

Massive BH binaries as periodically variable AGN

Luke Zoltan Kelley^{1,2★}, Zoltán Haiman,³ Alberto Sesana⁴ and Lars Hernquist¹

¹Harvard-Smithsonian Center for Astrophysics; Harvard University, Cambridge, MA 02138, USA

²Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA); Northwestern University, Evanston, IL 60208, USA

³Department of Astronomy, Columbia University, New York, NY 10027, USA

⁴School of Physics and Astronomy and Institute of Gravitational Wave Astronomy, University of Birmingham, Edgbaston B15 2TT, UK

Accepted 2019 January 11. Received 2018 December 6; in original form 2018 August 7

ABSTRACT

Massive black hole (MBH) binaries, which are expected to form following the merger of their parent galaxies, produce gravitational waves that will be detectable by pulsar timing arrays at nanohertz frequencies (year periods). While no confirmed, compact MBH binary systems have been seen in electromagnetic observations, a large number of candidates have recently been identified in optical surveys of active galactic nucleus (AGN) variability. Using a combination of cosmological, hydrodynamic simulations; comprehensive, semi-analytic binary merger models; and analytic AGN spectra and variability prescriptions; we calculate the expected electromagnetic detection rates of MBH binaries as periodically variable AGN. In particular, we consider two independent variability models: (i) Doppler boosting due to large orbital velocities and (ii) hydrodynamic variability in which the fuelling of MBH accretion discs is periodically modulated by the companion. Our models predict that numerous MBH binaries should be present and distinguishable in the existing data. In particular, our fiducial models produce an expectation value of 0.2 (Doppler) and 5 (hydrodynamic) binaries to be identifiable in the Catalina survey (CRTS), while 20 and 100 are expected after 5 yr of LSST observations. The brightness variations in most systems are too small to be distinguishable, but almost 1 per cent of AGN at redshifts $z \lesssim 0.6$ could be in massive binaries. We analyse the predicted binary parameters of observable systems and their selection biases, and include an extensive discussion of our model parameters and uncertainties.

Key words: quasars: supermassive black holes – galaxies: kinematics and dynamics.

1 INTRODUCTION

Active galactic nuclei (AGNs) are known to often be triggered by interactions and mergers between their host galaxies (e.g. Kauffmann & Haehnelt 2000; Comerford et al. 2015; Barrows et al. 2017; Goulding et al. 2018) that drive large amounts of gas towards the galaxy cores and massive black holes (MBHs) within (Barnes & Hernquist 1992). Many examples of ‘dual-AGN’, pairs of observably accreting MBHs in the same system, have been identified in radio, optical, and X-ray surveys (e.g. Komossa 2006; Rodriguez et al. 2006; Comerford et al. 2012; Koss et al. 2012). After a galaxy merger, the two MBHs are expected to sink towards the centre of the post-merger galaxy due to dynamical friction, which is very effective on $\sim 10^3$ pc scales (Begelman, Blandford & Rees 1980; Antonini & Merritt 2012). Once the MBHs reach $\sim$ pc separations and smaller, and eventually become gravitationally bound as an MBH binary (MBHB), the continued merging of the system depends sensitively on individual stellar scatterings extracting energy from

the binary (e.g. Merritt, Mikkola & Szell 2007; Sesana, Haardt & Madau 2007).

The effectiveness of stellar scattering in ‘hardening’ MBHBs remains unresolved, although studies are beginning to reach a consensus that the population of stars available for scattering (the ‘loss-cone’) is efficiently refilled (e.g. Sesana & Khan 2015; Vasiliev, Antonini & Merritt 2015). Of particular interest is whether and which systems are able to reach the $\lesssim 10^{-3}$ – 10^{-1} pc separations¹ at which point gravitational wave (GW) emission can drive the systems to coalesce within a Hubble time (Begelman et al. 1980). While dual-AGNs have been observed, there are no known AGN in confirmed gravitational bound binaries. If MBHBs are able to reach periods of $\sim$ yr (frequencies $\sim$ nHz), their GW emission should be detectable by pulsar timing arrays (PTAs; Hellings & Downs 1983; Foster & Backer 1990) – the European (EPTA; Desvignes et al. 2016), NANOGrav (The NANOGrav Collaboration et al. 2015), Parkes (PPTA; Reardon et al. 2016), and the International PTA (IPTA; Verbiest et al. 2016). The most recent and comprehensive

★ E-mail: lzkelley@northwestern.edu

¹For masses $\sim 10^6$ – $10^9 M_\odot$.

models for the cosmological population of merging MBHBs suggest that PTAs will plausibly make a detection within roughly a decade (e.g. Rosado, Sesana & Gair 2015; Taylor et al. 2015; Kelley et al. 2017b), and indeed, the most recent PTA upper limits on GW signals – particularly on the presence of a power-law, gravitational-wave background (GWB) of unresolved, cosmological sources – have already begun to inform the astrophysical models (Simon & Burke-Spolaor 2016; Taylor, Simon & Sampson 2017; but also, Middleton et al. 2018).

MBHBs form on sub-parsec scales, which, even using very-long baseline interferometry (VLBI), can only be spatially resolved at relatively low redshifts (e.g. D’Orazio & Loeb 2018). Spectroscopic and especially photometric methods, which do not require binaries to be spatially resolved, have recently put forward large numbers of binary candidates (Tsalman et al. 2011; Eracleous et al. 2012; Graham et al. 2015a; Charisi et al. 2016). On the theoretical side, few predictions have been made for the expected observability and detected rates of AGN in binary systems. On kpc scales, Steinborn et al. (2016) study dual- and offset-AGN during the (relatively) early stages of galaxy merger, providing interesting results on the nature and properties of the MBH in these systems. Volonteri, Miller & Dotti (2009) make rate predictions for tight MBHB, especially those with large orbital velocities that could be observable as dual broad-line AGN (e.g. Boroson & Lauer 2009). The authors predict an upper limit to the detection rate of such systems between 0.6 and 1.0×10^{-3} per unabsorbed AGN.

The focuses of this investigation are binaries and candidates identified by periodic variability in photometric surveys of AGN. In particular, Graham et al. (2015a) find 111 candidates in $\sim 240\,000$ AGN using the CRTS survey; Charisi et al. (2016) find 33 in $\sim 35\,000$ AGN using PTF; and Liu et al. (2016) initially identify 3 candidates in 670 AGNs using PanSTARRS; however, none are persistent in archival data. These three detection rates are roughly consistent at 5×10^{-4} , 9×10^{-4} , and $\lesssim 4 \times 10^{-3}$ AGN $^{-1}$. It is worth noting that these detection rates are very similar to the limits from orbital-velocity selected populations in Volonteri et al. (2009), which have similar binary parameters to the orbital-period selection for variability candidates.

The connection between EM and GW observations of MBHBs has already begun to be leveraged using these photometric variability candidates. While none of the individual candidate systems can be excluded by PTA measurements, Sesana et al. (2018) demonstrate that the *population* of MBHBs that they imply leads to a GWB amplitude in tension with existing PTA upper limits. In Sesana et al. (2018), a phenomenological approach, with as few physical assumptions as possible, is used to connect EM and GW observations. In this follow-up analysis, we rely instead on physically motivated, theoretical models to explore the repercussions on EM observations. Specifically, we use binary populations based on the Illustris hydrodynamic, cosmological simulations (e.g. Vogelsberger et al. 2014a; Nelson et al. 2015) coupled with comprehensive semi-analytic merger models (Kelley, Blecha & Hernquist 2017a; Kelley et al. 2017b) and synthetic AGN spectra to make predictions for the occurrence rates of periodically variable AGN. In Section 2, we summarize the binary population, the AGN spectra we use to illuminate them, and the models of variability we consider. In Section 3, we present our results of expected detection rates and the parameters that determine binary observability. Finally in Section 4, we discuss the limitations of our study and discuss its implications for identifying and confirming MBHBs through photometric variability studies.

2 METHODS

2.1 MBHB population and evolution

Our MBHB populations are based on the MBHs and galaxies in the Illustris simulations. Illustris is an $(108\text{ Mpc})^3$ volume of gas cells and particles representing dark matter, stars, and MBHs that is evolved from the early universe to redshift zero (e.g. Genel et al. 2014; Torrey et al. 2014; Vogelsberger et al. 2014a,b; Nelson et al. 2015; Rodriguez-Gomez et al. 2015). The simulations include sub-grid models for star formation, stellar nucleosynthesis and metal enrichment, and stellar and AGN feedback. MBH particles are initialized with a seed mass of $\sim 10^5 M_\odot$ in massive halo centres, after which they grow via accretion of local gas using a Bondi model. Details of the BH prescription and resulting MBH and AGN populations are presented in Sijacki et al. (2015). In the Illustris simulations, after or during a galaxy merger, once MBHs come within $\sim 10^2$ – 10^3 pc of one-another – roughly their gravitational smoothing length – they are manually merged and moved to the potential minimum of the halo. To more carefully examine the MBHB merger process and dynamics, we ‘post-process’ the MBH mergers using semi-analytic models.

In this section, we outline some of the key components of the merger models and the resulting merger dynamics that are described thoroughly in Kelley et al. (2017a,b). MBH–MBH ‘merger-events’ are identified in Illustris on $\sim$ kpc scales. We then consider each of these events independently by extracting the MBH masses, and spherically averaged galaxy density and velocity profiles for each constituent (dark matter, stars, gas) of the host. These profiles are then used to calculate hardening rates of the semimajor axis (da/dt) based on prescriptions for dynamical friction (Chandrasekhar 1942; Binney & Tremaine 1987), stellar ‘loss-cone’ scattering (Magorrian & Tremaine 1999), viscous drag from a circumbinary disc (Haiman, Kocsis & Menou 2009; Tang, MacFadyen & Haiman 2017), and GW emission (Peters & Mathews 1963; Peters 1964). Dynamical friction is required to harden the system on 10 – 10^3 pc scales, after which stellar scattering is typically dominant until the GW-dominated regime on $\sim 10^{-2}$ – 10^{-4} pc. In some systems, viscous drag can be the primary hardening mechanism near $\sim 10^{-2}$ pc. Our population of binaries is roughly 10^4 systems with total masses between $2 \times 10^6 M_\odot$ and $2 \times 10^{10} M_\odot$, with a steeply declining mass function. The mass ratio of the systems is inversely correlated with total mass: systems with low total masses can only have near-equal mass ratios due to the minimum MBH mass, and high total mass systems are dominated by extreme mass ratio mergers.

Both stellar scattering and viscous drag remain highly uncertain processes. The largest uncertainty affecting merger outcomes is likely the effectiveness of stellar scattering: in particular, how efficiently the stellar ‘loss-cone’ – those stars able to interact with the binary – are repopulated. Typical coalescence lifetimes are gigayears. Binaries that are both very massive $M \equiv M_1 + M_2 \gtrsim 10^9 M_\odot$, and near equal mass ratio $q \equiv M_2/M_1 \gtrsim 0.1$, are generally able to coalesce within a Hubble time. Systems with both lower total masses ($M \lesssim 10^8 M_\odot$) and more extreme mass ratios ($q \lesssim 10^{-2}$) often stall at either kpc or pc separations.² The fate of the

²E.g. for binaries with $M \lesssim 10^7 M_\odot$, ~ 30 per cent of $q \gtrsim 0.3$ systems coalesce before redshift zero, and only ~ 10 per cent of those with $q \lesssim 0.3$. For the latter, low mass ratio systems, this is unsurprising as dynamical friction is often ineffective at hardening to below $\sim$ kpc separations (e.g. McWilliams, Ostriker & Pretorius 2014). In our models, we find that more comparable mass systems are more likely to stall at $\sim$ pc scales, with stellar

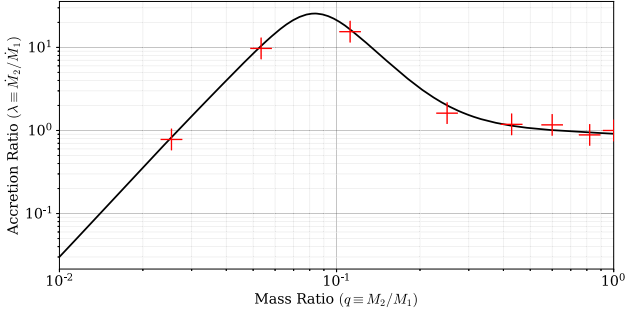


Figure 1. Accretion ratio data points are from hydrodynamic simulations of MBHB in circumbinary accretion discs by Farris et al. (2014). The line is fit with the function and parameters that are shown in equation (1).

remaining, intermediate systems depend more sensitively on the assumed dynamical parameters (i.e. the loss-cone refilling rate). Note that this differs from some previous studies finding that more massive systems merge *less* effectively (e.g. Yu 2002; Cuadra et al. 2009; Dotti, Merloni & Montuori 2015).³ The systems that reach $\sim$ yr periods are thus somewhat biased against low total masses and extreme mass ratios.

Predictions for the GWB and its prospects for detection by PTAs are presented in Kelley et al. (2017b), along with a description of our formalism for eccentric binary evolution. Most models predict GWB amplitudes at periods of 1 yr, $A_{\text{yr}^{-1}} \approx 0.5\text{--}0.7 \times 10^{-15}$, roughly a factor of 2 below current sensitivities, and detectable within about a decade. Predictions for GW signals from individually resolvable ‘single-sources’ are presented in Kelley et al. (2018), and are comparable in detectability to the GWB. The results we present in this paper are relatively insensitive to variations in binary evolution parameters, compared to those of the electromagnetic and observational models we describe below. For reference, the evolutionary model used here assumes an always full loss-cone and initial binary eccentricities of $e_0 = 0.5$.

