

Predicting Electromagnetic Signatures of Merging Compact Objects

Daniel John D'Orazio

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

© 2016

Daniel John D'Orazio

This work is licensed under a Creative Commons Attribution 4.0 International License.

ABSTRACT

Predicting Electromagnetic Signatures of Merging Compact Objects

Daniel John D’Orazio

We develop two new methods to measure the structure of matter around the Milky Way using stellar tidal streams from disrupting dwarf galaxies and globular clusters. The dark matter halo of the Milky Way is expected to be triaxial and filled with substructure, but measurements of the shape and profile of dark matter around the Galaxy are highly uncertain and often contradictory. We demonstrate that kinematic data from near-future surveys for stellar streams or shells produced by tidal disruption of stellar systems around the Milky Way will provide precise measures of the gravitational potential to test these predictions. We develop a probabilistic method for inferring the Galactic potential with tidal streams based on the idea that the stream stars were once close in phase space and test this method on synthetic datasets generated from N-body simulations of satellite disruption with observational uncertainties chosen to mimic current and near-future surveys of various stars. We find that with just four well-measured stream stars, we can infer properties of a triaxial potential with precisions of order 5–7 percent.

We then demonstrate that, if the Milky Way’s dark matter halo is triaxial and is not fully integrable (as is expected), an appreciable fraction of orbits will be chaotic. We examine the influence of chaos on the phase-space morphology of cold tidal streams and show that streams even in weakly chaotic regions look very different from those in regular regions. We discuss the implications of this fact given that we see several long, thin streams in the Galactic halo; our results suggest that long, cold streams around our Galaxy must exist only on regular (or very nearly regular) orbits and potentially provide a map of the regular regions of the Milky Way potential. We then apply this

understanding of stream formation along chaotic orbits to the interpretation of a newly-discovered, puzzling stellar stream near the Galactic bulge. We conclude that the morphology of this stream is consistent with forming along chaotic orbits due to the presence of the time-dependent Galactic bar.

These results are encouraging for the eventual goal of using flexible, time-dependent potential models combined with larger data sets to unravel the detailed shape of the dark matter distribution around the Milky Way.

Contents

List of Figures	ii
List of Tables	iii
Acknowledgements	iii
1 Introduction	1
1.1 Part I: Massive Black Hole Binaries	1
1.2 Part II: Stellar Black Hole + Neutron Star Binaries.	2
1.3 Outline of thesis	2
I Massive Black Hole Binaries	3
2 Accretion into the Central Cavity	5
2.1 Introduction	5
2.2 Details of Numerical Simulations	8
2.2.1 Numerical Implementation and Assumptions	9
2.2.2 Numerical Parameter Choices	11
2.2.3 Initial Conditions	13
2.2.4 Disc Parameters and Accretion Rate	14

2.2.5	Tests of Code Implementation	16
2.3	Results	16
2.3.1	Equal-Mass Binary	16
2.3.2	Unequal-Mass Binaries	30
2.4	Summary and Discussion	37
2.4.1	Comparison with MM08	38
2.4.2	Viscosity Study	39
2.4.3	Resolution Study	44
2.4.4	Physical Regime: Black Hole Binary Parameters	49
2.4.5	Caveats	53
2.5	Conclusions	55
2.6	Viscous implementation in polar coordinates	58
2.6.1	Viscosity and the Momentum Equation	58
2.6.2	The Form of the Viscous Stress Tensor	59
2.6.3	Components of σ in Polar Coordinates	60
2.6.4	Divergence of σ in Polar Coordinates	61
2.6.5	Implementation in FLASH	63
2.6.6	Viscously Spreading Ring	64
2.6.7	Cartesian Shear Flow	64
3	A transition in circumbinary discs	69
3.1	Introduction	69
3.2	Restricted 3-body Analysis	72
3.2.1	Restrictions on Orbits from Conserved Integrals of Motion	72
3.2.2	Restrictions on Orbits from Equations of Motion	78

3.2.3	Hydrodynamical Effects	79
3.3	Hydrodynamical simulations	87
3.3.1	Fiducial Simulations	87
3.3.2	Hydrodynamic Parameter Study	93
3.4	Discussion and Summary	99
3.5	Viscous, restricted three-body problem	103
4	A Reduced Orbital Period for PG 1302-102	111
4.1	Introduction	111
4.2	Implications of a shorter orbital period	113
4.2.1	Binary-Disc Decoupling	115
4.2.2	Binary Fraction among Quasars	117
4.2.3	Detectability of GWs	119
4.3	Testing the Binary BH Scenario for PG 1302	119
4.3.1	Broad Line Variability and Asymmetry	119
4.3.2	Relativistic Effects	123
4.3.3	Orbital time-scale Variability	124
4.4	Conclusions	125
5	Relativistic boost	127
5.0.1	V-band emission from a three–component system in PG 1302-102.	131
5.0.2	Model fitting to the PG 1302 optical light curve.	136
5.0.3	Disk Precession.	139
5.0.4	Archival UV data.	140

II Neutron Star - Black Hole Binaries	144
6 Rindler Dipoles	146
6.1 Introduction	146
6.2 Set-up and Limits	149
6.2.1 Rindler Spacetime	149
6.2.2 Electrodynamical Properties of an Event Horizon and the Horizon Battery .	152
6.3 A magnetic Dipole in arbitrary motion	155
6.3.1 The Electromagnetic Four-Potential	155
6.4 A Freely Falling Dipole Solution	163
6.5 A Boosted, Freely Falling Dipole Solution	168
6.6 Rindler Dipole	177
6.7 Rindler Dipole Boosted Parallel to the Horizon	182
6.8 Consequences for the BH-NS Binary	190
6.8.1 Voltage, Luminosity, and Energy	190
6.8.2 NS plummet into a SMBH	198
6.9 Conclusions	202
6.10 Detailed Solution to the Field Equations	205
6.10.1 Point Charge	206
6.10.2 A General Dipole Solution	208
6.11 Dipole Moments	212
7 Bright Transients from Black Hole - Neutron Star Mergers	217
7.1 introduction	217
7.1.1 The power of the battery	219
7.2 Curvature radiation	224

7.3	Fireball	227
7.3.1	pair production	228
7.3.2	Expansion and Emission	232
7.4	Post Merger	236
7.5	Observability	239
7.6	Conclusion	242
7.7	Parameter Dependence of Curvature Spectra	244

List of Figures

- 2.1 The distortion of a disc of massless test particles, initially in circular orbits around the center-of-mass of an equal-mass binary, with a central cavity. The panels show snapshots of the locations of the disc particles after 0.25, 0.5, and 0.75 binary orbits, as labeled. The orbits of the particles were followed by solving the restricted three-body problem, and are shown in a frame co-rotating with the binary. The binary point masses are marked by the two [red] dots at $x = \pm 0.5a$. The figure illustrates the tendency of the binary to create streams of particles entering the central cavity, due to gravity alone.

