ELECTROMAGNETIC SIGNALS FOLLOWING STELLAR–MASS BLACK HOLE MERGERS

S. E. de Mink$^{1,2}$ & A. King$^{3,4}$

$^1$Anton Pannekoek Institute for Astronomy, University of Amsterdam, 1090 GE Amsterdam, The Netherlands
$^2$Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA
$^3$Theoretical Astrophysics Group, Department of Physics & Astronomy, University of Leicester, Leicester LE1 7RH, UK
$^4$Leiden Observatory, Leiden University, Niels Bohrweg 2, NL-2333 CA Leiden, The Netherlands

ABSTRACT

It is often assumed that gravitational wave (GW) events resulting from the merger of stellar–mass black holes are unlikely to produce electromagnetic (EM) counterparts. We point out that the progenitor binary has probably shed a mass $\gtrsim 10 M_\odot$ during its prior evolution. If a tiny fraction of this gas is retained until the merger, the recoil and sudden mass loss of the merged black hole shocks and heats it within hours of the GW event. Whether the resulting EM emission is detectable is uncertain. The optical depth through the disk is likely to be high enough that the prompt emission consists only of photons from its optically thin skin, while the majority may take years to emerge. However, if some mechanism can release more photons in a time comparable to the few–hour energy production time, the peak luminosity of the EM signal could be detectable. For a disk retaining only $\sim 10^{-3}$ of the mass shed in the earlier binary evolution, medium–energy X–rays to infrared emission would be observable hours after the GW event for source distances $\sim 500$ Mpc. Events like this may have already been observed, but ascribed to unidentified active galactic nuclei. Improved sky–localization should eventually allow identification based on spatial coincidence. A detection would provide unique constraints on formation scenarios and potentially offer tests of strong–field general relativity. Accordingly we argue that the high scientific payoff of an EM detection fully justifies search campaigns.

Subject headings: gravitational waves — black hole physics — binaries: close — X-rays: general

1. INTRODUCTION

The direct detection of gravitational waves from binary black hole mergers GW150914, GW151224, and a possible third event LVT151012, has drawn wide attention (Abbott et al. 2016a,b). The general anticipation that the first events would involve neutron stars rather than black holes mobilized massive coordinated campaigns to search for accompanying electromagnetic (EM) counterparts (e.g. Abbott et al. 2016c). For merger events involving at least one neutron star, a plethora of accompanying EM signatures is expected at a range of timescales and wavelengths. In contrast, for binary black hole mergers the common consensus has been that no significant EM counterpart is expected, except for "those in highly improbable environments pervaded by large ambient magnetic fields or baryon densities" as Abbott et al. (2016c) state.

Understanding the report of a transient signal detected by the Fermi Gamma-ray Burst Monitor 0.4 seconds after the first event (Connaughton et al. 2016) attracted considerable attention. It encouraged several theoretical speculations for a possible origin of this unanticipated EM signal. But the lack of a corresponding detection by INTEGRAL/SPI-ACS (Savchenko et al. 2016), careful reanalysis of the data (Xiong 2016) and reassessment of the low count statistics (Greiner et al. 2016) all lead to the conclusion that the Fermi trigger is consistent with a background fluctuation, and unlikely to be of astrophysical origin. This is perhaps not surprising: the ultra–prompt nature of the Fermi signal implies an extremely small EM source and probably the near–simultaneous formation of the second black hole, ruling out the usual formation channels. Proposed scenarios include the exotic formation of a binary black hole inside a massive star (Loeb 2016), which has been ruled implausible (Woosley 2016; Dai et al. 2016). Alternatively, the survival of a minidisk surrounding one of the individual black holes activated shortly before the merger (Perna et al. 2016) implies timescales and luminosities inconsistent with the Fermi signal (Kimura et al. 2016). Nevertheless, the question whether binary black hole mergers can have EM counterparts is natural and interesting.