2.2 MBH accretion and AGN spectra

Our merger models follow the constituent MBHs of a given binary for long after it has ‘merged’ in Illustris. After the ‘merger’, Illustris records the accretion rate of ambient gas on to the single, remnant MBH. We use this accretion rate as a measure of the fuelling to the binary system as a whole, feeding the circumbinary disc: $\dot{M} = \dot{M}_1 + \dot{M}_2$. In our post-processing models, however, the two MBHs are *unmerged*, leaving an ambiguity in the feeding rate to each individual component. To resolve this, we use the results from the detailed circumbinary disc simulations in Farris et al. (2014), which give the ratio of accretion rates for a variety of binary mass ratios: $\lambda = \lambda(q) \equiv \dot{M}_2 / \dot{M}_1$. The simulation data points are plotted in Fig. 1, along with a fit described by the function,

scattering becoming ineffective – likely due to less centrally concentrated stellar mass distributions (see below).

³The discrepancy between Illustris-based models and those of some semi-analytic merger populations likely has to do with typical stellar densities in the cores of low-mass galaxies. In particular, the densities from Illustris may be systematically lower, but this has yet to be examined in detail. The fraction of observable systems with low masses ($M \lesssim 10^7 M_\odot$) drops rapidly (e.g. Fig. 11), implying that this difference between models has little effect on our results.

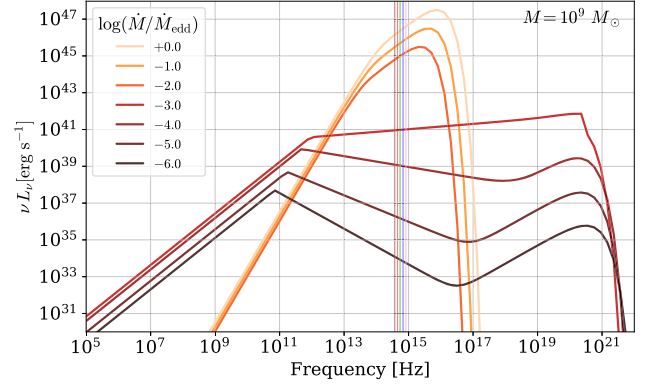


Figure 2. AGN spectra for an MBH of mass $M = 10^9 M_\odot$, and a variety of accretion rates. For Eddington ratios $f_{\text{Edd}} \equiv \dot{M} / \dot{M}_{\text{Edd}} < 10^{-2}$, we use the ADAF emission model from Mahadevan (1997), while for larger accretion rates we use a thermal, Shakura–Sunyaev spectrum. For reference, the vertical coloured lines are the optical bands: [i, r, g, v, b, u].

$$\lambda = q^{a_1} e^{-a_2/q} + \frac{a_3}{(a_4 q)^{a_5} + (a_4 q)^{-a_5}},$$

$$a_1 = -0.25, \quad a_2 = 0.1, \quad a_3 = 50, \quad a_4 = 12, \quad a_5 = 3.5. \quad (1)$$

We assume that the system is Eddington limited on large scales, i.e. $\dot{M} \leq \dot{M}_{\text{Edd}} \equiv \dot{M}_{\text{Edd},1} + \dot{M}_{\text{Edd},2}$, where $\dot{M}_{\text{Edd}} = 1.4 \times 10^{18} \text{ g s}^{-1} \left(\frac{M}{M_\odot} \right) \left(\frac{\epsilon_{\text{rad}}}{0.1} \right)^{-1}$, and ϵ_{rad} is the radiative efficiency which we take as 0.1. We let each MBH individually exceed Eddington⁴ (e.g. Jiang, Stone & Davis 2014) that can occur for the secondary when $\lambda > 1.0$, corresponding to $q \gtrsim 0.03$. The secondary accretion rate is maximized at $\lambda_{\text{max}} = \lambda(q \approx 0.08) \approx 25$.

The parameters for MBH evolution in Illustris are calibrated to match the observed M – σ relation, and the AGN (bolometric) luminosity function based on a constant radiative efficiency of 0.05 (see Sijacki et al. 2015). For our analysis, we calculate full spectra for each MBH based on its mass and Eddington ratio, $f_{\text{Edd}} \equiv \dot{M} / \dot{M}_{\text{Edd}}$, obtained from the accretion rates described above. For $f_{\text{Edd}} \geq 10^{-2}$, we assume the accretion flow is radiatively efficient and use a Shakura & Sunyaev (1973) ‘thin’-disc solution, which assumes emission is purely thermal from each annulus of the disc. For $f_{\text{Edd}} < 10^{-2}$, we assume radiatively inefficient accretion in the form of an ADAF (Narayan & Yi 1995) and use the emission model from Mahadevan (1997). The ADAF model includes self-absorbed synchrotron emission, bremsstrahlung, and inverse-Compton of synchrotron photons. We thus calculate AGN spectra as

$$F_\nu = F_\nu^{\text{thin}}(M, f_{\text{Edd}}) \quad f_{\text{Edd}} \geq 10^{-2}, \quad (2)$$

$$= F_\nu^{\text{ADAF}}(M, f_{\text{Edd}}) \quad f_{\text{Edd}} < 10^{-2}. \quad (3)$$

Spectra for a variety of accretion rates on to an $M = 10^9 M_\odot$ BH are shown in Fig. 2. The bolometric luminosity and the luminosities in the B and V bands are shown in Fig. 3, along with the effective radiative-efficiency ($L_{\text{bol}} / L_{\text{Edd}}$) and luminosity fractions (the inverse of the bolometric corrections). The change in spectral shape and optical luminosity at the transition of $f_{\text{Edd}} = 10^{-2}$, corresponding to the change from thick to thin accretion flows, is clear in both Figs 2 and 3. While this transition is likely artificially abrupt, it is consistent with transitions in the state of X-ray binaries, associated

⁴The alternative is explored in Section 3.3.

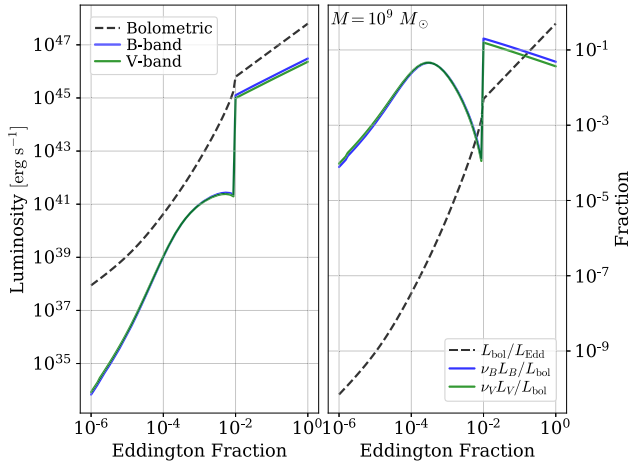


Figure 3. Luminosity and radiative efficiency versus Eddington ratio. The left-hand panel shows bolometric (dashed), *B*- and *V*-band luminosities (solid lines), calculated for a $M = 10^9 M_\odot$ MBH, against a variety of accretion rates. The right-hand panel gives the overall radiative efficiency $\epsilon_{\text{rad}} = L_{\text{bol}}/L_{\text{Edd}}$, as well as the fraction of energy emitted in the *B* and *V* bands (i.e. the inverse of the bolometric corrections).

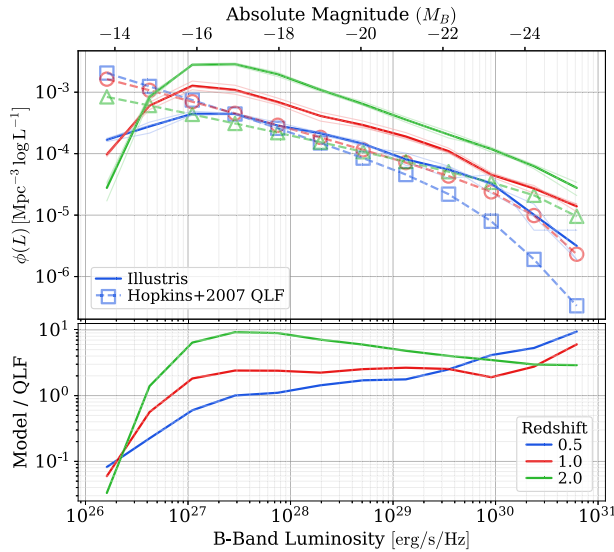


Figure 4. AGN luminosity functions constructed from Illustris MBH (solid lines) and the observationally derived QLF from Hopkins et al. (2007, points/dashed lines). The top panel shows the luminosity functions at three different redshifts, $z = 0.5, 1.0$, and 2.0 , and the bottom panel shows the ratio of QLFs between Illustris and observations. While our models are consistent with observations to an order of magnitude, our MBH populations and synthetic spectra noticeably overpredict the luminosity function. Note that this comparison is shown for the *B* band, while most of our analysis focuses on the *V* band.

with the same changes in accretion regime (e.g. Esin, McClintock & Narayan 1997). Regardless, the specific location and sharpness of this transition has little effect on our results as observable systems are predominantly at $f_{\text{Edd}} \gtrsim 0.1$.

The luminosity function of Illustris AGN, using our spectral models, is compared to the observationally determined quasar luminosity function (QLF) from Hopkins, Richards & Hernquist (2007) in Fig. 4. Comparing observed and simulated AGN observations is non-trivial. On the observational side, there are extinction,

bolometric corrections, *K*-corrections, and selection biases; while on the simulation side, we are using disc-integrated quantities based on semi-analytic models instead of either radiative transfer calculations or full disc simulations. None the less, our models agree with observations to well within an order of magnitude, although the Illustris AGNs tend to be systematically overluminous compared to the observed QLF. Throughout our analysis, we present our results primarily in terms of observable binary detections normalized to the predicted number of observable AGN. This is to reduce errors from our reproduction of the luminosity function. In Section 3.3, we also present results for a model in which all accretion rates have been lowered by a factor of 3, which does produce a better match to the observed QLF.

The luminosity functions from Hopkins et al. (2007), shown in Fig. 4, correct for obscuration. In our analysis, we use the same model that assumes that a luminosity-dependent fraction of systems are observable,

$$f(L) = \min \left[1, f_{46} \left(\frac{L}{10^{46} \text{ erg s}^{-1}} \right)^\beta \right], \quad (4)$$

$$B \text{ band: } f_{46} = 0.260, \beta = 0.082,$$

where L is the bolometric luminosity, and the fit values are for the *B* band. We convert between bolometric and spectral luminosity using a luminosity-dependent bolometric correction (Hopkins et al. 2007),

$$\frac{L}{L_i} = c_1 \left(\frac{L}{L_\odot} \right)^{k_1} + c_2 \left(\frac{L}{L_\odot} \right)^{k_2}, \quad (5)$$

$$B \text{ band: } c_1 = 7.40, c_2 = 10.66, k_1 = -0.37, k_2 = -0.014.$$

This bolometric correction, which would differ in general from that in our spectral models, is used here for the sole purpose of computing the obscuration fraction in a way consistent with Hopkins et al. (2007). It is not used elsewhere in our analysis.

One additional adjustment is required for spectra from discs in binary systems: the presence of a companion leads to truncation of each ‘circum-single’ disc at a radius comparable to the Hill radius. Specifically, we set the outer edge of each disc to

$$r_{\text{max}} = \frac{a}{2} \left(\frac{M_i}{M} \right)^{1/3}, \sim 74 R_s \left(\frac{P}{5 \text{ yr}} \right)^{2/3} \left(\frac{M}{10^9 M_\odot} \right)^{-2/3}, \quad (6)$$

where M_i is the mass of the primary or secondary, a is the semimajor axis, P is the orbital period, and R_s is the Schwarzschild radius of the MBH in question. A typical geometry for the binary is shown schematically in Fig. 5, showing the two circum-single discs around each MBH, within a larger circumbinary disc. While the bright, optical emission in AGN tends to come from relatively small radii, disc truncation can be important for especially massive BHs and those in short-period binaries. For example, the optical luminosity of a $10^9 M_\odot$ MBH, in a 5 yr period binary, can be decreased by ~ 10 – 20 per cent. When calculating the total luminosities of binaries, we also include contributions from the circumbinary portion of the accretion disc, where, fiducially, we assume that the inner-edge occurs at twice the binary separation, and the inner-edge of each circum-single disc is located at $3 R_s$ (i.e. the inner-most stable circular orbit for a non-spinning BH).

2.3 Populations and observations

We wish to calculate the number of binaries observable at a given log-interval of period, and a particular interval of redshift,

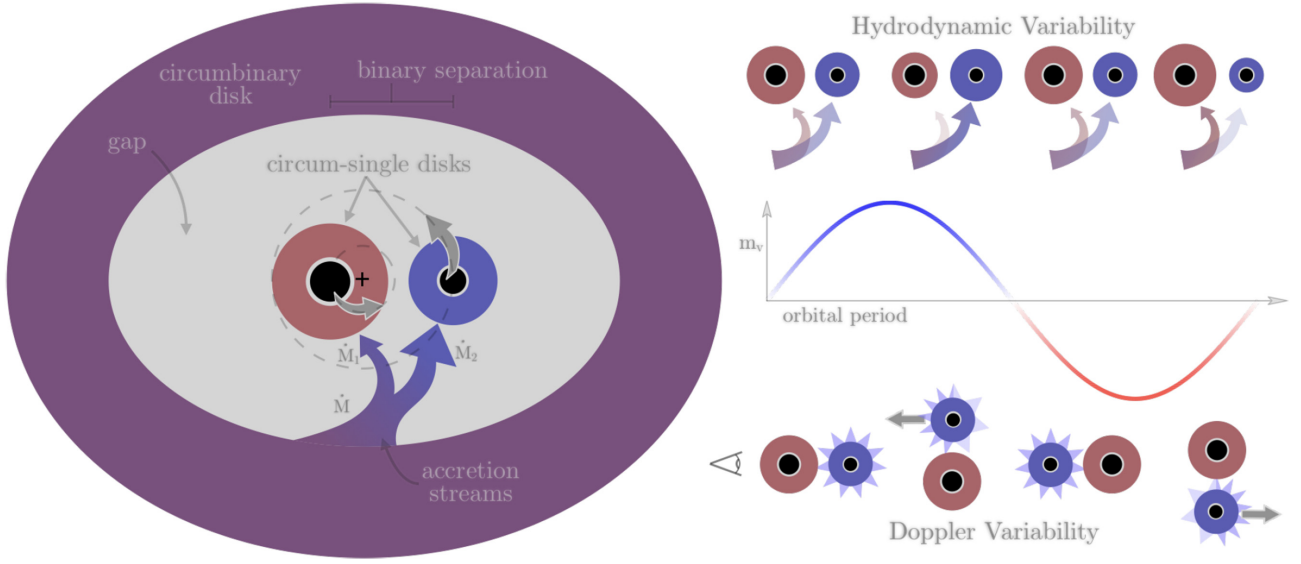


Figure 5. Schematic representation of the binary, disc, and accretion geometries assumed in our models; informed from the results of hydrodynamic simulations (e.g. Farris et al. 2014). *Left:* between the binary and the circumbinary disk is a ‘gap’ with a radius roughly twice the binary separation. Around each MBH is a ‘circum-single’ disc, fed by time variable accretion streams extending from the circumbinary disc. Because the secondary MBH is farther from the centre of mass, and closer to the circumbinary disc, it tends to receive a disproportionate share of the accretion rate. *Right:* the hydrodynamic and Doppler mechanisms for producing photometric variability are depicted on the top and bottom, respectively. The circumbinary disc orbits at longer periods than the circum-single discs that it feeds, causing periodic variations in accretion rate, and thus luminosity. For observers oriented near the orbital plane, Doppler boosting of the faster moving, and typically more luminous, secondary MBH can also produce brightness variations.