2.2 Snapshots of the surface density for the $q = 1$ disc in the point-symmetric transient state after ~ 1350 binary orbits. The snapshots are sharpened (masked to include only high Fourier frequencies in the image) in order to see the streams more clearly. The connected outer circles are drawn to guide the eye. These circles rotate with the disc structure at a period of $\approx 3t_{\text{bin}}$. Streams are generated on either side of the cavity, (shown in the left panel; the streams shown in this panel are moving inward, away from the cavity wall). These streams ultimately crash into the cavity wall on the diagonally opposite sides of the cavity, near the site where the opposing stream was generated (shown in the right panel; the streams shown in this panel are moving outward, toward the cavity wall). If a small asymmetry causes one stream to become stronger, then a runaway process would ensue pushing one side of the disc further from the binary and allowing the other side to come closer; this process could ultimately be responsible for the observed lopsided shape of the cavity. 20

- 2.3 *Top row:* Surface density distributions for the equal-mass ratio ($q = 1.0$) binary during a transient, point-symmetric state after ~ 1000 binary orbits (left) and during the quasi-steady asymmetric state after ~ 4000 binary orbits (right). The inset in the top left panel zooms in to the inner $\pm 2.5r/a$ of the disc in order to show the stream morphology. *Bottom two rows:* snapshots at ~ 4000 binary orbits, during the quasi-steady-state phase, for mass ratios $q = 0.5, 0.1, 0.075$, and 0.01 , as labeled. Each panel shows the inner $\sim 6\%$ of the simulated disc, extending $\pm 6r/a$ in both directions. The solid circles mark the inner boundary of the simulation at $r = r_{\min} = a$. The larger dotted circle at $r \simeq 2.08a$ is the position of the $(m, l) = (2, 1)$ outer Lindblad resonance (shown only for reference). Surface densities are plotted with the same linear grayscale in each panel, with the darkest regions corresponding to a maximum density of $0.8\Sigma_0$ ($0.4\Sigma_0$ for the top left panel). Orbital motion is in the clock-wise direction. 22
- 2.4 Snapshots of the azimuthally averaged disc surface density at different times (shown by different curves in each panel, from 0 to 4000 orbits, as labeled), and for different mass ratios (shown in different panels, from $q = 1.0$ to 0.01 , as labeled). In each panel, the solid [black] curve shows, for reference, the density profile in the point-mass ($q = 0$) case after 4000 orbits. The vertical dotted lines mark the radius where binary and viscous torques balance (from Figure 2.5); these lie close to the observed cavity edges. The vertical solid lines mark the inner edge of the integration domain ($r = r_{\min} = a$). In each case, the inner circumbinary disc spreads inward with time, but the density profile remains sharply truncated, with a low-density central cavity inside $r \lesssim 2a$ 24

2.6 The time variable accretion rate across the inner boundary of the simulation measured at $r = r_{\min} = a$ (top of each pair of panels) and the corresponding Lomb-Scargle periodogram (bottom of each pair) computed over 100 orbits. Panels are displayed in order of decreasing binary mass ratio, starting from $q = 1.0$ at top left (on the previous page) to $q = 0.003$ at the bottom right (on this page). The average accretion rate in each panel is denoted by the solid [blue] horizontal line. The dashed [green] horizontal line in each plot shows the average accretion rate for the point mass ($q = 0$) case for reference. For $q > 0.05$, the accretion rate is strongly modulated by the binary, with either one, two, or three distinct periods present simultaneously, depending on the value of q (see text for detailed explanations). For $q \lesssim 0.05$, the binary still reduces the mean accretion rate noticeably, but does not imprint strong periodicity; the $q = 0.003$ binary is nearly indistinguishable from a single BH.	29
2.7 The average mass accretion rate as a function of q , and for simulations with different spatial resolution. This is the same information as in Table 2.2, except here shown graphically and with only the simulations at the fiducial magnitude of viscosity.	30
2.8 The position of the cavity wall as a function of q , in runs with different radial and azimuthal resolutions, as labeled. The points mark the radii r_{ce} at which the azimuthally averaged viscous and gravitational torques balance (equation 2.10). The black line is an empirical fit to the fiducial resolution data points (equation 2.14).	33

2.9 Accretion rates at the inner boundary $r_{\min} = a$ for the equal-mass binary as in the top left panels of Figure 2.6, except for the lowest resolution runs used in this study which matches the highest resolution used in the disc simulated by MM08. The solid horizontal [blue] line in the top panel is the average accretion rate, and the bottom panel shows the Lomb-Scargle periodogram computed over 100 binary orbits.	40
2.10 Two-dimensional surface density distributions during the quasi steady-state phase, as in Figure 2.3, except for the single mass ratio $q = 1.0$, and for four different values of the viscosity parameter α , as labeled. Increasing α causes ripples created by streams impacting the cavity wall to smear out more quickly causing the surface density snapshots to appear smoother. For all values of α shown here, the over dense lump still survives long enough to create the $\sim (5 - 6)t_{\text{bin}}$ modulation of the accretion rate. Also for larger α , the near side of the disc extends in closer to the binary causing a larger fraction of streams to exit the integration domain at $r = a$. This results in higher measured accretion rates relative to the point mass values for the same α 's.	41
2.11 The time-dependent accretion rates as in Figure 2.6, except for the single mass ratio $q = 1.0$ and four different values of the viscosity parameter α , as shown in Figure 2.10. Additionally the top panel of each figure shows the entire accretion rate history for the $q = 1$ (black) and $q = 0$ (red) cases. Notice that long-term (once per $\sim 400t_{\text{bin}}$) variability appears in the larger α runs and is coincident with the period at which the elongated cavity precesses..	42

- 2.12 Two-dimensional surface density distributions during the quasi steady-state phase, as in Figure 2.3, except for the single mass ratio $q = 0.1$, and for four different combinations of high/low radial and azimuthal resolutions, as labeled. Increasing the spatial resolution decreases numerical diffusion, leads to sharper features, and allows stronger accretion streams. The stronger streams lead to more over-dense lumps where the regurgitated streams hit the cavity wall. As a result, the cavity becomes larger and more lopsided as the resolution is increased. In the lowest resolution case, the cavity never becomes lopsided. 45
- 2.13 The time-dependent accretion rates as in Figure 2.6, except for the single mass ratio $q = 0.1$ and the four different resolutions also shown in Figure 2.12. The $q = 0.1$ binary is at the cusp of the transition from the three-timescale to the single-timescale regime, and is particularly sensitive to resolution, as seen especially in the maxima of the accretion spikes. 47

3.2 Zero-velocity curves for four different values of the Jacobi constant for a binary with mass ratio $q = M_s/M_p = 0.1$ (level curves of the potential plotted in Figure 3.1). For $C_J \geq C^{\text{crit}}$ (blue thick dashed and solid black), the zero-velocity curves connect, separating the binary plane into distinct inner and outer regions (as well as a third region around the secondary for large enough values of the Jacobi Constant). For $C_J < C^{\text{crit}}$ (red lines), the zero-velocity curves open at L2 and then at L3 for even smaller C_J . The critical zero-velocity curve (black) passes through the Lagrange point L2. The five Lagrange points are labeled for reference.	74
3.3 The dark green regions are bounded by the zero-velocity curve which passes through L2, delineating the smallest restricted regions which connect and separating the binary plane into distinct inner and outer regions. Particles trapped outside (inside) of the dark green region are labeled blue (red). Depending on their velocity vectors, light- and dark-green particles are free to move from inner to outer regions.	77
3.4 Each panel is the result of evolving an initially spatially random distribution of particles, within radius $r < 2.5a$, via the R3Bp equations, for one binary orbital period. The colouring of particles refers to the initial placement of a particle as in Figure 3.3. The black diamonds mark the Lagrange points (see Figures 3.1 and 3.2). These snapshots, after only one binary orbit, show the formation of streams acting to deplete green particles.	80
3.5 Same as Figure 3.4 except after 100 binary orbital periods.	81
3.6 The same as Figure 3.5 except zooming in on the mass-ratio range $0.02 \leq q \leq 0.08$	82

3.7 Delineation of different phases in a non-viscous circumbinary disc. The y-axis records the inverse disc aspect ratio, equivalent to the orbital Mach number; a smaller value signifies larger pressure forces; large pressure forces preclude reasoning based on a purely gravitational analysis. Points represent the parameters of hydrodynamical simulations run in this study (§3.3). Red denotes a simulation with a filled gap/cavity.	85
3.8 Snapshots of the surface density distribution (shown in units of the unperturbed value, with a logarithmic colour scheme) for an equal mass binary with disc aspect ratios ($r_s/H \equiv \mathcal{M}$) surrounding the pressure dominated condition Eq. (3.5). Here we set the viscosity to be very small (the coefficient of kinematic viscosity is $\nu = 10^{-6}a_0^2\Omega_{\text{bin}}$, where Ω_{bin} is the binary angular frequency) in order to examine the analytic R3Bp prediction (3.5). The top row is for an isothermal equation of state $P = (c_s^{\text{iso}})^2\Sigma$ and the bottom row is for an adiabatic equation of state $P = (c_s^{\text{ad}})^2\Sigma^{5/3}$	86
3.9 Snapshots of the surface density distribution (shown in units of the unperturbed value, with a logarithmic colour scheme) from hydrodynamical simulations for a disc with orbital Mach number $\mathcal{M} \equiv r/H \equiv v_{\text{eff}}/c_s = 20$, and constant coefficient of kinematic viscosity $\nu = 0.01a_0^2\Omega_{\text{bin}}/\mathcal{M}^2$. The binary mass ratio increases from left to right, top to bottom, as labeled. For small mass ratios, the system is in nearly steady-state and an annular gap is cleared in the orbit of the secondary black hole. For $q \gtrsim 0.03$, the gap morphs into an even lower density time-dependent, precessing central cavity. The critical zero-velocity curve, which passes through L2, is over-drawn in white. The relatively shallow annular gap in the $q = 0.001$ case is difficult to see on this scale because the accretion prescription and inner boundary cause the inner disc to drain onto the primary.	91