Here we question the common consensus that the typical stellar–mass binary black hole merger is always dark. We consider a simple possibility for EM signals following binary black hole mergers, with a possible delay of hours. This mechanism requires the BH binary to have a circumbinary disk at the time of merger whose mass need only be a very small fraction of that shed as the system evolved. In section 2 we argue that the initial formation of such a disk is natural in the two progenitor formation scenarios for isolated binary evolution, but unlikely if the binary formed through dynamical interactions in a cluster. We discuss the evolution of the disk under viscous and tidal stresses. If any such disk survives until coalescence, it is perturbed both by the GW recoil and by the sudden drop in mass as energy and momentum is radiated away by gravitational waves. In section 3 we use simple analytic scaling arguments, calibrated against earlier detailed simulations for supermassive black hole mergers, to provide crude estimates of the possible luminosity, delay time, duration and approximate spectral
1. Slow orbital decay: \[ t_{\text{gw}} \gg t_{\text{visc}} \]
   resonant torques & viscous spreading

2. Fast orbital decay: \[ t_{\text{gw}} \ll t_{\text{visc}} \]
decoupling of the disk

3. Merger and recoil:
   sudden rearrangement of the disk

**Fig. 1.** Cartoon depicting the evolution of a circumbinary disk.

---

2. CIRCUMBINARY DISK

**Origin** — Circumbinary material is found around stellar binary systems in a wide variety of evolutionary phases. Examples include very young systems where the disk may be a remainder of the formation process (e.g., Rosenfeld et al. 2013; Janson et al. 2016). In more evolved systems, circumbinary disks can be formed by mass loss from one or both stars through the outer Lagrangian points, as seen in recent simulations (e.g., Pejcha et al. 2016; Kuruwita et al. 2016). Observed examples include the subdwarf + Be star system HR 2142 (Peters et al. 2016), the massive system RY Scuti (Smith et al. 2002) and various post asymptotic giant branch stars including the famous Red Rectangle (e.g., Cohen et al. 2004; Hillen et al. 2016). The presence of circumbinary material has also been considered for the microquasar SS 433, which may contain a black hole of about 10\(M_\odot\) (Perez & Blundell 2010; Bowler 2013).

The progenitors of binary black holes shed mass during various stages of their evolution. Fast ejecta, such as mass loss by radiatively driven stellar winds (e.g., Vink et al. 2000) are likely to escape the system. But most mass leaves the system with velocities comparable to or below the orbital velocities, and so stays close to the binary at least initially. The most obvious example is common–envelope ejection (Ivanova et al. 2013). Other examples include eruptive mass–loss episodes (Humphreys & Davidson 1994), material centrifugally shed from rapidly rotating stars after spin–up (Krtička et al. 2011), mass shedding during non–conservative Roche–lobe overflow (de Mink et al. 2009b) and mass loss through a weak fallback supernova (e.g., Fryer 2006; Lovegrove & Woosley 2013). The amount of mass shed and the ejecta velocities are uncertain in each case, but generally expected to amount to many\(M_\odot\). For example in the standard evolutionary path presented by Belczynski et al. (2016) for the first LIGO detection, about 90\(M_\odot\) is lost from the system, about 45\(M_\odot\) of it during the common envelope ejection. Most promising for our picture may be mass shed during the formation of the second black hole in a failed explosion. Mass that remains bound to the system is not exposed to the ionizing radiation and stellar winds of the progenitor stars.

We will argue in the next section that a long–lived disk of less than a percent of a solar mass already gives potentially detectable EM signals. The expected mass ejection in the scenario by Belczynski et al. (2016) is 3–4 orders of magnitude larger, so even a very inefficient mechanism for gas retention could in principle lead to observable signals. Gas retention seems most likely in formation channels for isolated binaries: (i) the classical common envelope scenario, which involves multiple phases of highly non-conservative mass transfer (e.g., Tutukov & Yungelson 1973, 1993; Kalogera et al. 2007; Kinugawa et al. 2016; Belczynski et al. 2016) and the (ii) chemically homogeneous evolution scenario in which two tidally distorted binary stars stay compact as they experience strong internal mixing (de Mink et al. 2009a), which has recently been proposed as a new channel for compact binary black hole formation (Mandel & de Mink 2016; Marchant et al. 2016; de Mink & Mandel 2016). Circumbinary gas may also be expected for binary black hole pairs formed through the (iii) assisted inspiral in the disk of active galactic nuclei (Stone et al. 2016; Bartos et al. 2016). In contrast, we do not expect circumbinary material to survive in the case of (iv) the dynamical formation channels, where tight binary black holes are formed through multiple three body interactions in dense star clusters (e.g., Sigurdsson & Hernquist 1993; Kulkarni et al. 1993, and subsequent papers). The sequence of near collisions needed to produce a tight binary black hole will probably also strip it of any circumbinary material. We conclude that EM signals from a surviving circumbinary disk can potentially distinguish between formation scenarios.