$d^2N/dz d \log_{10} p$. By using the number density in a comoving volume, $n \equiv dN/dV_c$, we can write

$$\frac{d^2N}{dz d \log_{10} p} = f_{\text{obs}} \frac{p}{\ln 10} \frac{dn}{dp} \frac{dV_c}{dz}. \quad (7)$$

We have introduced the factor f_{obs} , the fraction of systems that are observable at a given redshift, which is generally a function of any binary parameter (e.g. mass, orbital period, and inclination).

We consider particular periods of interest p_j , and for each simulated binary i in our population, we find the redshift at which it reaches those periods: z_{ij} . To calculate the number or number density of sources, we consider discrete bins in redshift, $\Delta z'_k \in [z'_k, z'_{k+1})$, and identify all binaries reaching the period of interest in that bin. A given binary will be counted at all periods that it reaches before redshift zero. In this way, we effectively treat each moment in time, for each binary, as a separate data sample – i.e. each represents an independent population of astrophysical binaries (at a slightly different redshift).

Because we are sampling in orbital period, instead of evenly in time, we must explicitly account for the time evolution of binaries by considering the fraction of time binaries spend emitting at each period. The temporal evolution of binaries can be written as

$$\frac{dn}{dp} = \frac{dn}{dt} \frac{dt}{df} \frac{df}{dp}. \quad (8)$$

Using Kepler’s law along with the hardening time, $\tau_h \equiv a/(da/dt)$, we can write

$$\frac{dt}{df} = \frac{2}{3} \tau_h p. \quad (9)$$

The number of binaries at period $p = p_j$, in redshift bin k , is then

$$\begin{aligned} \frac{dN_{jk}}{d \log_{10} p_j} &= \sum_i \frac{f_{\text{obs},i}}{\ln 10} \delta(z'_k \leq z_{ij} < z'_{k+1}) \mathcal{T}_{ijk} \mathcal{V}_k, \\ \mathcal{T}_{ijk} &\equiv \min \left(\frac{2}{3} \frac{\tau_{h,ij}}{\Delta t_k}, 1 \right), \\ \mathcal{V}_k &\equiv \frac{1}{V_{\text{ill}}} \frac{dV_c}{dz} \Delta z'_k. \end{aligned} \quad (10)$$

The differential, comoving volume of the universe as a function of redshift, dV_c/dz , is given in Hogg (1999, equation 28).

The fraction of systems observable at a given redshift and period, f_{obs} , depends on the emission and variability model. We require that the flux from the source is above the flux threshold, $F_v \geq F_{v,\text{sens}}$, and that the variability, $\delta_F \equiv \Delta F_v/F_v$, is above the variability threshold, $\delta_F \geq \delta_{F,\text{sens}}$, i.e.

$$f_{\text{obs}} = \Theta(F_v \geq F_{v,\text{sens}}) \cdot \Theta(\delta_F \geq \delta_{F,\text{sens}}). \quad (11)$$

Here, Θ is the Heaviside function: unity when the argument is true, and zero otherwise. The variability threshold depends on the signal-to-noise ratio (SNR) and a minimum variability floor, $\delta_{F,\text{min}}$. Based on the smallest variability amplitudes seen in Graham et al. (2015a), we use a fiducial value of $\delta_{F,\text{min}} = 0.05$. We calculate the variability threshold as

$$\delta_{F,\text{sens}} = \text{SNR}^{-1} + \delta_{F,\text{min}}, \quad (12)$$

which is consistent with the results of variability studies in *HST* by Sarajedini, Gilliland & Kasm (2003, see their fig. 3) and in PanSTARRS by Liu et al. (2016, see their fig. 4). In both cases, the authors find minimum detectable variabilities of ~ 1 –2 per cent, and then select systems using a cut which is some factor larger, at ~ 5 per cent. Similarly, we calculate the SNR by assuming the flux-sensitivity threshold is a factor of 5 above the noise, i.e. $\text{SNR} = 5F_v/F_{v,\text{sens}}$.

Our results are calculated for a range of detector sensitivities that encompass current and near-future instruments. For convenience, we frequently present our results in terms of two characteristic values based on CRTS and LSST sensitivities. In particular, we use V-band sensitivities of $F_{v,\text{sens}}^{\text{CRTS}} = 4 \times 10^{-28} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ cm}^{-2}$ ($m_V \approx 19.5$) and $F_{v,\text{sens}}^{\text{LSST}} = 3 \times 10^{-30} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ cm}^{-2}$ ($m_V \approx 24.9$) for CRTS and LSST, respectively.⁵ CRTS is considered in particular because of their large sample of binary candidates, but we consider this to be representative of current survey capabilities in general.

The overall number density of sources (i.e. in units of Mpc^{-3}) is the metric most directly extracted from our models. To better compare with observations, and to reduce the impact of systematic uncertainties in our luminosity functions, we focus on the number of simulated, observable binaries ($N_{\text{bin}}^{\text{sim}}$) as a fraction of the expected number of all observable AGN ($N_{\text{AGN}}^{\text{sim}}$), which we report in units of AGN^{-1} . To compute $N_{\text{AGN}}^{\text{sim}}$, we calculate the observability of all Illustris MBHs, using the same spectral models described in Section 2.1, out to a redshift $z_{\text{max}} = 4.0$. For the aforementioned sensitivities, our fiducial simulations predict 10^6 all-sky AGN to be observable by CRTS, and 4×10^7 by LSST. CRTS actually observes $\approx 3 \times 10^5$ spectroscopically confirmed AGN (Graham et al. 2015a). The spectroscopic data come primarily from SDSS, which covers only about one-quarter of the sky, suggesting an all-sky number a little above 10^6 , and consistent with our calculation.⁶

For a given sensitivity, our simulations predict an expected number of detected AGN and periodically variable binaries. A more robust prediction for a given survey can be calculated from the binary detection rate per AGN, along with the actual number of monitored AGN in the survey. We refer to this prediction as a ‘rescaled’ number of expected detections,

$$N_{\text{bin}}^{\text{rescaled}} = N_{\text{bin}}^{\text{sim}} \frac{N_{\text{AGN}}^{\text{obs}}}{N_{\text{AGN}}^{\text{sim}}}. \quad (13)$$

When making predictions for LSST, we assume a completeness $f_{\text{complete}}^{\text{LSST}} = 2$, relative to CRTS,⁷ i.e.

$$N_{\text{AGN}}^{\text{obs,LSST}} = N_{\text{AGN}}^{\text{sim,LSST}} \frac{N_{\text{AGN}}^{\text{obs,CRTS}}}{N_{\text{AGN}}^{\text{sim,CRTS}}} f_{\text{complete}}^{\text{LSST}}. \quad (14)$$

2.4 Models of variability

The luminosity of an object in a binary system will not necessarily vary on the orbital period or at all. The premise of photometric identification of MBHBs is that the binary period is somehow imprinted into variations of the observed luminosity. In the particularly convincing example of PG 1302-102 (Graham et al. 2015b), sinusoidal variations in the light curve can be well explained

by Doppler boosting from a mildly relativistic orbital velocity (D’Orazio, Haiman & Schiminovich 2015b; but see also Liu, Gezari & Miller 2018). Additionally, purely hydrodynamic modulations to accretion rates have been observed in simulations (e.g. Farris et al. 2014). Here, we describe models for both types of variability mechanisms. Throughout our analysis, each variability mechanism is considered entirely independently, i.e. we do not consider systems which may be both hydrodynamically and Doppler variable.

The physical scenario is depicted schematically in Fig. 5, assuming thin-disc geometries and a mass ratio $q \gtrsim 10^{-2}$ (see Section 2.4.2). On the left, the two MBHs, each with a circum-single disc, are shown within a cavity (‘gap’) of material evacuated by the binary orbit that separates them from the circumbinary disc. The radii of the circum-single discs are determined by the Hill radius of each MBH (see equation 6). Despite the presence of the gap, the circumbinary disc continues to transport angular momentum outwards, requiring material to be accreted inwards. Material from the disc overflows across the gap as accretion streams on to each circum-single disc. The orbital period at the inner edge of the circumbinary disc is longer than that of the binary, causing periodic variations in the accretion rate on to each circum-single disc (Fig. 5, upper right). The secondary MBH is farther from the centre of mass and closer to the circumbinary disc edge, which leads to it receiving a disproportionate fraction of the accreting material [as described by equation (1) and shown in Fig. 1]. Each circum-single disc, in addition to the circumbinary disc (although typically to a lesser extent), will produce AGN-like emission.

2.4.1 Doppler variability

Any component of the binary orbital-velocity along the observer’s line of sight will lead to both a relativistic boost and a Doppler shift in the observed spectrum of each circum-single disc. The boost is calculated using the Doppler factor, $D \equiv [\gamma(1 - v_{\parallel}/c)]^{-1}$, where the Lorentz factor $\gamma \equiv (1 - v^2/c^2)^{-1/2}$. The line-of-sight velocity, $v_{\parallel} \equiv v \sin(i)$, depends on the binary inclination i , which we define as zero for face-on systems. The observed flux is calculated as (D’Orazio et al. 2015b)

$$F_v = D^3 F'_v, \quad (15)$$

where the observed frequency ν is related to the rest-frame frequency as, $\nu = D \nu'$. Assuming a power law of index α_v for the section of the spectrum being observed, the Doppler variation in flux from the source will be (Charisi et al. 2018)

$$\frac{\Delta F_v^d}{F_v} = (3 - \alpha_v) \left| \frac{v}{c} \right| \sin i, \quad (16)$$

where v is the orbital velocity and c is the speed of light. The sensitivity of Doppler boosting to frequency and thus spectral shape offers a powerful method of testing it as a variability mechanism. D’Orazio et al. (2015b) have shown that, in both the optical and ultraviolet, this model explains the periodic variations observed in PG 1302-102.

In full generality, an AGN spectra may not be a power law at the frequency of interest, so we construct a full spectrum for each MBH in our simulations and numerically calculate the change in flux using equation (15). Additionally, the Doppler-boosting of each MBH in a binary is necessarily π out of phase, thus we determine the overall system variation as

$$\delta_F^{\text{dop}} \equiv \frac{\Delta F_{v,1}^d - \Delta F_{v,2}^d}{F_{v,1} + F_{v,2}}. \quad (17)$$

⁵The CRTS value we get from the cut-off in the flux distribution of candidates from Graham et al. (2015a), while the LSST value is from Ivezić et al. (2008). Our detection rates are not strongly dependent on the particular flux-threshold, as discussed in Section 3.3.

⁶We emphasize, however, that this degree of consistency is largely fortuitous, and even somewhat surprising as Illustris tends to overestimate the AGN luminosity function (e.g. Fig. 4).

⁷CRTS binary candidates are identified from SDSS spectroscopically confirmed AGN, which have a high completeness over roughly one-fourth of the sky (e.g. Richards et al. 2009). Studies suggest that through a combination of photometric and variability selection, LSST can achieve > 90 per cent completeness over its entire field of view (e.g. Schmidt et al. 2010; Butler & Bloom 2011; MacLeod et al. 2011) – roughly half of the sky, or twice that of SDSS (and thus the CRTS sample).

To handle the inclination dependence, for each simulated binary we calculate the SNR based on the un-boosted flux, F_v , and determine the minimum observable inclination $i_{\min}$, such that the variability is observable. The fraction of solid angles at which the system is observable, which, for randomly oriented inclinations is $\cos(i_{\min})$, then contributes linearly to the observability fraction, i.e.

$$f_{\text{obs}}^{\text{dop}} = \delta(F_v \geq F_{v,\text{sens}}) \cdot \delta(\delta_F \geq \delta_{F,\text{sens}}) \cdot \cos(i_{\min}). \quad (18)$$

2.4.2 Hydrodynamic variability

Periodic variations in accretion rates are frequently observed in hydrodynamic simulations of circumbinary discs (e.g. Artymowicz & Lubow 1994, 1996; Hayasaki, Mineshige & Sudou 2007; Roedig et al. 2012; D’Orazio, Haiman & MacFadyen 2013; Farris et al. 2014; Muñoz & Lai 2016). While significant uncertainties remain in understanding these accretion flows, the general pattern emerging is that three distinct mass-ratio regimes exist. For extreme mass ratios, $q \lesssim q_{\min}$, where $q_{\min} \sim 10^{-2}$, the secondary is a minor perturbation to the circumbinary disc, and the accretion flow remains steady. At intermediate mass ratios, $q_{\min} \lesssim q \lesssim q_{\text{crit}}$, where $q_{\text{crit}} \approx 1/3$, a gap is opened by the secondary and the accretion rate on to it varies by a factor of order unity, on the binary orbital period (D’Orazio et al. 2016).