3.10 Evolution of two passive scalars for $q = 0.001$, $q = 0.01$, and $q = 0.05$ (rows) at different times (columns) during gap opening as well as velocity vectors showing fluid motion in the co-rotating frame. The red scalar starts inside of the critical ZVC (overlaid, black curve) and the blue scalar starts outside. The green dots denote the Lagrange points of the binary potential (See Figures 3.1 and 3.2). Published with this article are <i>movies</i> of the above three simulations with the red and blue passive scalars plotted on different panels and with the same initialization as Figure 3.3 (these movies and the corresponding movies of surface density evolution can also be found at http://user.astro.columbia.edu/~dorazio/moviespage). . .	92
3.11 The standard deviation of \dot{M} onto the secondary (left), primary (center), and both (right) BHs, computed over the final 100 orbits of each simulation. The vertical solid line is drawn at the $q = 0.04$ transition, the dashed horizontal line is the mean of the standard deviation of each data set.	96
3.12 The lopsidedness, Eq. (3.9), of the circumbinary disc, spatially averaged over the region outside of the binary orbit, $r > a$, and time averaged over the final 100 orbits of each simulation. The vertical sold line denotes $q = 0.04$	98
3.13 Snapshots of the surface density distribution (shown in units of the unperturbed value, with a logarithmic colour scheme) from hydrodynamical simulations for a disc with $\mathcal{M} = 30$, and constant coefficient of kinematic viscosity $\nu = 10^{-3}a_0^2\Omega_{\text{bin}}$. For these high-viscosity, low-pressure simulations, we find an asymmetric-disc shape even for small mass ratio binaries (left). The transition to a time-dependent, lopsided cavity at $q \sim 0.04$ still takes over for larger mass ratio binaries (right). . .	99

3.14 Eigenvalues of a linear-stability analysis of the L4 and L5 points with an added viscous force. The top panel shows the real (solid lines) and imaginary (dashed lines) parts for different values of the constant coefficient of kinematic viscosity, $\nu = 0.0$ (black), $\nu = 0.005$ (blue), $\nu = 0.01$ (red). The bottom panel displays instability timescales normalized to the binary orbital time.	108
3.15 The difference in orbital out-spiral of a test-particle perturbed from the L4 and L5 points. λ_{L4} and λ_{L5} refer to the maximum real part of the complex eigenvalues found from Eq. 3.14. The inset zooms in on the region below the $q = 0.04$ transition where weak asymmetry between the L4 and L5 points exists. A value of -0.01 in this plot means that, after 100 binary orbital times, the difference in the final position of a particle perturbed from L4 is an e-fold farther from its starting position than a particle identically perturbed from L5. It is likely that a higher order perturbation in the mass ratio is required to capture the return of $\lambda_{L5} - \lambda_{L4} = 0$ for $q = 1$	109
3.16 Same as Figure 3.5 except for 10^4 particles obeying the viscous R3Bp Eqs. (3.10) and (3.10) with constant $\nu = 0.001a_0^2\Omega_0$	110
4.1 Results of 2D hydrodynamical simulations of a binary BH surrounded by a circumbinary accretion disc. The BHs clear out a central cavity and form their own minidisks. <i>Left-hand panels:</i> snapshots of the (logarithmic) surface-density of the gas discs, after reaching quasi-steady state, with mass ratios of $q = 0.5$ (top) and $q = 0.1$ (bottom). <i>Right-hand panels:</i> corresponding LSPs of the total accretion rate on to the secondary + primary BHs. The discs are locally isothermal with a Mach number of 10 and an alpha viscosity prescription ($\alpha = 0.1$).	113

4.2	The ratio of binary separation a to the decoupling radius r_{GW} , for three different values of the ratio between the rest frame optical period and the true binary period, $\chi = t_{\text{opt}}/t_{\text{bin}}$. The shaded region marks the binary mass range inferred from the widths of broad lines measured by G15. For $\chi > 3$, the PG 1302 binary is past decoupling, for any choice of mass M .	116
4.3	Combinations of total binary mass M and $\chi = t_{\text{opt}}/t_{\text{bin}}$ for which the predicted binary fraction of quasars matches 10 CRTS candidates with luminosity above that of PG 1302 (dark grey, bounded by solid curves), and for which it consists only of PG 1302 (light grey, bounded by red dashed curves). Vertical lines delineate the range of masses preferred by broad line widths. Each shaded region is bounded by the fraction f_{bin} of quasars which are triggered by a binary. In each case, the lines correspond to $f_{\text{bin}} = 0.01, 0.1, 1$ (left to right).	118
4.4	Predicted variations of the centroid V_0 (left) and FWHM Γ (right) of an emission line emanating from the inner CBD. The total accretion rate on to both black holes is over plotted in the right-hand panel in arbitrary units (orange). Dark black lines are smoothed versions of the light grey simulation data.	123
5.1	Binary parameters producing the optical flux variations of PG 1302-102 by relativistic boost. Shaded regions mark combinations of binary mass M , mass ratio $q = M_2/M_1$, and inclination i causing $>13.5\%$ flux variability (or line-of-sight velocity amplitude $(v/c) \sin i \geq 0.07$), computed from the Doppler factor $D^{3-\alpha}$ with the effective spectral slope of $\alpha = 1.1$ in the V band, including emission from the primary, as well as from the secondary black hole. The three panels assume fractions $f_2 = 1.0, 0.95$, or 0.8 of the total luminosity arising from the secondary black hole; these values are consistent with fractions found in hydrodynamical simulations? (see <i>Methods</i>).	131

5.4 Extended Data Figure 1 Model spectrum of PG1302. Circumbinary (dashed blue) and circumsecondary (solid black) disk spectra for a total binary mass of $10^{9.4}$, binary mass ratio $q = 0.05$, and ratio of accretion rates $\dot{M}_2/\dot{M}_1 = 20$. A vertical dashed line marks the center of the V-band and the approximate flux from an advection-dominated accretion flow (ADAF) is shown as a red dot for the V-band contribution of the primary. The spectrum for a radiatively efficient, thin disk around the primary is shown by the thin red dashed curve for reference.	142
5.5 Extended Data Figure 2 Parameter combinations for which the combined V-band luminosity of the three-component system varies by the required 0.14 mag. M is the binary mass, q is the mass ratio, and i is the orbital inclination angle. This figure is analogous to Fig. 1, except instead of adopting a fractional luminosity contribution f_2 by the secondary, the luminosities of each of the three components are computed from a model: the primary's luminosity is assumed to arise from an ADAF, while the secondary's luminosity is generated by a modestly super-Eddington thin disk. Emission from the circumbinary disk is also from a thin disk, and is negligible except for binaries with the lowest mass ratio $q \lesssim 0.01$ (see text).	143
5.6 Extended Data Figure 3 Model fits to PG1302's optical light curve. Best-fitting curves assuming relativistic boost from a circular binary (solid black curves), a pure sinusoid (red dotted curves) and accretion rate variability adopted from hydrodynamical simulations? (blue dashed curves) for a $q = 0.075$ (a) and a $q = 0.1$ (b) mass-ratio binary. The grey points with 1σ errors bars show the data for PG1302?.	143