**Disk evolution** — The evolution and fate of a possible circumbinary disk surrounding two black holes is subject to...
substantial uncertainties and poses a hard computational problem. Whether a disk can survive until coalescence of the black holes is a question that will ultimately be answered by observations. Here we assume that the disk has been formed by the progenitor stars through one of the scenarios above, and give simple physical considerations specifying the evolution and possible long-term survival of a circumbinary disk.

After the formation of the second black hole, there are no internal sources of stellar radiation. A fraction of the circumbinary disk may survive an initial phase of settling and cool to very low temperatures, probably even cooler than a normal protoplanetary disc (~10 K, e.g. Dullemond et al. 2007), implying an extremely small disc aspect ratio H/R ~ 10^{-3}. The disk mass is limited by self-gravity to a value H/R times the total binary mass M, otherwise fragmentation occurs (Armitage et al. 2001).

Cold disks require an external ionizing source to activate the magneto–rotational instability and so evolve viscously. A plausible source is the ambient cosmic ray flux. This ionizes a skin depth of order ~ 100 cm, giving a Shakura & Sunyaev (1973) viscosity parameter \( \alpha \sim 10^{-2} \) (Gammie 1996; Armitage et al. 2001) in the skin, where matter can at least initially flow inwards. The inflow rate is much lower near the mid plane of the disk, so there is a tendency for matter to accumulate in dead zones characterized by the local grain opacities (Gammie 1996).

Adapting this picture to the circumbinary disk of a black hole binary is clearly complex. The initial conditions of the disk are very uncertain, and inflowing material is dammed up at orbital resonances with the central binary. The inner region of the disk is subjected to resonant torques (cf. Lynden-Bell & Pringle 1974; Pringle 1991), which opens an inner cavity with a radius of about \( R_{in} \sim 2a \), where \( a \) is the binary separation (Artymowicz & Lubow 1994, e.g.).

A crucial question is whether material can flow through the inner cavity. This is important since the resulting accretion luminosity can further ionize the disk. This would in turn activate the magneto–rotational instability and might potentially lead to a runaway effect emptying the disk through the inner cavity. Various authors, e.g. Shi & Krolik (2015, and references therein), have argued for high accretion rates. Ragusa et al. (2016) instead claim that the flow of material from the inner portion of the disc through the cavity is drastically reduced for very thin and inactive disks. They argue that the higher accretion rates found in earlier studies are computational artefacts resulting from the artificially large scale heights adopted for numerical convenience. The findings by Ragusa et al. (2016) favor the possible long-term survival of the disk, which is needed for the scenario we propose here.

At present, it appears unlikely that any theoretical calculation can robustly answer the question of whether a fraction of the original circumbinary material can survive as a disk until the BH merger occurs. Given this uncertainty, we simply parametrize the disk mass \( M_d \) as a fraction \( q_{sd} \equiv M_d/M \) of the total binary mass.

The orbit of the binary system decays as a result of gravitational wave radiation. The timescale for this process is given by Peters (1964)

\[
t_{GW} \approx 1.1 \times 10^4 (a/R_\odot)^4 M_{60}^{-3} \text{ yr}
\]

where we have assumed equal masses. Initially, the timescale for orbital decay is long compared to viscous timescale and the disk can viscously respond to the change of the orbital separation by spreading and shrinking the inner cavity. (step 1 in Figure 1).