For near-equal mass ratio systems ($q \gtrsim q_{\text{crit}}$), a highly distorted cavity is evacuated around the binary, out to roughly twice the binary separation. At the outer edge of the cavity, a significant over density of material develops (see also Shi & Krolik 2015; Noble et al. 2012). The Keplerian orbital period of that over density sets the variation time-scale as five to six times the binary period. The binaries we are considering (i.e. $M > 10^6 M_{\odot}$, $\tau_{\text{orb}} \sim \text{yr}$) are almost always in the GW-dominated regime in which the hardening time-scale – the duration a given binary spends at that separation – decreases rapidly with decreasing orbital period. Thus, if a given variational time-scale is probing binaries at shorter periods, the number of observable systems decreases (D’Orazio et al. 2015a).

We assume that hydrodynamic variability takes place for binaries in the thin-disc state, i.e. $f_{\text{Edd}} \geq 10^{-2}$. While fluctuations in the accretion rate are also likely to occur for ADAF discs, the simulations exploring this phenomenon (e.g. D’Orazio et al. 2016) are primarily applicable to thin discs. Modelling only variations in high-accretion rate systems has negligible effects on our overall results, as the low accretion rate systems are much harder to observe. Based on Farris et al. (2014), we assume for our fiducial models that all binaries with mass ratios above $q_{\min} = 0.05$ exhibit hydrodynamic variability, and those above $q_{\text{crit}} = 1/3$ are observable at $\tau_{\text{var}} = 5 \tau_{\text{orb}}$. The accretion rate variations in simulations predominantly affect the secondary MBH (e.g. Farris et al. 2014), so we model the overall hydrodynamic variations as

$$\delta_F^{\text{hyd}} \equiv \frac{\Delta F_{v,2}^h}{F_{v,1} + F_{v,2}} = \frac{F_{v,2}(M_2, \chi f_{\text{Edd},2}) - F_{v,2}(M_2, f_{\text{Edd},2})}{F_{v,1} + F_{v,2}}, \quad (19)$$

where we take the effective enhancement to the accretion rate as $\chi = 1.5$. In Section 3.3, we explore alternative values of χ and the importance of the τ_{var} assumption.

3 RESULTS

Flux variability amplitudes for a grid of mass ratio and total mass are shown in Fig. 6 for both Doppler (left) and hydrodynamic variability (right). Here, we use fixed values of orbital period: $\tau_{\text{orb}} = 3.14 \text{ yr}$, the system accretion rate: $f_{\text{Edd},\text{sys}} = 10^{-1}$ and redshift: $z = 1$. The

accretion rate on to the secondary MBH, determined by the accretion partition function (equation 1), is illustrated by the green colour bar. For both variability mechanisms, the secondary’s variations tend to dominate across the parameter space, although for shorter periods and very high total mass systems, the primary can occasionally dominate.

Fig. 6 shows a sharp discontinuity in variability amplitudes at $q \approx 2.5 \times 10^{-3}$, corresponding to the point at which the secondary’s accretion rate drops below $f_{\text{Edd}} = 10^{-2}$, and its disc transitions to the ADAF state. Note that the location at which this transition occurs is sensitive to the two simulated data points constraining the fitting function in this $q \ll 1$ domain (see Fig. 1). In the case of hydrodynamic variability, recall that our model considers only variations occurring in the thin-disc state, and additionally, only systems with $q \geq q_{\min} = 0.05$ are considered as variable in our fiducial configuration. Systems near this transitional mass ratio will alternate between thin and ADAF discs, which produces very high variability amplitudes.⁸ Doppler variability shows a similar, although more mild, discontinuity at the same transition mass ratio due entirely to the drop in optical luminosity of the secondary as it transitions to the ADAF state.

Doppler variability amplitudes are significantly larger for higher total mass systems that have larger orbital velocities. At the largest total masses, however, truncation of the secondary’s accretion disc becomes significant, and overall variability amplitudes again decline. Besides the mass ratio and accretion rate cut-offs imposed in our model, the hydrodynamic variability amplitudes have no explicit dependence on total mass or mass ratio. The trends seen in the right-hand panel of Fig. 6 are instead due to the relative brightness of the secondary compared to that of the primary and circumbinary disc. At large total masses, the secondary’s disc is significantly truncated, which decreases its luminosity and thus the overall variability amplitude. When most of the emitting region of the secondary’s accretion disc is preserved (i.e. at lower total masses), the relative accretion rate on to the secondary determines its brightness and the variability amplitude.

The total luminosity of the system is tied closely to its total mass. Near $q \approx 0.1$, the accretion rate on to the secondary is enhanced by more than the inverse mass ratio, leading to the secondary outshining the primary. Near this mass ratio sweet spot, systems remain observable even at lower total masses. The hatched regions of Fig. 6 show systems with V-band spectral fluxes below the LSST (circles) and CRTS (dots) thresholds for systems at a fixed luminosity distance of $d_L(z=1) \approx 6.5 \text{ Gpc}$. At this distance, systems with $f_{\text{Edd}} = 10^{-1}$ can be seen down to $M \approx 2 \times 10^8 M_{\odot}$ and $M \approx 3 \times 10^6 M_{\odot}$ for CRTS and LSST, respectively.

3.1 Event rates

The number of detectable, periodically variable MBHBs are shown as a function of survey sensitivity and redshift in Fig. 7. The detection of Doppler variable binaries (left-hand panel) becomes plausible only for sensitivities $m_v \gtrsim 21$. Hydrodynamic variables are much more common, and models predict that sources could be identifiable even near $m_v \approx 18.5$. At low sensitivities, observable sources from both models of variability are likely to appear near $z \approx 0.1$. With much higher sensitivities, sources could be identified out to $z \approx 1.0$, but likely not much farther. To study the observability of variables in more detail, we focus on two sensitivities representative

⁸This is the case even allowing for variability in the ADAF state (not shown).

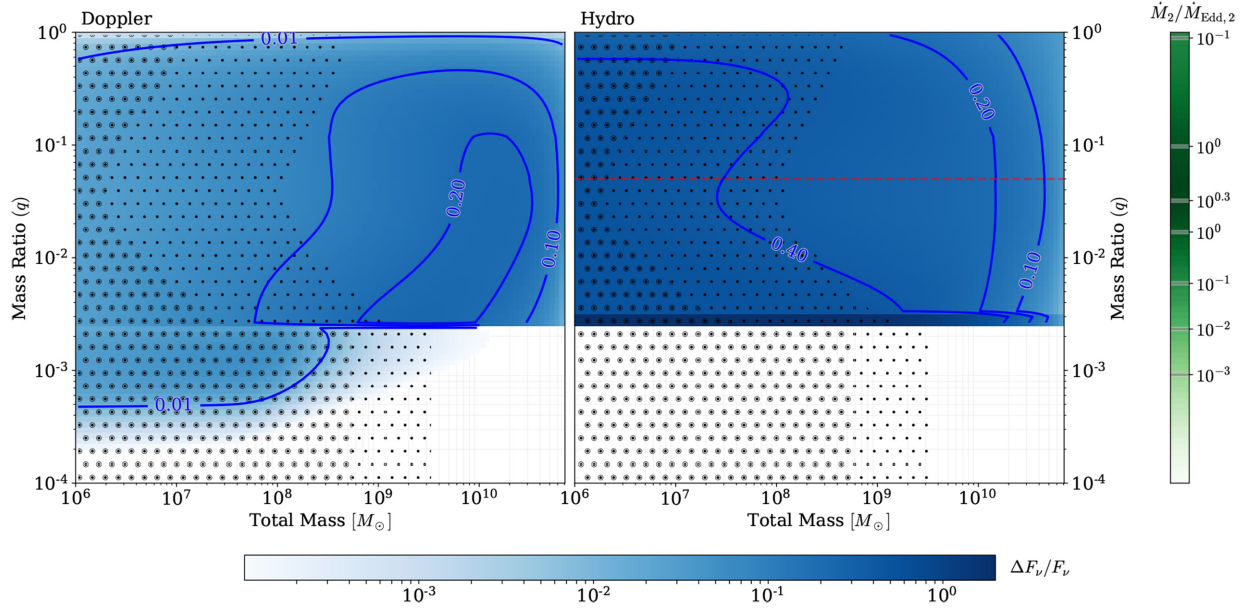


Figure 6. Variability amplitudes in total-mass–mass-ratio space for Doppler and hydrodynamic variability. In this example, we use a fixed orbital period of $\tau_{\text{orb}} = 3.14$ yr and a system accretion rate of $f_{\text{Edd,sys}} = 10^{-1}$. The accretion rate on to the secondary is shown in the green colour bar to the right. The hashed regions correspond to luminosities below the detection threshold for LSST (larger circles) and CRTS (smaller dots) at a uniform redshift $z = 1$. Note that fiducially, based on the results of Farris et al. (2014), only systems with $q > q_{\text{min}} = 0.05$ (red, dashed line) are considered as variables, although here we show variability at lower mass ratios. The contours correspond to the indicated variability amplitudes.

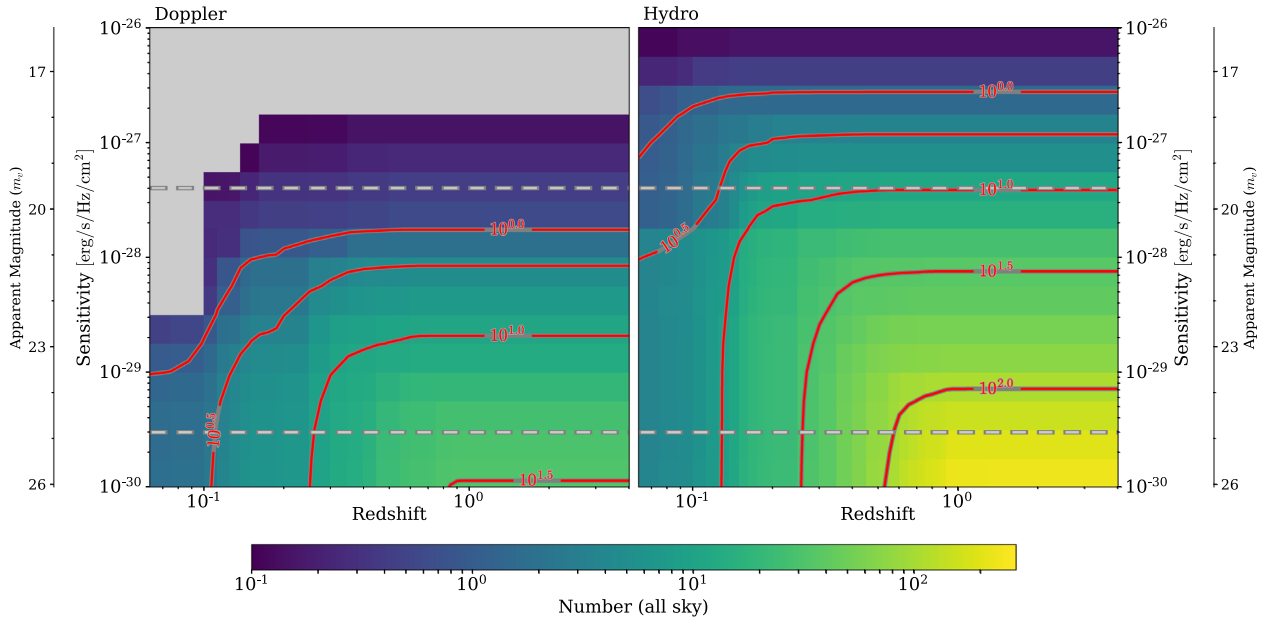


Figure 7. Cumulative number of detectable, periodically variable MBHBs in sensitivity versus redshift space. Doppler variable binaries (left) are considered independently from hydrodynamic ones (right). The plotted values are all-sky detection rates for binaries with observer-frame periods between 0.5 and 5.0 yr. The sensitivities of CRTS, $F_{\text{v,sens}}^{\text{CRTS}} = 4 \times 10^{-28} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ cm}^{-2}$ ($m_v \approx 19.5$), and LSST, $F_{\text{v,sens}}^{\text{LSST}} = 3 \times 10^{-30} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ cm}^{-2}$ ($m_v \approx 24.9$), are marked with dashed, grey lines.

of CRTS and LSST, $F_{\text{v,sens}}^{\text{CRTS}} = 4 \times 10^{-28} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ cm}^{-2}$ ($m_v \approx 19.5$) and $F_{\text{v,sens}}^{\text{LSST}} = 3 \times 10^{-30} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ cm}^{-2}$ ($m_v \approx 24.9$), respectively, which are marked in Fig. 7 with dashed, grey lines. At the sensitivities of CRTS, our models predict 0.5 and 10 Doppler and hydrodynamic variables, respectively, to be observable on the full sky; for LSST, those numbers increase substantially to 30 and 200.