6.1	Neutron star - Rindler horizon effective circuit diagram. Magnetic field lines act as wires connecting the neutron star to the horizon. Current flows in (out) of the horizon via positively (negatively) charged particles spiraling in tight Larmor radii around magnetic field lines into the horizon.	147
6.2	Rindler space is the shaded wedge given by $T > \pm Z$, $Z > 0$ on the Minkowski spacetime diagram. The dotted vertical line is the trajectory of a Minkowski observer and the dashed hyperbolic line is that of a Rindler observer as viewed by a Minkowski observer. In the frame of the Rindler observer an event horizon exists at $T = \pm Z$	151
6.3	Diagram demonstrating the relationship between source trajectory coordinates $X_S(\tau)$, observer coordinates X , and the retarded proper time $\tau_*(X)$ at the intersection of the past light cone of an observer at X and the source trajectory.	157
6.4	Spacetime diagrams for the infalling Rindler dipole of section §6.4. Also shown is the worldline of a Rindler observer. The top panel is drawn by Minkowski observers, the bottom panel is drawn by Rindler observers. Note that $Z_S = z_S$ at $T = t = 0$, hence the labeling of the initial source position.	165
6.5	The $x = x_S = 0$ slice (plane containing the dipole) of the Poynting flux for the infalling dipole as viewed by Rindler observers. The axes are in units of Z_S	167
6.6	3D visualization of the magnetic dipole field lines of a dipole falling from initial height $z_S(t = 0) = Z_S$ above the Rindler horizon, denoted by the gray plane at $z = 0$. The visualization region is a cube with side length $2Z_S$. On the left, magnetic field lines and the corresponding horizon current densities $\mathbf{J}_{\mathcal{H}}$ are plotted. On the right, electric field lines and corresponding charge densities $\sigma_{\mathcal{H}}$ (0 here) are plotted on the stretched horizon located at $z_H = 0.01Z_S$	169

- 6.7 Spacetime diagram depicting the x -component of the infalling, boosted dipole worldline of §6.5 from the Rindler observer's perspective, for three different values of β_S . The z -component of the worldline is identical to that portrayed for the infalling dipole in the bottom panel of Figure 6.4 except the light cone structure is altered. Because the infalling boosted dipole approaches the speed of light in the z -direction, the motion in the x -direction must go to zero ($dx/dt \rightarrow 0$). This is evident from the worldlines in this figure which asymptote to vertical lines. 171
- 6.8 3D visualization of the field lines of a dipole spiraling into the Rindler horizon from initial height $z_S(t = 0) = Z_S$ with an initial boost of $\beta_S = 0.9$ in the x -direction. The visualization region spans from $-2Z_S$ to $2Z_S$ in the x - and y -directions and extends $2Z_S$ above the Rindler horizon. Surface currents \mathbf{J}_H and surface charge densities σ_H are plotted on the stretched horizon located at $z_H = 0.01Z_S$ 176
- 6.9 An $x = x_S = 0$ slice of the Poynting flux for the infalling boosted dipole of §6.5 as viewed by Rindler observers for three different boost magnitudes in the x -direction, $\beta_S = 0.1, 0.5, 0.9$. The Poynting flux is 0 at infinity despite outward components of the field in the region plotted here. The axes are in units of Z_S 178

6.10 Current density vectors (white) overlaid on contours of charge density on the stretched horizon ($\alpha_{\mathcal{H}} = 10^{-4}$) of the infalling boosted dipole with rest frame magnetic moment in the y -direction. From top to bottom, the magnitude of the boost in the x -direction increases from $\beta_S = 0.1, 0.5, 0.9$. As inferred from the last of Eqs. (6.59), the magnitude of the charge density increases with β_S . Also the shape of the charge separation is squeezed in the direction of source boost as indicated by Eq. (6.60). All of the snapshots are taken at $g_H t = 1$ and the contour labels are arbitrarily scaled. The gray regions are regions of steeply increasing $\sigma_{\mathcal{H}}$ which have been removed to more clearly view the contour structure. The axes are in units of Z_S .	179
6.11 3D visualization of the magnetic dipole field lines hovering at constant height z_S above the Rindler horizon, denoted by the gray plane at $z = 0$. The visualization region is a cube with side length $2z_S$.	183
6.12 Spacetime diagram in the Rindler frame depicting the worldline of a source boosted parallel to the Rindler horizon (§6.7). The light ray (dotted line) has slope $dt/dz = (g_H z_S)^{-1}$ and the worldline has slope $(v_{S,x} g_H z_S)^{-1}$, where z_S is the constant position of the source above the Rindler horizon.	184

6.13 Streamlines of the magnetic (left) and electric (right) fields in the plane $x = 0$ for a magnetic dipole with dipole moment $\mathbf{m} \propto \hat{\mathbf{e}}_y$ and with three different boost velocities increasing from top to bottom $v_{S,x} = 0.1, 0.5, 0.9$ in the x -direction. Plotted over the streamlines are contours of the x -components of the fields relative to the $y-z$ magnitude. Darker regions represent negative values and lighter regions represent positive values. The white regions are clipped to better view the contour structure. The snapshots here are taken at $g_H t = 1/4$, however the fields retain the same structure for all time except for their motion in the x -direction (out of the page). The axes are in units of z_S . Since the source is boosted, observers near the horizon see the fields as they would have been when the dipole was further away in the negative x -direction. The result is an observed dragging of the fields along the horizon in the negative x -direction. The observed larger radius of curvature of the dipole lobes manifests itself as the flattening of the field lines. As can be gathered from Figure 6.12, this effect is intensified for larger boost factor $v_{S,x}$. In the $v_{S,x} = 0.9$ case, plotted at the bottom of the figure, the 2D slice of the magnetic field loses its dipolar structure in most of the region below the source. As $v_{S,x}$ approaches 1, the slope of the source worldline approaches the light cone slope and an observer at a given z will see further and further into the relative past of the dipole. Note also that the contours in the left panels show that the circulation direction of the dipole lobes changes sign at a value of z which gets larger for larger $v_{S,x}$. This change in sign results since observers near the horizon see fields from further in the past when the fields were pointing in a different x -direction. The increase in z -location of this turning point for larger $v_{S,x}$ can again be understood from Figure 6.12. 186

- 6.14 A 3D visualization of the magnetic fields lines and corresponding horizon currents $\mathbf{J}_{\mathcal{H}}$ (left) and electric field lines with the corresponding horizon charges $\sigma_{\mathcal{H}}$ (right) for the boosted Rindler dipole. The case shown is for $v_{S,x} = 0.2$ 187
- 6.15 An identical plot to Figure 6.10 but for the boosted case. Current density vectors (white) are overlaid on contours of charge density on the stretched horizon ($\alpha_{\mathcal{H}} = 10^{-4}$) of the boosted dipole with magnetic moment in the y -direction. The bottom panel also plots streamlines of the currents. From top to bottom, the magnitude of the boost in the x -direction increases from $v_{S,x} = 0.1, 0.5, 0.9$. Each snapshot is taken at $g_H t = 10$. The configuration drags along the stretched horizon keeping a constant lag distance behind the moving source. The induced currents can be thought of as redistributing charge in order to slide the charge distribution along behind the boosted dipole. The gray regions are regions of steeply increasing $\sigma_{\mathcal{H}}$ which have been removed to more clearly view the contour structure. The axes are in units of z_S 188
- 6.16 An $x = x_S = 0$ slice through the Poynting flux. The plane contains the source as viewed by Rindler observers for three different boost magnitudes in the x -direction, $v_{S,x} = 0.1, 0.5, 0.9$. The Poynting flux is 0 at infinity despite outward components of the field in the region plotted here. The axes are in units of z_S 189
- 6.17 Log-luminosity computed from Eq. (7.5) (blue, solid line and leftmost y-axis labels) and representative log-voltage drop on the horizon (red, dashed line and right-most y-axis labels). Luminosities and voltages are computed for $M = 10M_{\odot}, 10^2 M_{\odot}$, and $10^3 M_{\odot}$ with the dipole at Rindler height $z_S = 3M$ as a function of $v_{S,x}$ varying from 0.01 to 0.95. The last stable circular orbit in the Schwarzschild spacetime would have $v_{S,x} = 0.5$ 194