Eventually, \( t_{GW} \) becomes shorter than the viscous timescale \( t_{visc}(R_{in}) \) and the evolution of the disk decouples from the inner binary (e.g. Milosavljević & Phinney 2005, step 2 in Figure 1). This happens very shortly before the merger. The size of the inner radius \( R_{in} \) of the disk can be estimated by solving \( t_{GW}(a) = t_{visc}(R_{in}) \) where \( R_{in} \sim 2a \). The inner edge of the disk is now deep in the potential well of the BH binary, so must be hotter and ionized. We therefore now use the parameterization \( \alpha_{-1} = \alpha/10^{-1} \) appropriate for an ionized disk, which gives

\[
R_{in} \approx 3 \times 10^{10} \alpha_{-1}^{-2/5} [(H/R)_{-3}]^{-4/5} M_{60} \text{ cm}.
\]

Response to sudden mass loss and recoil at coalescence—When the binary black hole coalesces, the emission of gravitational waves produces a sudden reduction of the total central mass and it imparts an impulsive kick to the newly formed merged black hole (illustrated in step 3 in Figure 1, see Peres 1962; Fitchett 1983; Campanelli et al. 2007; Tichy & Marronetti 2008). Barausse et al. (2012) and Lousto & Zlochower (2014, and references therein) provide semi–analytic expressions for the viscous dissipation away and the recoil velocities (see also Schnittman & Buonanno 2007; Baker et al. 2007). They find significant fractions of mergers with \( v_{rec} > 500 - 1000 \text{ km s}^{-1} \), depending on the orientations and magnitudes of the spins of the coalescing black holes. These in turn bear the imprint of the spin evolution of the progenitor stars, modified by their subsequent implosion and possible birth kicks when forming black holes. These could be substantial in some cases (e.g. Repetto et al. 2012), but definitely not in all (Mirabel & Rodrigues 2003). Post–Newtonian effects can later align the binary further in some cases (e.g. Schnittman 2004). For the progenitor formation scenarios for isolated binary evolution (i and ii, see the origin section) one may expect spins with a substantial component out of the plane of the binary. The newly merged black hole resulting from such a system is generally expected to have a recoil with a substantial component directed perpendicular to the plane of the binary (e.g. Campanelli et al. 2007; Lousto et al. 2012).

Particles that remain bound to the newly formed black hole suddenly find themselves on elliptical orbits. The gas disk is expected to be cool and thin, so the orbital motion of the gas is hypersonic and susceptible to prompt shocks. These can, in principle, produce a transient EM signature (Milosavljević & Phinney 2005). Various groups have studied some or all of these effects in the context of supermassive binary black hole mergers, where bright prompt–merger flares because of shocks have been suggested as possibly detectable transient EM signatures (e.g. Bode & Phinney 2007; Lippai et al. 2008; Schnittman & Krolik 2008; Shields & Bonning 2008; Megevand et al. 2009; O’Neill et al. 2009; Krolik 2010; Rossi et al. 2010; Corrales et al. 2010).
Particularly large effects are expected for recoils with a substantial component \(v_{\text{rec},||}\) in the binary plane (Rossi et al. 2010). If this is in the 12 o’clock direction for a counterclockwise disk, gas at the 9 o’clock position suddenly has a greatly reduced angular momentum with respect to the (now moving) center of mass, particularly if \(v_{\text{rec},||}\) is comparable to the Kepler velocity \(v_K\) at the inner edge of the disk. This material falls deep into the potential well of the newly formed black hole and releases energy as it circularizes much closer to the black hole.

3. Electromagnetic Signal

We give crude estimates of the main characteristics of the resulting EM emission based on simple scaling relations calibrated against the numerical simulations by Rossi et al. (2010). We caution that the numerical estimates below are subject to large uncertainties.

1. Time delay and duration: Heating of the disk by shocks occurs on a characteristic timescale of order the dynamical time

\[
t_{\text{dyn}} \sim \frac{GM}{v^3} \sim 2.2 \frac{M_{60}}{v_3^3} \text{ hrs},
\]

where \(v_3 = v/(10^3 \text{ km s}^{-1})\) and \(v\) is the greater of the Keplerian velocity at the inner orbit and the recoil velocity imparted to the center of mass, i.e. \(v = \max(v_K(R_{\text{in}}), v_{\text{rec}})\). The light–travel time required to tell the disk about the GW event is much shorter than \(t_{\text{dyn}}\). Therefore, \(t_{\text{dyn}}\) describes the delay with respect to the GW event and energy deposition in the disk. This means that we expect a minimum time delay of hours between the merger and the EM signal, if photons are radiated away promptly. The time delay becomes longer, and can be years if a significant fraction of the radiation is trapped, as we discuss in 3. below.