We compare the populations of AGN, MBHBs, and observably variable binaries in Fig. 8. The black curves show the distribution of AGN, while all binaries are shown in purple, and those above the flux limit of each instrument are shown in red. Binaries that are observably variable are plotted in light and dark blue for Doppler and hydrodynamic variability, respectively. The most accurate predictions from our models are likely the number of binaries per

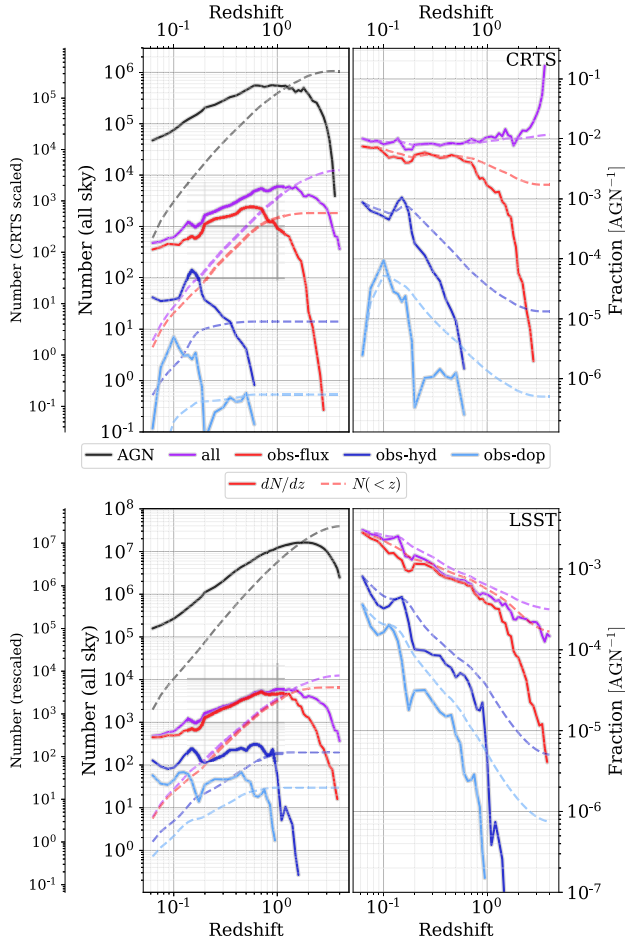


Figure 8. Observability of MBHBs versus redshift for CRTS (top) and LSST (bottom). Observable AGNs are shown in black, while all binaries are in purple, and those above the flux limit of each instrument are in red. Binaries with observable variability are shown in light blue for Doppler variables, and dark blue for hydrodynamic variables. The left-hand panels give the absolute number of systems in the full sky, while the right-hand panels give the number of binaries as a fraction of observable AGN. Differential distributions are shown with solid lines, and cumulative ones with dashed lines. Values for binaries include all systems with observer-frame periods between 0.5 and 5.0 yr.

observable AGN (right-hand panels) that should reduce systematic uncertainties in the typical luminosities of our simulated AGN and binaries. For convenience, we also include a second, left y-axis with the predicted number of sources rescaled to the number of studied, spectroscopically confirmed AGN in CRTS.⁹

From Fig. 8, we can see that the number of all binaries (purple) closely traces that of AGN, except two orders of magnitude fewer. For $z \lesssim 0.6$, our models predict that most MBHBs are above the flux limit for CRTS (top, red) which implies that in the order of 1 per cent of AGN out to similar distances could be in binaries – $\approx 10^3$ systems. The number of observably bright binaries falls off rapidly above $z \approx 1$ for CRTS, and above $z \approx 1.5$ for LSST. In the full volume ($z \approx 4.0$), the fraction of AGN in binaries decreases to about 10^{-3} for CRTS and 10^{-4} for LSST.

⁹In other words, the ‘CRTS scaled’ y-axis is shifted such that the total number of predicted AGN matches the observed 3.3×10^5 of CRTS (see equation 13).

Out to a redshift of $z \approx 0.2$, roughly 10 per cent of binaries observable above the CRTS flux limit are also identifiable as hydrodynamic variables (dark blue), corresponding to $\approx 10^{-3} \text{ AGN}^{-1}$ hydrodynamically variable binaries. This decreases to $\approx 10^{-4.5} \text{ AGN}^{-1}$ at $z \approx 1$, and $\approx 10^{-5} \text{ AGN}^{-1}$ in the full volume. Scaling to the number of monitored CRTS AGN, our models predict five hydrodynamically variable binaries should be visible. For LSST, ≈ 25 per cent of binaries above the flux limit are also identifiable as hydrodynamically variable systems to $z \approx 0.2$ and ≈ 10 per cent to $z \approx 1$. The total number of systems with hydrodynamic variability above the LSST threshold is predicted to be $\approx 2 \times 10^2$ in the full sky, or ≈ 100 assuming twice the completeness of CRTS (i.e. complete for roughly half of the sky).

The number of binaries with Doppler variability above the variability threshold (light blue) is drastically fewer than hydrodynamic ones, as only binaries with nearly relativistic velocities produce sufficiently large brightness modulations. Our models predict that, optimistically, of order unity Doppler variable binaries should be present in the Graham et al. (2015a) CRTS data set, specifically an expectation value of 0.5 all-sky, and 0.2 after rescaling to the number of CRTS AGN. While Doppler variables are expected at a rate of $\sim 10^{-6} \text{ AGN}^{-1}$ for both CRTS and LSST, this corresponds to a much more promising number of LSST sources: 30 Doppler variable binaries all-sky, or 20 after rescaling.

The predicted number of detectable variables are shown versus period in Fig. 9, grouped in bins of different total masses. The horizontal dashed lines provide an estimate of the minimum plausibly-observable event rates. For CRTS, this is the inverse of the number of monitored AGN from the variability survey, and for LSST we again rescale to twice the completeness of CRTS (see equation 13). The shaded region highlights a decade of period between 0.5 and 5.0 yr, corresponding to the values plotted Fig. 8 and representative of the identified CRTS candidates. The number of binaries declines sharply with decreasing orbital period for all masses, reflecting the GW-hardening time-scale, $\tau_{\text{gw}} \propto \tau_{\text{orb}}^{8/3}$, by which nearly all binaries at these periods will be dominated (see e.g. Haiman et al. 2009; Kelley et al. 2017a).

The most obvious feature of Fig. 9 is that binaries with both types of variability, and observed by either current CRTS-like instruments or even future LSST-like surveys, will be dominated by systems at the longest orbital periods. This is unfortunate as the same is naturally expected from red-noise contaminated single AGN.¹⁰ Our models suggest that Doppler variable systems are only marginally observable by CRTS, with sources only likely to occur near and above $p \approx 5$ yr. LSST on the other hand will likely detect systems down to $p \approx 2$ yr, where multiple orbital cycles could be observed, thereby decreasing the chance of red-noise contamination. Hydrodynamic variables are far more common, and could be observed down to $p \approx 1$ yr by CRTS and $p \approx 0.5$ yr by LSST.

Doppler variability depends on the orbital velocity, and thus favours larger total masses at a fixed orbital period. This is clearly seen in Fig. 9, where systems detectable by CRTS are dominated by those in the 10^8 – $10^9 M_\odot$ and 10^9 – $10^{10} M_\odot$ mass bins. Because LSST is much more sensitive, it begins to probe lower total masses,

¹⁰Most AGN variability is well fit by a damped random walk, i.e. red-noise at high frequencies, but flattening to white below some critical frequency. While the transition is typically at inverse months, the distribution has tails extending to many years containing binaries which would still seem red (MacLeod et al. 2010).

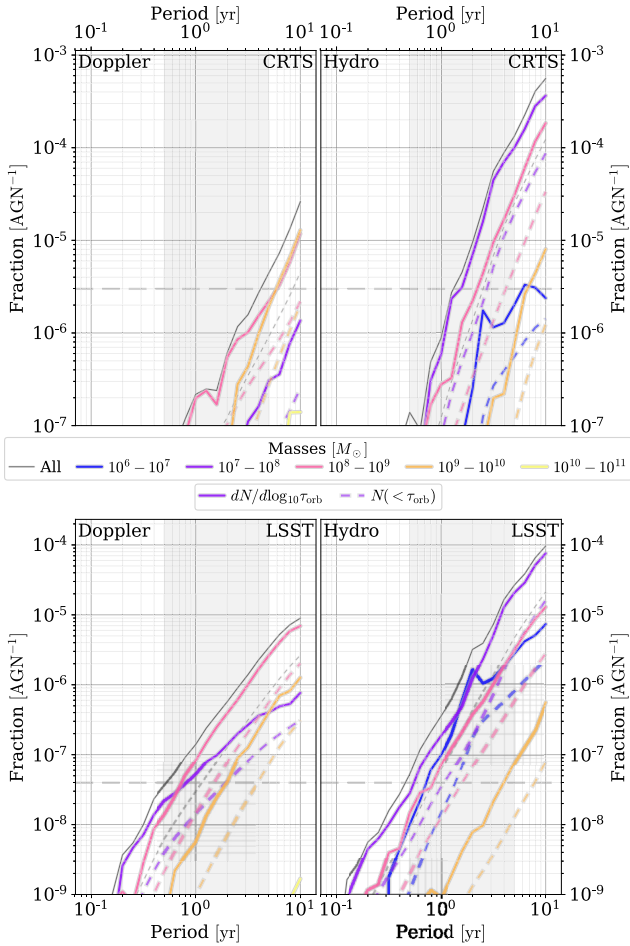


Figure 9. Observability of variable MBHBs versus period for CRTS (top) and LSST (bottom). Doppler variable systems are shown in the left column, while hydrodynamic variables are shown on the right. Binaries are broken into bins of different total masses indicated by the different coloured lines, while the distribution of all binaries is in black. Solid lines indicate distributions per log-decade of period, while dashed lines are cumulative. The grey, horizontal dashed lines provide an estimate of the minimum observable occurrence rates, scaled to the CRTS completeness and twice that for LSST.

especially at smaller periods where the orbital velocity increases at fixed mass. Apparent in the Doppler variable LSST population (bottom left), lower mass systems are relatively more observable at shorter periods, where their velocities are larger. Hydrodynamic variability has no explicit total mass dependence, and thus those binaries tend to be dominated by the much more numerous, lower mass systems. Hydrodynamic variables are still not representative of all binary masses as the more massive systems can be more luminous, and thus observed in a larger volume.

3.2 Observation efficiency

To get a better sense of the fraction of binaries that are detectable, and their parametric dependences, we plot detection efficiency versus total mass and mass ratio in Fig. 10. The top panels, which show the number of binaries per AGN, demonstrate that the overall number of binaries falls quickly with increasing total mass (left), and thus also with decreasing mass ratio (right) as these quantities are strongly inversely correlated. Not only is the mass function of

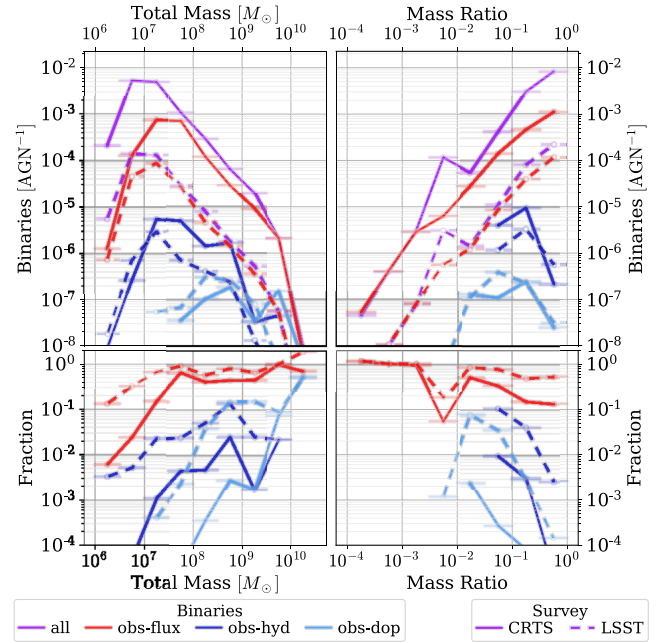


Figure 10. Detection efficiency of MBHBs as periodic variables in bins of total mass (left) and mass ratio (right) for sensitivities of both CRTS (solid) and LSST (dashed). The top panels show binary occurrence rates per observable AGN, and the bottom panels show the fraction of binaries that are observable: above each survey's flux limit (red), and hydrodynamically variable (dark blue) and Doppler variable (light blue) above each survey's variability threshold. Values are for binaries at all redshifts with observer-frame periods between 0.5 and 5.0 yr.

MBH sharply decreasing with mass, but also for a fixed orbital period, higher mass systems harden faster – further decreasing their number.¹¹ The lowest total mass bin is an outlier, with noticeably fewer binaries due primarily to the mass cut at $10^6 M_\odot$, but also to the difficulty for low-mass systems to merge effectively and reach the binary separations corresponding to $\sim$ yr periods.

The lower panels of Fig. 10 show the number of observable binaries (red and blues) divided by the total number of binaries in each bin. This highlights that while observable binaries tend to follow the distribution of all binaries, their completeness drops significantly at lower masses. Doppler variable binaries (light blue) have an especially strong total-mass dependence, such that the efficiency of CRTS falls from $\approx 10^{-1}$ at $10^{10} M_\odot$ to well below 10^{-2} at $10^9 M_\odot$. LSST maintains an efficiency of $\approx 10^{-1}$ down to roughly $10^{8.5} M_\odot$ before dropping. While the peak recovery fraction of hydrodynamic variables (dark blue) is no better than for Doppler ones, the orders of magnitude higher efficiency for the most numerous binaries at $\lesssim 10^8 M_\odot$ leads to vastly more, detectable hydrodynamically variable binaries overall.

Both variability mechanisms have strong and similarly shaped mass ratio dependence, but largely for different reasons. Our models assume that for mass ratios $q < q_{\min}$ (fiducially, $q_{\min} = 0.05$), there is no hydrodynamic variability as the secondary acts as a minor perturber (see Farris et al. 2014). For mass ratios $q > q_{\min}$, however, there is no dependence in our model between the amplitude of variability and binary mass ratio, though a strong trend is evident in

¹¹While this point does not affect our results, note again that some studies find that more massive binaries do not harden faster.