6.18 Maximum curvature radiation energies computed from Eq. (6.72) (blue, solid line and leftmost y-axis labels) and corresponding maximum γ_p (Eq. 6.75) to which electrons/positrons can be accelerated (red, dashed line and rightmost y-axis labels). Both are computed for $M = 10M_\odot$, 10^2M_\odot , and 10^3M_\odot with the dipole at Rindler height $z_S = 3M$ as a function of $v_{S,x}$ varying from 0.01 to 0.95.	199
6.19 Luminosity, voltage (top) and energy, Lorentz-factor (bottom) plots identical to those portrayed in Figures 6.17 and 6.18 respectively. Here we have plotted both panels for a 10^6M_\odot BH and at a much smaller horizon distance than in the previous Figures: $z_S = R_{lc}$, corresponding to the maximum separation where the BH-NS circuit remains connected.	201
7.1 Schematic of a Faraday loop as seen by an observer external to the horizon. The black sphere depicts the BH horizon orbiting out of the page. In green is a schematic of the instantaneous closed loop defining one of infinitely many circuits made up of electrons and positrons moving along magnetic fields lines which trace the BH horizon.	220
7.2 Total possible power supplied by the BH battery via Eq. (7.5) as a function of time until merger for two point masses undergoing orbital decay via gravitational radiation reaction (Eq. 7.8). The solid, dashed, and dot-dashed lines indicate NS surface magnetic-field strengths of 10^{12}G , 10^{14}G , and 10^{16}G respectively, for a BH mass of $10M_\odot$. The plot extends to a binary separation of GM/c^2 , the size scale of the event horizon for the maximally spinning BH we consider. We have dropped factors of G and c in the axis labels.	224

7.3	The spectra of primary curvature radiation at times corresponding to binary separations $10GM/c^2$, $6GM/c^2$, and $3GM/c^2$ (dot-dashed, dashed, solid) scaled to $B_{\text{NS}} = 10^{12}G$ (factors of G and c are omitted in the labels). We use an electron energy power law index of $p = 2.0$ and a minimum Lorentz factor set by radiation reaction in the outer magnetosphere. Dependence on both parameters is minimal (see the Appendix). The red dots indicate photon energies above which the magnetosphere is opaque to pair production via $\gamma + B$ interactions.	228
7.4	The optical depth to the different pair producing processes. The magnetosphere curvature photons are trapped by $\gamma + B$ early on, $\gamma + \gamma$ also becomes relevant for magnetosphere photons just before merger. The $\gamma + B$ optical depth is computed at a time dependent frequency near the peak of the primary curvature spectrum. Factors of G and c are omitted in the upper x-axis label.	230
7.5	The radius of the photosphere as a function of NS magnetic-field strength, for different assumed radii of energy injection $R_0 = GM/c^2, 2GM/c^2$. Factors of G and c are omitted in the figure labels.	231
7.6	The power available to the post-merger, spinning BH remnant as a function of remnant spin and NS magnetic-field strength. This power is generated from the Blandford-Znajek process and the flux of open NS magnetic-field lines, Eq. (7.28). This maximal power will decay as the remnant magnetosphere decays on the resistive timescale.	238

7.7 The spectrum of primary curvature radiation at times corresponding to binary separations of $10M$, $6M$, and $3M$ (dot-dashed, dashed, solid). Each panel is for the labeled minimum electron Lorentz factor and power law index p of electron energies. $\gamma_{\min} = \text{RadRx}$ refers to the radiation reaction limited Lorentz factor at the point of weakest electric field in the region connecting NS and BH (of order a few to 10 times smaller than the maximum Lorentz factor near merger). 245

List of Tables

2.1	<i>Summary of our simulation runs. Low, medium, and high radial and azimuthal resolutions are denoted by “Lo$\Delta r\Delta r$”; “HiΔr” and “Lo$\Delta\phi$”; “Mid$\Delta\phi$”; “Hi$\Delta\phi$” and correspond to $\Delta r/a = 0.035, 0.24, 0.017$ and $\Delta\phi/2\pi = 0.0078, 0.0052, 0.0039$, respectively. The viscosity parameter α is set to 0.01 unless otherwise specified.</i>	12
2.2	The mean accretion rate \dot{M}_{bin} , averaged over 1000 orbits in the quasi-steady-state, for binaries with different mass ratios. The rates are shown in units of the corresponding rate \dot{M}_{q0} found in a single-BH ($q = 0$) simulation. This ratio is computed as an average from 3500 to 4500 orbits unless the quasi-steady state isn’t reached until after (or for large α much before) 3500 orbits; in this case the value in the table is denoted by *. The first four rows show results for different combinations of radial and azimuthal resolutions. The first row is our fiducial resolution. The last three rows are for runs at the fiducial resolution but different magnitudes of the viscosity parameter α .	31
3.1	Parameters for the simulations run in this study. The columns, from left to right, are the binary mass ratio, the orbital Mach number, and the coefficient of kinematic viscosity.	95

7.1 Expected number of Fermi GBM events in units of [yr ⁻¹] $f_{\text{fb}}(B_{\text{NS}})$ where $f_{\text{fb}}(B_{\text{NS}})$ is the fraction of BHNS coalescences with NS magnetic-field strength B_{NS} and which will not tidally disrupt the NS and will generate the signal predicted here. B_{NS} is the NS surface magnetic-field strength.	241
--	-----

Acknowledgements

Thank my advisors

People: Family: In nyc: Nickie, Ryan, David alex, uncle steve, aunt patty Grandparents

Friends: John O'Donnell Ben Gilmour Adam Bisogne Matt Wilde Marina Kevan Crotan
(grads) Adrian Price-Whelan Jeff Andrews Lia Andrea Sarah Munier

Colleagues: Brian Farris Paul Duffell Geoffrey Ryan Nick Stone

Old grads: Destry Saul Cameron Hummels Duane Lee Erika Hamden

Faculty:

Things: DD lounge The woods The desert

2016, New York, NY

Chapter 1

Introduction

“When making an axe handle, the pattern is not far off.”

– Gary Snyder quoting Ezra Pound para-phrasing Lu Ji

- 1) Introduce the concept of gravitational radiation and its importance. a) what are GWs b) What are the expected sources of GWs c) How are GWs detected - list of experiments and their sensitivity (Figure?)
- 2) Introduce the role EM counterparts can play a) with GW detection b) without GW detection

The specific sources of GWs studied in this thesis are the inspiral and merger of MBHBs in galactic nuclei and the merger of magnetized NSs with $\gtrsim 10M_{\odot}$ BHs.

1.1 Part I: Massive Black Hole Binaries

Origin story + Final Parsec Problem

List of past work: Theory of disk+binary interaction Simulations

The first observations- Graham, D’Orazio, Charisi

1.2 Part II: Stellar Black Hole + Neutron Star Binaries.

Tidal Disruption details - and EM signals from disruption

Explanation of BH battery - Review of UI models in astro
review of GJ69 and Force Free?

Review of Pop Synth for NSBH LIGO
prospects in the LIGO era

1.3 Outline of thesis

Part I

Massive Black Hole Binaries

“The bigger they are the harder they fall”

– David

Chapter 2

Accretion into the Central Cavity of a Circumbinary Disc

2.1 Introduction

Massive black holes (MBHs) appear to reside in the nuclei of most nearby galaxies (see, e.g., reviews by Kormendy & Richstone 1995 and Ferrarese & Ford 2005). In hierarchical structure formation models, galaxies are built up by mergers between lower-mass progenitors, which deliver nuclear MBHs (e.g. Springel et al. 2005; Robertson et al. 2006), along with a significant amount of gas (Barnes & Hernquist 1992), to the central region of the newly born post-merger galaxy. Since mergers are common (e.g. Haehnelt & Kauffmann 2002), it follows that massive black hole binaries (MBHBs) should also be common in galactic nuclei.