2. Disk dimensions: We define a typical radius \(R_{\text{rec}}\) such that the Kepler velocity at this radius is equal to the recoil velocity

\[
R_{\text{rec}} = \frac{GM}{v_{\text{rec}}} \sim 10 \frac{M_{60}}{v_3} R_{\odot}.
\]

In the typical case we expect \(v_{\text{rec}} < v_K(R_{\text{in}})\), so that the inner radius of the disk falls within this radius (cf. 2). The outer disk radius is much larger than when the disk was formed (i.e. \(R_{\text{out}} > R_{\odot}\), as it must absorb the angular momentum of the gas which has spiraled inwards. Therefore we expect the outer radius to be significantly larger than \(R_{\text{rec}}\).

3. Peak luminosity: the expected rate of dissipation of kinetic energy is

\[
L \sim f \frac{M_{60}v^3}{t_{\text{dyn}}} \sim f \frac{v^5}{G} q_d
\]

where \(f\) is a scaling factor. We calibrate the factor \(f \sim 0.1\) against the numerical simulations by Rossi et al. (2010), which assume an angle of \(\theta = 15^\circ\) between the recoil direction and the orbital plane

\[
L \sim 5 \times 10^{42} \left(\frac{f}{0.1}\right) v_3^5 \frac{q_d}{10^{-3}} \text{ erg s}^{-1},
\]

because the dissipation lasts several dynamical times (cf. Fig 22 of Rossi et al. 2010). The peak EM luminosity depends on how promptly the radiation appears. The luminosity (6) is highly super–Eddington for any stellar–mass merger. If the shocked disc reacts by expanding homologously it would retain a high optical depth until significantly expanded. This means that it can take years for the photons to appear, which would strongly reducing the luminosity potentially rendering the signal undetectable. If instead the photons are released on a timescale that is comparable to the few–hour energy production time, then the peak luminosity of the EM signal is comparable to the rate of energy dissipation given in 6. This would give EM fluxes of up to \(10^{-15} v_3^2 f/(0.1)\) erg s\(^{-1}\) at the \(\sim 500\) Mpc distance of GW150914. This is potentially detectable at the expected photon energies (X–rays, see below).

The peak EM luminosity also depends very sensitively on the kick velocity \(v\) and its direction, and on the disk–to–black hole mass ratio \(q_d\). Note however that it is otherwise independent of the black hole mass. Remarkably, this means that the signal of stellar–mass events can potentially be as luminous as for mergers of supermassive black holes (but note that a more massive merger lasts longer (cf. Eq. 3), and so emits more energy in total).

4. Spectral Energy Distribution: The characteristic temperature \(T\) associated with the burst satisfies

\[
T_b < T < T_s
\]

where \(T_b\) is the blackbody temperature and \(T_s\) is the shock temperature given by

\[
T_b = \left(\frac{L}{2 \pi \sigma R_o^2}\right)^{1/4}
\]

\[
\sim 3 \times 10^6 v_3^{9/4} M_{60}^{-1/2} \left(\frac{q_d}{10^{-3}}\right)^{1/4} \left(\frac{f}{0.1}\right)^{1/4} \text{ K},
\]

\[
T_s = \frac{3 \pi m_H v^2}{16 k} \sim 2 \times 10^7 v_3^2 \text{ K}.
\]

Comparing (8, 9) we see that for \(v_3 \sim 1\) and assuming prompt emission the EM signal is likely to peak in medium–energy X–rays. Detection may also be possible at longer wavelengths, for example as the Rayleigh–Jeans tail in the infrared. This may be important if the emission is heavily reddened.

Given the luminosity estimate (6) and the timescale (3) it seems possible that events like these may already have been observed, but ascribed to variability in unidentified AGN.

5. Additional Spectral Features

The EM luminosity can be greatly enhanced for kicks in the orbital plane where \(v_{\text{rec}} \lesssim v_K(R_{\text{in}})\) and the material falls deep into the potential of the merged black hole. This gives red– or blueshifts of the same order (say \(\sim 1000\) km s\(^{-1}\)) for the EM emission, which can in principle be used to probe the post merger recoil. Further, the composition of the material of the disk may be enriched if the material results from ejecta that were processed by nuclear burning in the progenitor stars. For the closest events this might eventually lead to detectable spectral diagnostics.