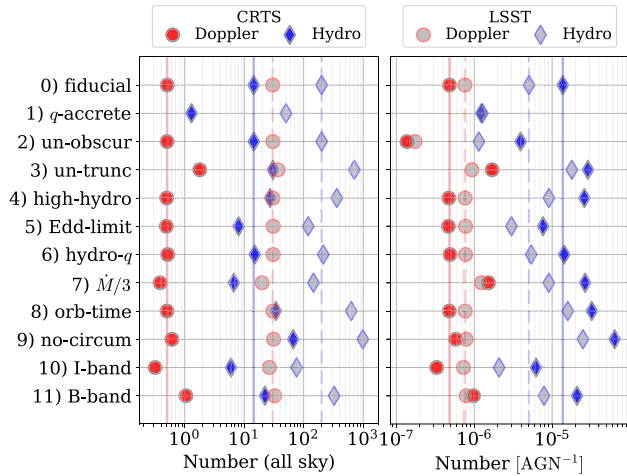


Figure 11. Comparison of detection rates for parametric changes from our fiducial model for CRTS (coloured faces) and LSST (coloured edges). The left-hand panel shows all-sky expected detection rates, and the right-hand panel shows rates normalized to the number of observable AGN by each instrument. Red points indicate Doppler variables, while blue points indicate hydrodynamic ones. The vertical lines denote the detection rates of the fiducial models. See the surrounding text for a description of each alternative model. The change in detection rates is typically well under an order of magnitude, except for the ‘ q -accrete’ model that eliminates all Doppler detections and decreases hydrodynamic ones by $\approx 10\times$ and $\approx 3\times$ for CRTS and LSST, respectively.

the detection efficiencies (dark blue, the lower right-hand panel of Fig. 10). While mass ratio does not affect the amplitude of variability, at high mass ratios, $q > q_{\text{crit}}$ (fiducially, $q_{\text{crit}} = 0.3$), the period of variability is taken to be five times longer than the orbital period. This means that high mass ratio hydrodynamic variables observed at $p \sim 5$ yr, for example, are actually produced by the much smaller population of systems at $p \sim 1$ yr. That shift in period leads to a significant drop in the number of hydrodynamic variables observed in the highest mass ratio bin. The tendency for higher mass ratio systems to be lower in total mass likely also contributes to the high mass ratio decline, as evidenced by a similar (though more subtle) decline for systems above the flux limit (red).

Binaries are easier to observe as Doppler variables when they have larger orbital velocities and thus smaller mass ratios at a fixed total mass. At the same time, at very low mass ratios, the variabilities that are produced by the secondary become washed out by the brighter primaries. This leads to a strong peak in the detection efficiency of Doppler variables with mass ratio near $q \approx 0.03$ – 0.07 . For both Doppler and hydrodynamic variability, there is also a boost for systems near $q \approx 0.05$ where the accretion rate on to the secondary peaks (see Fig. 1).

3.3 Model effects and parameters

Our models of binary AGN variability include a variety of effects and uncertain parameters. Fig. 11 compares our fiducial detection rates (top row and vertical lines) with variations from different parametric changes to our models. Each row gives the expected all-sky (left) and per-AGN (right) detection rates for each alternative configuration. Doppler variables are shown in red and hydrodynamic variables in blue. The variations we consider are meant to illustrate both parametric uncertainties and the significance of

particular physical effects. Each alternative model is described below:

(i) ‘ q -accrete’: The accretion rate of each component MBH is scaled proportionally to its mass, i.e. $\lambda \equiv \dot{M}_2/\dot{M}_1 = q$. While the relative accretion rate has been studied extensively in the planetary and binary communities, a consensus has not been reached. For both Doppler and hydrodynamic variability, the secondary MBH is almost always the source of observable periodicity, making the particular form of the accretion partition function (Fig. 1) important. We test this, very simple model to explore our sensitivity to the accretion function. Scaling the accretion rate of each MBH to its mass produces significantly lower secondary accretion rates than using the Farris et al. (2014) model. Because Doppler variable sources are preferentially lower mass ratio, the decreased accretion rate on to the secondary means that its variability is hidden by the brighter primary, and effectively never detectable. The rate of hydrodynamic variable detections is decreased by an order of magnitude for CRTS and roughly a factor of 3 for LSST.

(ii) ‘un-obscur’: Neglecting AGN obscuration. Interestingly, the overall number of detectable variables is virtually unaffected by the presence of obscuration, while, without obscuration, the total number of observable AGN is increased by a factor of a few. This highlights the fact that the limiting condition for observing binaries photometrically tends to be in the variability sensitivity as opposed to the overall flux sensitivity. Improving the minimum detectable variation amplitudes (i.e. the floor variability sensitivity, $\delta_{F,\text{min}}$) could significantly increase the number of binary detections, even at a fixed flux sensitivity.

(iii) ‘un-trunc’: Each MBH’s accretion disc extends to indefinitely, instead of being truncated by the presence of the companion. By definition this model only affects the observability of binaries and not single AGN. All binary detection rates are increased by the presence of larger emitting regions, although Doppler variable binaries detected by LSST are only slightly increased, implying that the intrinsic variability amplitude is the limiting factor and not the brightness of the secondary. Doppler and hydrodynamic detections by CRTS are increased by about a factor of 4 and 2, respectively, while LSST hydrodynamic detections are increased by roughly a factor of 3. This model demonstrates the importance of disc truncation. While AGNs in the ADAF state (and thus, typically at low accretion rates) are unlikely to contribute significantly to the detectable population, better understanding if and how truncation occurs will be important in determining their relative observability.

(iv) ‘high-hydro’: The amplitude of hydrodynamic variability is increased from $\chi = 1.5$ to 2.0 . In the simulations of Farris et al. (2014), the authors find different variability amplitudes for different mass ratios, generally varying from ≈ 1.5 to ≈ 3.0 . In this model, we double the brightness increase that occurs from hydrodynamic variability and find that it produces only a moderate increase in hydrodynamic detections, ≈ 90 per cent. The systems that become detectable in this model are typically in either small mass ratio binaries, or at lower periods, where truncation of each MBH’s circum-single disc hampers its observability.

(v) ‘Edd-limit’: The accretion rate in each circum-single disc is limited to Eddington. In our fiducial model, only the overall accretion rate to both MBHs is Eddington limited. Because the secondary accretion rate is larger than the primary’s for $q \sim 0.1$, it can exceed the Eddington limit individually. Conceptually, the ‘Edd-limit’ model assumes that the gas inflow is regulated effectively even at small scales, or alternatively that the radiative efficiency

does not increase for Eddington fractions above unity.¹² Estimated Doppler variable detections by CRTS and LSST are both decreased negligibly. Hydrodynamic variable detection rates are decreased by ≈ 50 per cent for both instruments. Hydrodynamically variable systems are more sensitive to this limit likely due to their tendency to be at slightly higher mass ratios ($q \approx 0.1$), which are closer to the peak of the accretion partition function.

(vi) ‘hydro- q ’: The minimum mass ratio for variability is set to $q_{\min} = 0.0$ (fiducial $q_{\min} = 0.05$). Because the contribution from the secondary MBH in lower mass ratio systems is very small, they tend to be unobservable even if they produce hydrodynamic variability, and thus this model shows an effectively negligible change to detection rates.

(vii) ‘ $\dot{M}/3$ ’: The accretion rates from Illustris are uniformly scaled down by a factor of 3. The AGN luminosity function constructed from Illustris is generally consistent with observations, but noticeably overpredicts the number of observable systems. The luminosity functions fit better if the accretion rates (or radiative efficiencies) are systematically decreased by ≈ 50 per cent, but in that case the predicted number of AGN detectable by CRTS becomes too low. This model attempts to quantify the uncertainty in the normalization of the Illustris luminosity function by systematically and significantly decreasing MBH accretion rates.

For CRTS, the all-sky predicted number of identifiably variable systems decreases from 0.51 to 0.39 (Doppler) and 14 to 6.7 (hydrodynamic). At the same time, the predicted number of CRTS detectable AGN decreases from 1.1×10^6 to 2.5×10^5 , which is inconsistently low compared to the number monitored by CRTS. Rescaling to the actual number of CRTS observed AGN leads to an increase in the predicted variable detection rates: from 0.16 to 0.51 (Doppler) and 4.6 to 8.9 (hydrodynamic). Systematically decreasing accretion rates leads to fewer observable variables in our simulated sky, and far fewer AGN. Assuming that the number of observable variables *per AGN* is the most accurate predictor, however, actually leads to an increase in predicted detection rate for CRTS and LSST from the low accretion-rate model.

(viii) ‘orb-time’: Hydrodynamic variability always occurs at the orbital period, instead of shifted to longer periods for $q > q_{\text{crit}} = 0.3$. Some hydrodynamic simulations suggest that high mass ratio systems will exhibit variability at periods roughly five times longer than the binary period, corresponding to the orbital time of an overdensity of material at the outer edge of the gap in the disc. For these systems, observed periodicity at ~ 1 yr is actually coming from sources with orbital periods of ~ 0.2 yr, which are far less numerous. If variability always occurs at the orbital period, the number of hydrodynamically variable detections double for both CRTS and LSST. Specifically, rescaled detection rates increase from 4.6 to 11 for CRTS, and from 130 to 400 for LSST. It is worth noting that in the Farris et al. (2014) simulations, while variations are *predominantly* at $\approx 5 \times$ the orbital period, there is still a component directly at the orbital period itself which could be identifiable for sufficiently high SNR systems. To be conservative, we use the time-shifted configuration as our fiducial model.

(ix) ‘no-circum’: Emission from the circumbinary portion of the accretion disc is neglected. Without the circumbinary disc, fractional brightness variations increase as there is less steady

emission to compete with. This increases Doppler detection rates by only ≈ 20 per cent for CRTS and negligibly for LSST, but increased hydrodynamic detection rates by almost a factor of 4 for both CRTS and LSST. Circumbinary emission seems to primarily hamper the detection of near equal-mass binary systems, which are significantly less important for Doppler detection rates. This model highlights that circumbinary emission is an important source of flux to consider. Additionally, recent simulations suggest that the circumbinary portion of the disc can also contribute to periodic variability signatures (Tang, Haiman & MacFadyen 2018) which is not considered here.

(x) ‘ B and I bands’: Spectra are sampled in the rest-frame B and I bands instead of the V band. For both CRTS and LSST, more variables are detectable in the B band and fewer in the I band. Doppler detections by CRTS increase by almost a factor of 2, and negligibly for LSST, while hydrodynamic detections increase by ≈ 40 per cent for both instruments. Keeping in mind that our models do not take into account factors such as colour-dependent extinction, this effect seems to be driven by the AGN circum-single discs being intrinsically brighter on the bluer side of the spectrum.

4 DISCUSSION

In this paper, we make predictions for the electromagnetic detection of MBHBs as photometrically periodic AGN. We use a population of binaries drawn from cosmological hydrodynamic simulations of MBHs and galaxies, evolved using semi-analytic, post-processing models of the detailed merger physics. Employing synthetic AGN spectra, along with models of both Doppler and hydrodynamic variability, we have calculated detection rates for the flux and variability sensitivities of CRTS and LSST. Here, we present the results and implications of our study, after first discussing their limitations.

4.1 Caveats

Numerous limitations exist in our current methods, both in terms of our binary populations and our models of variability. In the former class, while the masses of MBHs in Illustris nicely reproduce the observed, redshift zero BH–galaxy scaling relations (Sijacki et al. 2015), there is still significant uncertainty in the full distribution of MBH masses (e.g. McConnell & Ma 2013). The MBH accretion rates have been calibrated to produce accurate masses and reproduce observational, bolometric luminosity functions (Sijacki et al. 2015). Using spectral models, and a simple model of obscuration, applied to the entire population of Illustris MBHs, we predict a total number of observable AGN (see Table 1) that are consistent with CRTS observations, but the AGN luminosity function from our model overpredicts the observed relation from Hopkins et al. (2007, reproduced in Fig. 4). While we overestimate the luminosity function even for the brightest systems, we still suffer from small number statistics and incompleteness in the most massive MBHs, due to the finite volume of the Illustris box.

Instead of using bolometric luminosities and corrections, or characteristic spectral indices, we have constructed full spectral models for each of the MBHs in our binary populations. Still, these spectra are highly simplified in the complex and actively developing field of AGN emission. Perhaps the most important deficiency of our spectral models are the lack of any lines, colour-dependent extinction, or non-thermal contributions from outside of the disc. We also do not consider any intrinsic AGN variability. Full spectroscopic observations of AGN, including variations between

¹²Some theoretical and numerical work have shown that the radiative efficiency increases only logarithmically for super-Eddington accretion rates in the ‘slim’-disc regime (e.g. Jaroszynski et al. 1980; Abramowicz 2005; Sądowski 2009).

Table 1. Expected observability of MBHB, periodically variable AGN. The first four columns give the expected detection rates of Doppler and Hydrodynamic variable binaries with periods between 0.5 and 5.0 yr. The last two columns give the expected detection rates for (single) AGN. Each cell includes in parenthesis the range of values from the models discussed in Section 3.3. The first row is the all-sky prediction from our simulations, while the second row is normalized to the predicted number of detectable AGN for each instrument. The last row rescales our results based on the number of AGN monitored in the CRTS variability study ($\approx 3.3 \times 10^5$; Graham et al. 2015a), and assuming twice that completeness for LSST.