Despite this expectation, observational evidence for MBHBs remains scarce (see, e.g., Ko mossa 2006; Tsalmantza et al. 2011; Eracleous et al. 2012). The dearth of MBHBs could be attributed to several factors: it is possible that typically only one of the two BHs is active at spatially resolvable separations; binaries may also lose their angular momentum efficiently due to the

surrounding stars and gas and quickly move to spatially unresolvable orbital separations. Another possible hindrance, which we address in this paper, is that the outward gravitational torques from the binary can balance the inward viscous torques and pressure forces, clearing a central cavity in a putative circumbinary gas disc ([Artymowicz & Lubow 1994](#)), possibly rendering the system too dim for detection. Overall, identifying MBHBs is difficult, and a better understanding of their expected observational signatures, especially those based on time-variability ([Haiman et al. 2009](#)), is needed. Merging MBHBs should be unambiguously identifiable by gravitational wave (GW) detectors, such as eLISA (?) or ongoing Pulsar Timing Arrays (e.g. [Lommen 2012](#)). Identifying the electromagnetic (EM) counterparts of these GW sources (among the many false candidate galaxies in the GW error box) will, however, likewise require an understanding of their observational signatures.

Recent studies have explored the gas-dynamics of circumbinary accretion discs around near-equal-mass binaries in some detail. Since the system is not axisymmetric, this requires a two- or three-dimensional treatment. [MacFadyen & Milosavljević \(2008\)](#) (hereafter MM08) have run two-dimensional hydrodynamical simulations for an equal-mass binary. Three-dimensional smoothed particle hydrodynamical (SPH) simulations have been carried out for equal-mass and 2:1 mass-ratio binaries by [Hayasaki et al. \(2007\)](#) and for a 3:1 mass-ratio binary by [Cuadra et al. \(2009\)](#) and ?. ? have followed up on the work of MM08 for an equal-mass binary by running 3D magnetohydrodynamical (MHD) simulations, and [Noble et al. \(2012\)](#) have further added a post-Newtonian treatment of general relativistic (GR) effects, and followed the disc through the late stages of orbital inspiral (from orbital separation $r = 20M$ to $8M$). [Farris et al. \(2011\)](#) followed the merger of an equal-mass binary and a surrounding disc through merger in full 3D general relativistic MHD, starting from $10M$. Finally, [Farris et al. \(2012\)](#) have added gas-cooling to GRMHD simulations of an equal-mass binary prior to decoupling, through decoupling and to merger starting from $10M$. A generic result of all of these studies is that a low-density cavity is carved out by the binary torque,

but gas leaks into the cavity through non-axisymmetric streams (as first discussed in the SPH simulations of [Artymowicz & Lubow 1996](#)). These streams can power significant accretion onto the binary components, and should lead to bright EM emission.

A particularly promising feature is that the accretion rate onto the BHs can be both high and strongly variable, modulated by the binary’s orbital motion. This could allow a detection of sub-pc binaries by looking for periodic variations in the luminosity of AGN-like objects ([Haiman et al. 2009](#)) or periodic shifts and intensity variations of spectral lines (e.g. [Haiman et al. 2009; Shen & Loeb 2010; Eracleous et al. 2012](#) and references therein). If the accretion remains significant and periodic down to \ll pc separations, then it could also enable the identification of EM counterparts of gravitational wave sources: either for precursors to eLISA sources in the $M = 10^5 - 10^7 M_\odot$ range ([Kocsis et al. 2006, 2008](#)) or by detecting periodic modulations of more massive $M = 10^8 - 10^9 M_\odot$ binaries discovered by pulsar timing arrays (PTAs; [Tanaka et al. 2012; Sesana et al. 2012](#)).

Although existing studies have focused on near-equal-mass MBHBs, in reality, coalescing MBHBs should have a distribution of mass ratios $q \equiv M_2/M_1$.^a Mergers occur between galaxies over a wide range of sizes, harboring central BHs of different masses, so that MBHBs resulting from galactic mergers should have a correspondingly wide range of mass ratios. Studies based on Monte-Carlo realizations of dark matter merger trees indeed find broad distributions between $10^{-2} \lesssim q < 1$, generally peaking in the range $q \sim 0.1 - 1$ (e.g. [Volonteri et al. 2003a; Sesana et al. 2005, 2012; Gergely & Biermann 2012](#)). However, the predictions depend on the occupation fraction of MBHBs, the redshift-evolution of the correlation between the masses of MBHBs and their host galaxies, as well as on the limit on the mass ratio of host galaxies whose nuclear MBHBs can coalesce; $q < 0.1$ mergers could in fact be most common (e.g. [Lippai et al. 2009](#)).

Here we follow up on the earlier work of MM08 and move beyond the near-equal-mass binary case. We study the periodicity and the time-averaged rate of accretion across the central cavity, by

^a[Nixon et al. \(2011\)](#) explored a range of mass ratios, using 3D SPH simulations, but restricted their study to retrograde discs.

running 2D hydrodynamical simulations of a circumbinary disc for 10 different binary mass ratios ranging from $q = 0.003$ to $q = 1$. Clearly, one expects that in the limit $q \rightarrow 0$, the accretion rate approaches that of an accretion disc around a single BH, and will no longer be time-variable. The main goal in this paper is to answer the following basic questions: *How does the mean accretion rate, and its fluctuations, depend on the mass ratio? In particular, down to what mass ratio is the mean accretion and/or its variability significantly affected by the binary torques?* We address these questions with the caveat that, throughout this paper, accretion is defined as the mass crossing the inner boundary of the simulation domain and not necessarily that accreted by either BH.

The rest of this paper is organized as follows: In §2.2, we describe the setup of our numerical simulations, including changes we made to the public version of the Eulerian grid code FLASH and the initial and boundary conditions we adopted. In §2.3 we present our main results, namely that we find four distinct patterns for the time-variability of the accretion rate as a function of the mass ratio q . In §2.4 we compare our findings with that of MM08 as well as investigate the dependence of our results on the magnitude of viscosity on the resolution. We also discuss scaling of the simulations to physical parameters, such as black hole mass and orbital separation, and discuss the corresponding orbital and residence times, as well as some caveats. Finally, in §6.9 we conclude by briefly summarizing our main results and their implications. The Appendix details our implementation of viscosity in polar coordinates, an important addition to FLASH.

2.2 Details of Numerical Simulations

To simulate a gas disc in the gravitational field of a binary, we use the Eulerian grid-based hydrodynamical code FLASH (Version 3.2; Fryxell et al. 2000).^b FLASH solves the volume-integrated fluid equations by solving the Riemann problem at each cell boundary. A piece-wise parabolic rep-

^bWe note that MM08 used an earlier release, Version 2, of the same code.

resentation of the fluid variables is used to interpolate between cells, i.e. FLASH is a PPM code, accurate to 2nd order in both space and time. FLASH uses a monotonicity constraint, rather than artificial viscosity, to control oscillations near discontinuities. This makes it well suited for following supersonic fluid dynamics in the inner regions of circumbinary disc. FLASH also supports polar coordinates, which is convenient for simulating discs.