4. Discussion
Our estimates are subject to various significant uncertainties. These include, but are not limited to, the evolution of the circumbinary disk and its response to heating by tides, spiral waves and cosmic rays, the pre-merger spins of the black holes, the size and direction of the recoil, and how promptly the dissipated energy appears from the shocked disk.

Given these large unknowns, we cannot state with certainty whether the signals are observable or not. Further detailed simulations will be needed for more accurate predictions, and even these face the difficulty that the behavior of the disk viscosity over time is inherently uncertain. A definitive answer will only come from observations.

Observing the signal — If the EM signal appears a few hours after the GW merger, a fairly accurate position is needed for suitable instruments to scan the error box on a similar timescale. This is not possible with the currently available combination of LIGO and SWIFT alone, but will become increasingly feasible as additional GW detectors come online and allow accurate triangulation of GW events. A particularly attractive possibility is that eLISA may be able to detect black hole binaries several years before coalescence (Sesana 2016), allowing targeted campaigns. Our estimates (6, 8, 9) of the luminosity and temperature, and the possible radial velocity signal, suggest that searches from X-rays to infrared could be fruitful. There are a large number of current, planned, and projected instruments which would be suitable for this (e.g. ZTF, ATLAS, GOTO, BlackGem, LSST, SVOM, Einstein Probe).

Other possible EM signals — Recently, various groups proposed the possibility of other types of observable EM signals, encouraged by the tentative Fermi signal. Perna et al. (2016) also consider the possibility that material ejected in the second weak SN explosion might form a dead disk, which we also consider, but our scenario differs from theirs in various aspects. They assume a minidisk around one of the individual black holes, heated by tidal torques and shocks during the pre-merger phase. This rapidly consumes the disk and powers a short gamma-ray burst coinciding with the merger event (see also Murase et al. 2016). Kimura et al. (2016) argue to the contrary that such a disk would be heated and activated a few years before the merger event and lead to a precursor event instead. Our scenario differs in the disk geometry, which is circumbinary in our picture. As a result the brightening arises from perturbations of the circumbinary disk itself, on its dynamical timescale. This is not directly related to viscous accretion on to the newly formed black hole, where the timescale is much longer. Other suggestions beyond those mentioned in the Introduction include signals from the merger of charged black holes (Zhang 2016; Fraschetti 2016; Liebling & Palenzuela 2016).

5. CONCLUSIONS

We have suggested that mergers of stellar–mass black hole binaries driven by gravitational wave emission potentially produce EM counterparts if a low–mass circumbinary disk survives until the coalescence. We argue that such a disk responds to sudden mass loss and recoil of the GW merger leading to shock, which heats it within hours of the merger. Whether the signal is detectable is uncertain. If the optical depth through the disk is high the prompt emission consists only of photons from its optically thin outer layers, while the majority may take years to emerge. If by some mechanism, photons are released promptly the resulting EM signal will probably appear similar to background AGN events, most likely in medium–energy X–rays, possibly extending to the infrared, and last at least a few hours. They would become identifiable once triangulation by future GW detectors produces error boxes small enough that prompt–response EM instruments can scan them on this timescale. In certain cases they may be pre–detected years in advance once the Evolved Laser Interferometer Space Antenna comes online. We also expect significant radial velocities (∼1000 km s⁻¹) in certain cases, which might be detectable in a rapid optical follow-up.

Despite the large uncertainties, we suggest that the potential rewards of a successful detection of the signal discussed here justify the effort of coordinated EM campaigns. It would place unique constraints on the binary evolution before the merger, and thus provide crucial information about the contribution of various progenitor channels. Detection of a radial velocity, as suggested above, would also constrain the recoil predicted by the GW data, and so directly test strong-field predictions of General Relativity.

Acknowledgements. We thank the anonymous referees for helpful comments, and A. Buonanno, Z. Haiman, A. MacFadyen, Y. Levin, I. Mandel, C. Miller, Ch. Nixon, P. O’Brien, M. Renzo, E. Rossi and R. Wijers for stimulating discussions. SdM acknowledges support from a Marie Sklodowska-Curie Action (H2020 MSCA-IF-2014, project id 661502) and National Science Foundation under Grant No. NSF PHY11-25915. Astrophysics research at the University of Leicester is supported by an STFC Consolidated Grant.

REFERENCES
