018) and references therein. [Left:] The X-ray and radio light curves of GW170817, and optical detection at ~100 days, together with a range of off-axis structured jet models. [Right:] the spectral energy distributions at 16 and 110 days, showing that there was little change in the spectrum of the non-thermal component between those times, and that the radio, optical and X-ray fit on a single power law.

Wallner, A., Faestermann, T., Feige, J. et al. 2015, Nature Communications, 6, 5956.