2.2.1 Numerical Implementation and Assumptions

We assume a geometrically thin accretion flow with angular momentum aligned with that of the binary. This permits a decoupling of the fluid equations in the z direction, perpendicular to the plane of the disc, so that we can define height-integrated fluid variables and set up simulations in two dimensions. In follow-up studies, we plan to extend these simulations to the full three dimensions, which we expect will be important in determining the amount of inflow into a putative circumbinary cavity (for recent 3D grid-based simulations, see ? and [Noble et al. 2012](#)). In the present study, we choose 2D polar coordinates (r, ϕ) and employ FLASH to solve the following standard set of 2D hydrodynamical equations:

$$\begin{aligned} \frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \vec{v}) &= 0 \\ \frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla) \vec{v} &= \\ &\quad - \frac{1}{\Sigma} \nabla P - \nabla \Phi_{\text{bin}} + \nabla \cdot \nu \nabla \vec{v} + \nabla \left(\frac{1}{2} \nu \nabla \cdot \vec{v} \right) \\ \frac{\partial (\Sigma E)}{\partial t} + \nabla \cdot [(\Sigma E + P) \vec{v}] &= \Sigma \vec{v} \cdot (-\nabla \Phi_{\text{bin}}) \end{aligned} \tag{2.1}$$

Here Σ is the vertically integrated disc surface density, $\vec{v} = v_r \hat{r} + v_\phi \hat{\phi}$ is the fluid velocity, P is the pressure, ν is the coefficient of kinematic viscosity, E is the total internal plus kinetic energy of the fluid, $E = \epsilon + \frac{1}{2} |\vec{v}|^2$, and Φ_{bin} is the gravitational potential of the binary. The gravitational

potential is inserted into the simulation by hand, and is given by

$$\begin{aligned}\Phi_{\text{bin}}(r, \phi) = & -\frac{GM(1+q)^{-1}}{\left[r^2 + \left(\frac{a}{1+q^{-1}}\right)^2 - \frac{2ra}{1+q^{-1}} \cos(\phi - \Omega_{\text{bin}}t)\right]^{1/2}} \\ & - \frac{GM(1+q^{-1})^{-1}}{\left[r^2 + \left(\frac{a}{1+q}\right)^2 + \frac{2ra}{1+q} \cos(\phi - \Omega_{\text{bin}}t)\right]^{1/2}}.\end{aligned}\quad (2.2)$$

Here $\Omega_{\text{bin}} = (GM/a^3)^{1/2}$ is the binary's orbital frequency, a is the separation of the binary, $M_p > M_s$ are the masses of the primary and the secondary, $M = M_p + M_s$ is the total mass, and $q = M_s/M_p \leq 1$ is the mass ratio. The origin of the coordinate system is chosen to coincide with the binary's center of mass. In the case of a single point mass, we use the limit of equation (2.2) as $q, a \rightarrow 0$, $\Phi_{\text{bin}} = -GM/r$.

Note that we do not evolve the orbital parameters of the binary nor do we allow its center of mass to wander; these simulations are numerical experiments which are physically motivated in the limit of small disc mass (this assumption is justified for our physical parameter choices; see discussion in §2.4.4 below).

We neglect self-gravity of the disc. Given the local sound speed c_s , the Toomre parameter $Q \equiv c_s \Omega / (\pi G \Sigma)$ can be written as $Q \sim (H/r)(M/M_d)$, for a disc with mass M_d and vertical scale-height H in hydrostatic equilibrium. For a thin disc $H/r \lesssim 0.1$, but in all of our simulations, we choose $M \gg M_d$ and thus $Q \gg 1$, making the disc stable to gravitational fragmentation. This is justified for standard Shakura-Sunyaev discs, when the simulations are scaled to physical BH masses and sufficiently small separations (see §2.4.4 below). For comparison, we note that in their SPH simulations, Cuadra et al. (2009) and ? studied more massive discs, with $M_d \sim 0.2M$, making self-gravity important.

The pressure is given, as in MM08, by a locally isothermal equation of state,

$$P = c_s^2(r)\Sigma \quad (2.3)$$

where the sound speed is assumed to scale with radius as $c_s \propto r^{-1/2}$. For a Keplerian potential, this corresponds to a constant disc scale–height to radius ratio, $H/r = c_s/v_\phi \equiv \mathcal{M}^{-1}$, where v_ϕ is the orbital velocity in the disc and \mathcal{M} is the corresponding Mach number. Throughout our simulations, we choose the disc sound speed such that, for a Keplerian azimuthal velocity, $H/r = 0.1$ (or $\mathcal{M} = 10$) everywhere.^c

To incorporate viscosity, FLASH calculates the momentum flux across cell boundaries due to viscous dissipation (the last two terms in the second of equations 2.1). To compute ν we adopt the α prescription, $\nu = \alpha c_s H$ ([Shakura & Sunyaev 1973](#)), where α is a dimensionless parameter indicating the scale of turbulent cells, and the scale height is computed from $H = c_s r / v_\phi$ with v_ϕ being the Keplerian value. Following MM08, we choose a fiducial, constant $\alpha = 0.01$ (although we explore the effects of increasing α in §2.4.2). Since the FLASH viscosity implementation is not fully supported in polar coordinates, we made adjustments to the routines that compute the momentum flux and viscous diffusion time from ν , originally in Cartesian coordinates. These modifications and tests of our polar viscosity implementation are detailed in the Appendix.

2.2.2 Numerical Parameter Choices

The inner edge of the computational domain is chosen at $r_{\min} = a$ and the outer edge at $r_{\max} = 100a$. Although we are only interested in the inner few r/a of the disc, extending the computations to larger radii acts as a buffer for small initial numerical transients, and also provides a potential

^cSince we add the binary’s quadrupole potential, and the non-zero pressure makes the azimuthal velocities slightly sub-Keplerian, in practice \mathcal{M} in the simulations approaches ~ 18 near $r = a$ and becomes a constant $\mathcal{M} \sim 10$ at $r > 5a$.

reservoir of gas from which the inner regions can be fed. As in other Eulerian codes, FLASH uses a boundary zone of ‘guard cells’ which enforces boundary conditions. We choose a diode-type inner boundary condition: the values of fluid variables in cells bordering the guard cells are copied into the guard cells, with the restriction that no fluid enter the domain, $v_r(r_{\min}) \leq 0$. We adopt an ‘outflow’, outer boundary condition: this is identical to the diode version, except that flow is allowed both into and out of the domain. In practice, with our initial conditions in §2.2.3, we find an outflow at the outer boundary of our disc. However, the simulation does not run for a significant fraction of a viscous time at the outer radius (see below) and we expect inflow at the outer boundary to establish itself eventually.

Unless specified otherwise, the spatial resolution is fixed throughout the grid, with the default values set at $[\Delta r/a, \Delta\phi/2\pi] \simeq [0.024, 0.0078]$, corresponding to a grid of $\sim 4096 \times 128$ cells. The wavelength of a density wave due to a Lindblad resonance is of order $2\pi H \sim 0.6r$ (e.g. Dong et al. 2011a; ?); as in MM08, the radial resolution is chosen to resolve this length scale with many cells. Note that our fiducial resolution results in cell aspect ratios which are approximately square at $r \sim 3a$. To test numerical convergence (see §2.4.3 below), we performed several additional runs, increasing the radial and azimuthal resolutions, by factors of 1.42 and $1.42^2 \sim 2.0$, from the lowest resolution runs. The time resolution is set to be half of the shortest propagation time (viscous or dynamical) across a cell. The effects of changing the resolution are discussed in §2.4.3.

We run our simulations for between 4×10^3 and 10^4 binary orbits. For reference, we note that the viscous time can be related to the orbital time as

$$t_{\text{visc}} = \frac{2r^2}{3\nu} = \frac{\mathcal{M}^2}{3\pi\alpha} t_{\text{orb}} \simeq 1060 \left(\frac{\mathcal{M}^2}{100} \right) \left(\frac{0.01}{\alpha} \right) t_{\text{orb}}. \quad (2.4)$$

Thus, our typical run of 4000 binary orbits corresponds to ~ 4 viscous times at the innermost regions, but less than one viscous time at $r \gtrsim 3a$.

Table 2.1: Summary of our simulation runs. Low, medium, and high radial and azimuthal resolutions are denoted by “ $\text{Lo}\Delta r$ ”; “ $\text{Mid}\Delta r$ ”; “ $\text{Hi}\Delta r$ ” and “ $\text{Lo}\Delta\phi$ ”; “ $\text{Mid}\Delta\phi$ ”; “ $\text{Hi}\Delta\phi$ ” and correspond to $\Delta r/a = 0.035, 0.24, 0.017$ and $\Delta\phi/2\pi = 0.0078, 0.0052, 0.0039$, respectively. The viscosity parameter α is set to 0.01 unless otherwise specified.