Number	Observable, variable binaries				AGN	
	Doppler		Hydrodynamic		CRTS	LSST
	CRTS	LSST	CRTS	LSST	CRTS	LSST
All-sky	$5 \times 10^{-1} \left(\frac{2 \times 10^{+0}}{3 \times 10^{-1}} \right)$	$3 \times 10^{+1} \left(\frac{4 \times 10^{+1}}{2 \times 10^{+1}} \right)$	$1 \times 10^{+1} \left(\frac{7 \times 10^{+1}}{6 \times 10^{+0}} \right)$	$2 \times 10^{+2} \left(\frac{1 \times 10^{+3}}{8 \times 10^{+1}} \right)$	$1 \times 10^6 \left(\frac{4 \times 10^6}{3 \times 10^5} \right)$	$4 \times 10^7 \left(\frac{2 \times 10^8}{2 \times 10^7} \right)$
AGN ⁻¹	$5 \times 10^{-7} \left(\frac{2 \times 10^{-6}}{1 \times 10^{-7}} \right)$	$8 \times 10^{-7} \left(\frac{1 \times 10^{-6}}{2 \times 10^{-7}} \right)$	$1 \times 10^{-5} \left(\frac{6 \times 10^{-5}}{4 \times 10^{-6}} \right)$	$5 \times 10^{-6} \left(\frac{2 \times 10^{-5}}{1 \times 10^{-6}} \right)$	$1 \times 10^0 (-)$	$1 \times 10^0 (-)$
Scaled	$2 \times 10^{-1} \left(\frac{6 \times 10^{-1}}{5 \times 10^{-2}} \right)$	$2 \times 10^{+1} \left(\frac{5 \times 10^{+1}}{6 \times 10^{+0}} \right)$	$5 \times 10^{+0} \left(\frac{2 \times 10^{+1}}{1 \times 10^{+0}} \right)$	$1 \times 10^{+2} \left(\frac{6 \times 10^{+2}}{4 \times 10^{+1}} \right)$	$3 \times 10^5 (-)$	$3 \times 10^7 \left(\frac{4 \times 10^7}{3 \times 10^7} \right)$

observing epoch, should be incorporated into calculations to carefully consider the effectiveness with which periodic photometric variability can be accurately classified.

We have also relied very strongly on the results of Farris et al. (2014), which use 2D, isolated, purely-hydrodynamic simulations. Other groups (e.g. Cuadra et al. 2009; Roedig et al. 2012; Shi et al. 2012) have supported the Farris et al. (2014) results whose conclusions seem robust for their simulated conditions. Accounting for thick-disc accretion (and mass flow out of the disc plane, possibly enhanced by magnetic fields and radiation), and turbulent flows with varying inflow rates on large scales, are likely to affect detailed predictions. Using the Farris et al. (2014) accretion rates also produces an inconsistency in our models: the large accretion rates to $q \sim 0.1$ binaries, combined with the typically large systemic accretion rates in Illustris, imply that binaries will grow towards $q \sim 1$ relatively quickly. Naively, the e-folding time for the secondary at these mass ratios is often as short as 10 Myr. At the same time, the accretion is typically super-Eddington in these systems, and whether those accretion rates accurately correspond to mass-growth rates (i.e. neglecting outflows) is unclear. If the secondary MBHs were allowed to grow self-consistently, it would likely decrease our predictions for both types of variables. At the same time, some studies in the context of planetary systems (e.g. Lubow, Seibert & Artymowicz 1999) find that even at relatively extreme mass ratios (e.g. $q \sim 10^{-3}$), the accretion rate on to the secondary can still be a significant fraction of the total accretion rate. If accretion rates remain high at low mass ratios, it could significantly increase the incidence of detectable systems.

We expect many of the simplifications and uncertainties in our models to tend towards *fewer* systems being observable as variables. For example, thick-disc and turbulent accretion flows are more likely to smooth out the periodic variations in emission rather than enhance them. Light from AGN host galaxies will also produce an additional background from which variability must be disentangled. AGNs are also known to exhibit strong intrinsic variability, especially at long periods that easily mimics periodicity. While our model for variability sensitivity is based on observational studies of signal identification, it is an extremely simplistic accounting of a very difficult task, as shown in the detailed analyses of Graham et al. (2015a), Charisi et al. (2016), and Liu et al. (2016) that include careful treatments of noise.

4.2 Conclusions

A summary of expected detection rates for variability periods from 0.5 to 5 yr are presented in Table 1. Our models predict that MBHBs should be detectable at rates of roughly 5×10^{-7} AGN⁻¹

and 10^{-5} AGN⁻¹ for Doppler and hydrodynamic variability, respectively. In our simulations, this corresponds to all-sky rates of 0.5 and 10 sources at the CRTS sensitivity, while scaling to the number of CRTS-monitored AGN yields 0.2 and 5 binaries. For the expected sensitivity of LSST, and assuming twice the completeness of AGN confirmation, we predict 20 and 100 Doppler and hydrodynamic binaries to be observable. Additional data are presented in Section A and online to facilitate detection-rate predictions for optical surveys with different sensitivities and durations.

Our predictions for current instruments are significantly lower than the rate of candidates put forward by Graham et al. (2015a) and Charisi et al. (2016). Our models suggest that having multiple Doppler variables are unlikely in CRTS, and only a small fraction of the candidates can be explained as hydrodynamic variables. This is consistent with GW limits that imply that the published candidates contain false positives (Sesana et al. 2018). But our predictions do indicate that there should exist numerous true MBHBs within the candidate populations. The systems painstakingly identified by Graham et al. (2015a) and Charisi et al. (2016) provide an extremely valuable opportunity to identify examples of MBHBs. The candidates deserve significant follow-up to find the binaries they contain, as no examples of gravitationally bound MBHBs have been confirmed to date. Additionally, the candidates put forwards are very convincing in demonstrating periodic variability *above that produced by the best-fitting models of intrinsic AGN variability*. The characterization and study of the mechanisms producing false positives thus presents an interesting opportunity to not only better identify binaries but also to explore the fundamental accretion processes at play.

Simply extending the temporal baselines of candidate observations will provide a determinant in distinguishing red-noise contaminated systems. Candidates exhibiting red-noise fluctuations misconstrued as periodicity are expected to eventually deviate from sinusoidal behaviour. Unfortunately, this test is not without issue. Disc turbulence and time-varying feeding rates of gas not only introduce their own luminosity fluctuations but also decrease the coherent, periodic variations from a binary. Some simulations also see accretion alternate from primarily feeding one MBH to then predominantly feeding the other (e.g. Dunhill, Cuadra & Dougados 2015; Bowen et al. 2018), even in otherwise smooth discs. These factors introduce a significant complication in separating AGN with significant red-noise from those which are binaries, but exhibit excursions from periodicity. While we find Doppler variables to be intrinsically very rare, the characteristic spectral dependence of their variations is a valuable test of their origin (D’Orazio et al. 2015b). Systems that may be seen edge-on, like Doppler variables, can also produce periodic lensing spikes (Haiman 2017;

D’Orazio & Di Stefano 2018). While our analysis does not consider identification of systems simultaneously exhibiting both Doppler and hydrodynamic variability, these phenomena could be coincident and observationally distinguishable. It is also worth considering ‘triggered’ searches, for example: from a candidate Doppler variable binary, a survey could search for signs of hydrodynamic variability at multiples of the orbital period with boosted SNR. Ultimately, we expect that time-varying spectroscopic features of binarity (e.g. Comerford et al. 2009; but see also Eracleous et al. 1997) may be the most robust identifier of spatially unresolved MBHBs.

Our results offer hints at what candidate system parameters may be most indicative of true binarity. We find that neither Doppler nor hydrodynamic variables are likely to be observed much beyond $z \approx 1$, as variability amplitudes are simply too low to be distinguished at larger redshifts. Out to a redshift 0.6, our models suggest that almost 1 per cent of AGN could harbour binaries, although only a small fraction are identifiable as such. To plausibly detect hydrodynamically variable binaries, a survey must have a sensitivity of $m_v \gtrsim 18.5$ and $m_v \gtrsim 21$ for Doppler variable binaries. Owing to the strong period dependence of the GW hardening rate, binaries are most likely to be observed at longer periods – unfortunately the same trend as for red-noise contamination.

The Doppler variability model depends on nearly relativistic velocities, which means that more massive binaries are strongly favoured. We expect both current and future instruments to detect Doppler variables predominantly above $10^8 M_\odot$, and LSST will likely see mostly 10^8 – $10^9 M_\odot$ binaries. Hydrodynamic variability is insensitive to the total binary mass, and thus is dominated by the far more numerous systems at lower masses. Below $\approx 10^7 M_\odot$, however, the lower luminosities begin to limit the sensitive volume. For both CRTS and LSST, hydrodynamically variable binaries should be mostly between $\approx 10^7$ and $10^8 M_\odot$, although systems between $10^{6.5}$ and $10^9 M_\odot$ should be observable, and higher mass systems are likely too rare.

Both Doppler and hydrodynamic variability strongly favour systems with mass ratios $q \sim 0.1$, in our fiducial model. In the case of Doppler variability, this trend is due to the larger orbital velocities of secondary MBHBs with lower mass ratios. The peak mass-ratio sensitivity occurs near $q \approx 0.05$, as secondaries become too faint to produce observable variability when their masses become much lower. The bias towards $q \approx 0.1$ is enhanced by the heightened accretion rates for secondary MBHBs near that mass ratio seen by Farris et al. (2014). Hydrodynamically variable binaries are most apparent near $q \approx 0.1$, both because of the enhanced accretion rate and because systems with $q \sim 1$ have variability periods shifted largely to longer periods. In binaries with $q \ll 1$, the secondary again becomes quite faint, and the variability induced in the accretion flow also begins to be negligibly small.

Selection biases in binary parameters identifiable as variables is important in estimating the GWB implied by a given binary population. Sesana et al. (2018) show that the CRTS candidates are in tension with current GWB upper limits from PTAs. The authors show that the tension can be decreased if the mass-ratio distribution of MBHBs is biased towards lower values, as systems with low q produce weaker GW signals. Our full binary populations have relatively flat mass ratio distributions in log space, but the selection effects for detecting variability prefer binaries with lower mass ratios. This observational bias has the opposite effect of an intrinsically low- q population. The high- q systems, which dominate the GWB, are absent from the observational candidates, but contribute more significantly to the GWB. The amplitude of this effect can be estimated as follows. For a given population

of candidates, their most likely binary parameters can be used to estimate a GW strain that they produce directly, h_c^{direct} . Our results suggest that over the full range of masses, less than 10 per cent of binaries, are observable. Furthermore, these systems should have $q \lesssim 0.1$, while binaries with $q \approx 1$ are intrinsically at least three times more common. Those nearer equal mass binaries produce GWs with strains roughly 10 times larger.¹³ All together, because GW strains add in quadrature, this implies that the true GWB amplitude is $h_c \approx 300^{1/2} h_c^{\text{direct}}$.

In conclusion, our models are able to explain only a fraction of previously identified, periodically variable binary candidates. The distributions of detectable-binary parameters that we find suggest that existing PTA constraints on the GWB also require a large fraction of candidates to be false positives. On the other hand, our results suggest that many of the candidates in CRTS and PTF could indeed be true MBHBs. We have presented the parameters of variables that we expect to be MBHBs in the hope that the candidates can be followed up with additional photometric and spectroscopic monitoring to find the binaries they contain. Confirmed examples of year-period MBHBs would present a boon to AGN and MBHB astrophysics. Such systems contain key information on MBH growth, MBHB evolution, and offer stringent constraints and insights into our predictions for low-frequency GW signals soon to be detectable by PTAs, and eventually by LISA.

ACKNOWLEDGEMENTS

The authors are grateful for insightful conversations and comments from Laura Blecha and Daniel D’Orazio, and helpful feedback from Edo Berger, Daniel Eisenstein, and Avi Loeb.

This research made use of *astropy*, a community-developed core PYTHON package for Astronomy (Astropy Collaboration et al. 2013), in addition to *scipy* (Jones et al. 2001), *ipython* (Perez & Granger 2007), and *numpy* (van der Walt, Colbert & Varoquaux 2011). All figures were generated using *matplotlib* (Hunter 2007).

ZH was supported in part by NASA through grants 16-SWIFT16-0015 and NNX17AL82G and by NSF through grant 1715661. AS is supported by the Royal Society.

REFERENCES

- Abramowicz M. A., 2005, *Growing Black Holes: Accretion in a Cosmological Context*. Springer, Berlin, p. 257
- Antonini F., Merritt D., 2012, *ApJ*, 745, 83
- Artymowicz P., Lubow S. H., 1994, *ApJ*, 421, 651
- Artymowicz P., Lubow S. H., 1996, *ApJ*, 467, L77
- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Barnes J. E., Hernquist L., 1992, *ARA&A*, 30, 705
- Barrows R. S., Comerford J. M., Greene J. E., Pooley D., 2017, *ApJ*, 838, 129
- Begelman M. C., Blandford R. D., Rees M. J., 1980, *Nature*, 287, 307
- Binney J., Tremaine S., 1987, *Galactic Dynamics*. Princeton Univ. Press, Princeton, NJ
- Boroson T. A., Lauer T. R., 2009, *Nature*, 458, 53
- Bowen D. B., Mewes V., Campanelli M., Noble S. C., Krolik J. H., Zilhão M., 2018, *ApJ*, 853, L17
- Butler N. R., Bloom J. S., 2011, *AJ*, 141, 93
- Chandrasekhar S., 1942, *Principles of Stellar Dynamics*. Univ. of Chicago Press, Chicago, IL.

¹³Strain is related to chirp-mass as, $h \propto \mathcal{M}^{5/3}$, and the chirp-mass is $\mathcal{M} = M q^{3/5} / (1 + q)^{6/5}$.

- Charisi M., Bartos I., Haiman Z., Price-Whelan A. M., Graham M. J., Bellm E. C., Laher R. R., Márka S., 2016, *MNRAS*, 463, 2145
- Charisi M., Haiman Z., Schiminovich D., D'Orazio D. J., 2018, *MNRAS*, 476, 4617
- Comerford J. M. et al., 2009, *ApJ*, 698, 956
- Comerford J. M., Gerke B. F., Stern D., Cooper M. C., Weiner B. J., Newman J. A., Madsen K., Barrows R. S., 2012, *ApJ*, 753, 42
- Comerford J. M., Pooley D., Barrows R. S., Greene J. E., Zakamska N. L., Madejski G. M., Cooper M. C., 2015, *ApJ*, 806, 219
- Cuadra J., Armitage P. J., Alexander R. D., Begelman M. C., 2009, *MNRAS*, 393, 1423
- D'Orazio D. J., Di Stefano R., 2018, *MNRAS*, 474, 2975
- D'Orazio D. J., Loeb A., 2018, *ApJ*, 863, 185
- D'Orazio D. J., Haiman Z., MacFadyen A., 2013, *MNRAS*, 436, 2997
- D'Orazio D. J., Haiman Z., Duffell P., Farris B. D., MacFadyen A. I., 2015a, *MNRAS*, 452, 2540
- D'Orazio D. J., Haiman Z., Schiminovich D., 2015b, *Nature*, 525, 351
- D'Orazio D. J., Haiman Z., Duffell P., MacFadyen A., Farris B., 2016, *MNRAS*, 459, 2379
- Desvignes G. et al., 2016, *MNRAS*, 458, 3341
- Dotti M., Merloni A., Montuori C., 2015, *MNRAS*, 448, 3603
- Dunhill A. C., Cuadra J., Dougados C., 2015, *MNRAS*, 448, 3545
- Eracleous M., Halpern J. P., M. Gilbert A., Newman J. A., Filippenko A. V., 1997, *ApJ*, 490, 216
- Eracleous M., Boroson T. A., Halpern J. P., Liu J., 2012, *ApJS*, 201, 23
- Esin A. A., McClintock J. E., Narayan R., 1997, *ApJ*, 489, 865
- Farris B. D., Duffell P., MacFadyen A. I., Haiman Z., 2014, *ApJ*, 783, 134
- Foster R. S., Backer D. C., 1990, *ApJ*, 361, 300
- Genel S. et al., 2014, *MNRAS*, 445, 175
- Goulding A. D. et al., 2018, *PASJ*, 70, S37
- Graham M. J. et al., 2015a, *MNRAS*, 453, 1562
- Graham M. J. et al., 2015b, *Nature*, 518, 74
- Haiman Z., 2017, *Phys. Rev. D*, 96, 023004
- Haiman Z., Kocsis B., Menou K., 2009, *ApJ*, 700, 1952
- Hayasaki K., Mineshige S., Sudou H., 2007, *PASJ*, 59, 427
- Hellings R. W., Downs G. S., 1983, *ApJ*, 265, L39
- Hogg D. W., 1999, preprint ([arXiv:astro-ph/9905116](https://arxiv.org/abs/astro-ph/9905116))
- Hopkins P. F., Richards G. T., Hernquist L., 2007, *ApJ*, 654, 731
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Ivezic Z. et al., 2008, preprint ([arXiv:0805.2366](https://arxiv.org/abs/0805.2366))
- Jaroszynski M., Abramowicz M. A., Paczynski B., 1980, *AcA*, 30, 1
- Jiang Y.-F., Stone J. M., Davis S. W., 2014, *ApJ*, 796, 106
- Jones E. et al., 2001, SciPy: Open Source Scientific Tools for Python
- Kauffmann G., Haehnelt M., 2000, *MNRAS*, 311, 576
- Kelley L. Z., Blecha L., Hernquist L., 2017a, *MNRAS*, 464, 3131
- Kelley L. Z., Blecha L., Hernquist L., Sesana A., Taylor S. R., 2017b, *MNRAS*, 471, 4508
- Kelley L. Z., Blecha L., Hernquist L., Sesana A., Taylor S. R., 2018, *MNRAS*, 477, 964
- Komossa S., 2006, *Mem. Soc. Astron. Ital.*, 77, 733
- Koss M., Mushotzky R., Treister E., Veilleux S., Vasudevan R., Tripp M., 2012, *ApJ*, 746, L22
- Liu T. et al., 2016, *ApJ*, 833, 6
- Liu T., Gezari S., Miller M. C., 2018, *ApJ*, 859, L12
- Lubow S. H., Seibert M., Artymowicz P., 1999, *ApJ*, 526, 1001
- MacLeod C. L. et al., 2010, *ApJ*, 721, 1014
- MacLeod C. L. et al., 2011, *ApJ*, 728, 26
- Magorrian J., Tremaine S., 1999, *MNRAS*, 309, 447
- Mahadevan R., 1997, *ApJ*, 477, 585
- McConnell N. J., Ma C.-P., 2013, *ApJ*, 764, 184
- McWilliams S. T., Ostriker J. P., Pretorius F., 2014, *ApJ*, 789, 156
- Merritt D., Mikkola S., Szell A., 2007, *ApJ*, 671, 53
- Middleton H., Chen S., Del Pozzo W., Sesana A., Vecchio A., 2018, *Nature Commun.*, 9, 573
- Muñoz D. J., Lai D., 2016, *ApJ*, 827, 43
- Narayan R., Yi I., 1995, *ApJ*, 452, 710
- Nelson D. et al., 2015, *Astron. Comput.*, 13, 12
- Noble S. C., Mundim B. C., Nakano H., Krolik J. H., Campanelli M., Zlochower Y., Yunes N., 2012, *ApJ*, 755, 51
- Peters P. C., 1964, *Phys. Rev.*, 136, 1224
- Peters P. C., Mathews J., 1963, *Phys. Rev.*, 131, 435
- Prez F., Granger B., 2007, *Comput. Sci. Eng.*, 9, 21
- Reardon D. J. et al., 2016, *MNRAS*, 455, 1751
- Richards G. T. et al., 2009, *ApJS*, 180, 67
- Rodriguez C., Taylor G. B., Zavala R. T., Peck A. B., Pollack L. K., Romani R. W., 2006, *ApJ*, 646, 49
- Rodriguez-Gomez V. et al., 2015, *MNRAS*, 449, 49
- Roedig C., Sesana A., Dotti M., Cuadra J., Amaro-Seoane P., Haardt F., 2012, *A&A*, 545, A127
- Rosado P. A., Sesana A., Gair J., 2015, *MNRAS*, 451, 2417
- Sarajedini V. L., Gilliland R. L., Kasm C., 2003, *ApJ*, 599, 173
- Schmidt K. B., Marshall P. J., Rix H.-W., Jester S., Hennawi J. F., Dobler G., 2010, *Astrophys. J.*, 714, 1194
- Sesana A., Khan F. M., 2015, *MNRAS*, 454, L66
- Sesana A., Haardt F., Madau P., 2007, *ApJ*, 660, 546
- Sesana A., Haiman Z., Kocsis B., Kelley L. Z., 2018, *ApJ*, 856, 42
- Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337
- Shi J.-M., Krolik J. H., 2015, *ApJ*, 807, 131
- Shi J.-M., Krolik J. H., Lubow S. H., Hawley J. F., 2012, *ApJ*, 749, 118
- Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, *MNRAS*, 452, 575
- Simon J., Burke-Spolaor S., 2016, *ApJ*, 826, 11
- Steinborn L. K., Dolag K., Comerford J. M., Hirschmann M., Remus R.-S., Teklu A. F., 2016, *Mon. Not. R. Astron. Soc.*, 458, 1013
- Sadowski A., 2009, *ApJS*, 183, 171
- Tang Y., MacFadyen A., Haiman Z., 2017, *MNRAS*, 469, 4258
- Tang Y., Haiman Z., MacFadyen A., 2018, *MNRAS*, 476, 2249
- Taylor S. R. et al., 2015, *Phys. Rev. Lett.*, 115, 041101
- Taylor S. R., Simon J., Sampson L., 2017, *Phys. Rev. Lett.*, 118, 181102
- The NANOGrav Collaboration et al., 2015, *ApJ*, 813, 65
- Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, *MNRAS*, 438, 1985
- Tsalmantza P., Decarli R., Dotti M., Hogg D. W., 2011, *ApJ*, 738, 20
- van der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22
- Vasiliev E., Antonini F., Merritt D., 2015, *ApJ*, 810, 49
- Verbiest J. P. W. et al., 2016, *MNRAS*, 458, 1267
- Vogelsberger M. et al., 2014a, *MNRAS*, 444, 1518
- Vogelsberger M. et al., 2014b, *Nature*, 509, 177
- Volonteri M., Miller J. M., Dotti M., 2009, *ApJ*, 703, L86
- Yu Q., 2002, *MNRAS*, 331, 935

APPENDIX

To facilitate comparisons with surveys beyond CRTS and LSST, we present our predicted observations of MBHBs as periodically variable AGN in sensitivity-variability time-scale parameter space. Fig. A1 gives detection rates in units of AGN⁻¹, and Fig. A2 as a cumulative number of all-sky observable systems. The associated tabular data for Figs A1 and A2 are included online. To estimate the number of detectable systems for a particular survey, the limiting variability period that it is sensitive to must be determined. Nyquist sampling implies the maximum period is at most the total temporal baseline of the survey, but likely numerous complete cycles of variability should be observed to reject red-noise contamination. Multiplying the predicted rates in Fig. A1 by the actual number of AGN observed by a particular survey, as opposed to using Fig. A2 directly, will provide a more robust prediction as it decreases systematic uncertainties from the simulated AGN luminosity function.

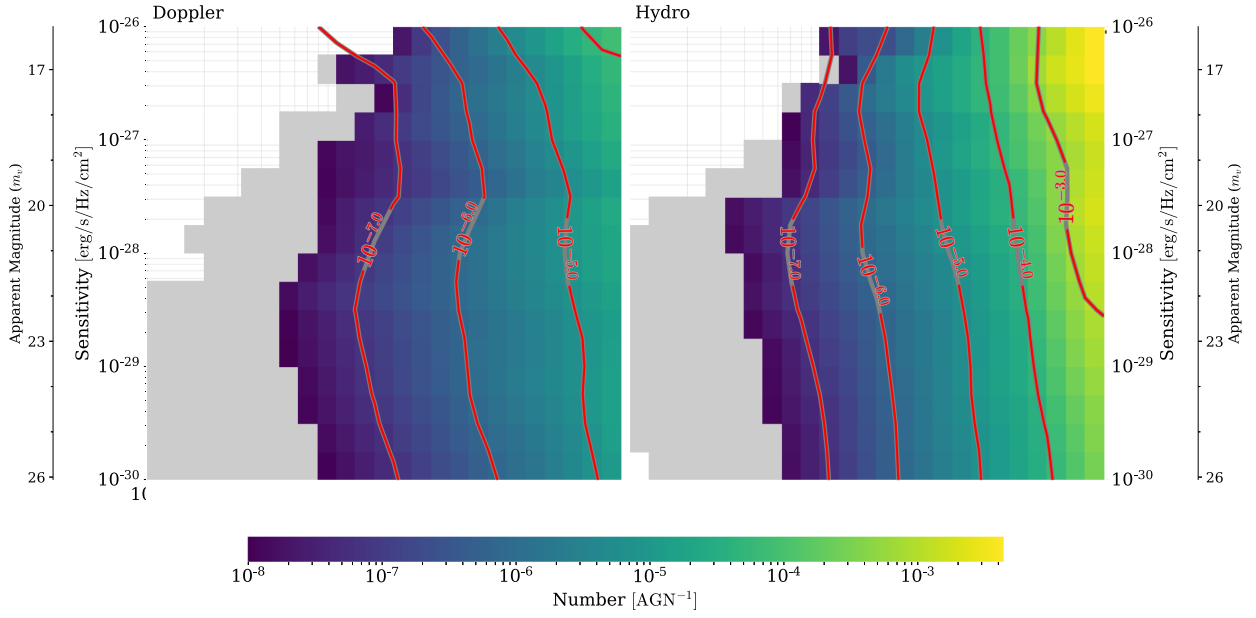


Figure A1. Predicted rate of observable binaries in sensitivity-period parameter space. Occurrence rates are given in units of AGN^{-1} , which leads to non-monotonic behaviour as a function of sensitivity. Cumulative binary occurrence rates are shown for variations between 0.1 yr and the indicated periods. The observational duration of a given survey determines the maximum variability period that it is sensitive to.

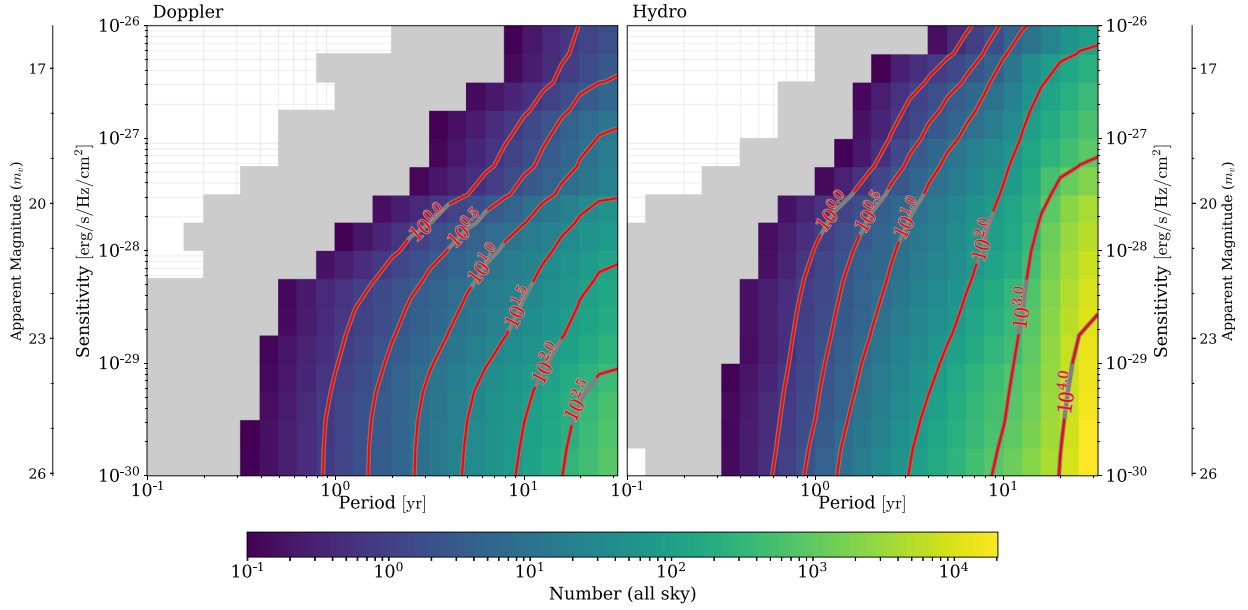


Figure A2. Predicted number of all-sky observable binaries in sensitivity-period parameter space.

This paper has been typeset from a $\text{\LaTeX}$ file prepared by the author.