Mass ratio q	Spatial Resolution $[\Delta r, \Delta\phi]$	No. Orbits ($N_{\text{orb}} = t_{\text{max}}$)
1.0	$[\text{Lo}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Mid}\Delta\phi], [\text{Hi}\Delta r, \text{Hi}\Delta\phi]$	5500, 6800, 4500, 4200
1.0 ($\alpha = 0.02$)	$[\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	4500
1.0 ($\alpha = 0.04$)	$[\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	5200
1.0 ($\alpha = 0.1$)	$[\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	1400
0.75	$[\text{Lo}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	6500, 4500
0.5	$[\text{Lo}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	6000, 4500
0.25	$[\text{Lo}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	5700, 4500
0.1	$[\text{Lo}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Mid}\Delta\phi], [\text{Hi}\Delta r, \text{Hi}\Delta\phi]$	7000, 8000, 4500, 4100
0.075	$[\text{Lo}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	4500, 5200
0.05	$[\text{Lo}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Mid}\Delta\phi], [\text{Hi}\Delta r, \text{Hi}\Delta\phi]$	4500, 7600, 4500, 4200
0.025	$[\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	4500
0.01	$[\text{Lo}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	5200, 5500
0.003	$[\text{Mid}\Delta r, \text{Lo}\Delta\phi]$	5500
0 (All α)	$[\text{Lo}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Lo}\Delta\phi], [\text{Mid}\Delta r, \text{Mid}\Delta\phi], [\text{Hi}\Delta r, \text{Hi}\Delta\phi]$	6000, 8000, 4500, 4200

We have performed 34 runs altogether, for 10 different binary mass ratios between $q = 0.003$ and $q = 1.0$ (including control runs for $q = 0$, i.e. a single BH). The mass ratio, resolution, and the number of binary orbits followed in each of our simulation runs are summarized in Table 2.1.

2.2.3 Initial Conditions

In our initial conditions, we insert a central cavity around a binary, and also include a density pile-up just outside the cavity wall. Such a surface density profile is expected to develop during the inward migration of the secondary. When the secondary arrives at the radius where the local disc mass is too small to absorb the secondary’s angular momentum, its migration stalls, the inner disc drains onto the primary, and continued accretion from larger radii causes a pile-up of gas outside the secondary’s orbit (Syer & Clarke 1995; Ivanov et al. 1999; Milosavljević & Phinney

2005; Chang et al. 2010; ?). As emphasized recently by Kocsis et al. (2012b), the details of this process are still uncertain, as the coupled time-dependent migration, cavity formation, and pile-up, has not been modeled self-consistently, even in one-dimensional calculations. However, using self-consistent *steady-state* solutions, Kocsis et al. (2012a) showed that in many cases, the pile-up can cause overflow already at large binary separations.

For simplicity, and for ease of comparison, we adopt the same initial surface density profile as in MM08. This profile is motivated by the earlier results of Milosavljević & Phinney (2005); it has very little gas inside of $r \simeq 3a$, and peaks at $\sim 8a$,

$$\Sigma(r, t_0) = \Sigma_0 \left(\frac{r_s}{r} \right)^3 \exp \left[- \left(\frac{r_s}{r} \right)^2 \right]. \quad (2.5)$$

Here Σ_0 is an arbitrary constant, and $r_s = 10a$. The initial profile is shown by the [blue] dashed curves in Figure 2.4 below.

We also follow MM08, and incorporate pressure gradients and the quadrupole contribution of the binary's potential into the initial azimuthal velocity,

$$\Omega^2 = \Omega_K^2 \left[1 + \frac{3}{4} \left(\frac{a}{r} \right)^2 \frac{q}{(1+q)^2} \right]^2 + \frac{1}{r\Sigma} \frac{dP}{dr} \quad (2.6)$$

and account for viscous drift in the initial radial velocities,

$$v_r = \frac{d}{dr} \left(r^3 \nu \Sigma \frac{d\Omega}{dr} \right) \left[r \Sigma \frac{d}{dr} (r^2 \Omega) \right]^{-1}. \quad (2.7)$$

We emphasize that with these initial conditions, the disc is not initially in equilibrium, and material diffuses away from the peak of the surface density, both inward and outward, due to pressure gradients. However, after running the simulations for several thousand orbits, the system relaxes to a steady pattern of accretion, and we do not expect the initial profile to significantly

influence our conclusions.

2.2.4 Disc Parameters and Accretion Rate

Our primary goal is to quantify the magnitude and variability of accretion across the central cavity. To this end we compute the time-dependent accretion rate at the inner edge of the simulation domain, $r_{\min} = a$

$$\dot{M}(t) = \int_0^{2\pi} \Sigma(r_{\min}) v_r(r_{\min}) r_{\min} d\phi. \quad (2.8)$$

Since we are unable to track the fate of the gas at smaller radii, nor do we allow the masses of the BHs to increase, this mass is effectively lost from the simulation. In practice, the total mass that is lost is a small fraction of the total initial disc mass (at most a few % by the end of each run).

Because we neglect the self-gravity of the disc, equations (2.1) are independent of Σ_0 , and there is no unique way to assign a physical normalization to \dot{M}_{sim} without appealing to a disc model which includes additional physics. Instead, MM08 compare the average accretion rate at the edge of the integration domain ($r_{\min} = a$) to that in a disc in a Keplerian potential with the same surface density at $r \simeq 3a$,

$$\frac{\dot{M}_{\text{bin}}}{\dot{M}_{\text{free}}} \simeq \frac{\langle \dot{M}_{\text{sim}}(a) \rangle_t}{6\pi\alpha} \left(\frac{H}{r}\right)^{-2} (GMr)^{-1/2} \Sigma^{-1}(3a). \quad (2.9)$$

MM08 find $\dot{M}_{\text{bin}}/\dot{M}_{\text{free}} \simeq 0.2$. To make this comparison meaningful, one must assume that $\dot{M}_{\text{free}}(3a) \sim \dot{M}_{\text{free}}(a)$, *i.e.* that the reference, fictitious point-mass disc is in steady-state. However, for a steady-state disc, specifying α and H/r (along with the dominant source of opacity) sets the physical value \dot{M}_{free} . For our fiducial values of $M = 10^7 M_\odot$ and $a = 10^3 r_S$ (and with $\alpha = 0.01$ and $H/r = 0.1$), we find the unphysically large value of $\dot{M}_{\text{free}} \sim 10^7 \dot{M}_{\text{Edd}}$, a result of the above choices (where \dot{M}_{Edd} is the accretion rate that would produce the Eddington luminosity, with a radiative efficiency of 10%). If we instead require $\dot{M}_{\text{free}} \sim \dot{M}_{\text{Edd}}$, then this would translate to a

much thinner disc, with $H/r \simeq 4 \times 10^{-3}$. In such a thin/cold disc, resolving density waves with the same number of cells as in MM08 would require a radial grid resolution ~ 17 times higher than in our highest-resolution run, and would be impractical.

Rather than attempting to compare our binary simulations to a hypothetical steady-state point-mass disc, we choose to leave the surface density normalization Σ_0 essentially arbitrary, and instead perform explicit reference simulations with a single BH (i.e. $q \rightarrow 0$), with the same initial conditions as the binary runs. This approach has the advantage of explicitly isolating the effect of turning on/off the binary. Note, however, that our point-mass runs should *not* be expected to produce the steady-state accretion rate for a corresponding single-BH system. This is because our initial conditions are far from this state, and we do not run the simulations long enough (i.e. for a few viscous time at the outer edge) to allow it to settle to the correct steady-state.

2.2.5 Tests of Code Implementation

To ensure that non-axisymmetric perturbations are not induced artificially in the disc, we simulate a disc around a single black hole for 10^4 binary orbits. To keep axisymmetry during these runs, we re-implemented the FLASH2 routine UNBIASED-GEOMETRY (UBG)^d into the FLASH3 routines. UBG cleans up round-off errors in the cell boundary positions, and forces the polar grid to keep cell sizes uniform in the azimuthal direction. Without UBG, small perturbations in the azimuthal direction grow to significant size after a few hundred orbits. We have verified that after re-implementing UBG, axisymmetry is preserved for 10^4 binary orbits.

A detailed description as well as multiple tests of our viscosity implementation are laid in out in the Appendix.

^dThis routine is titled *clean_last_bits* in the FLASH2 download.

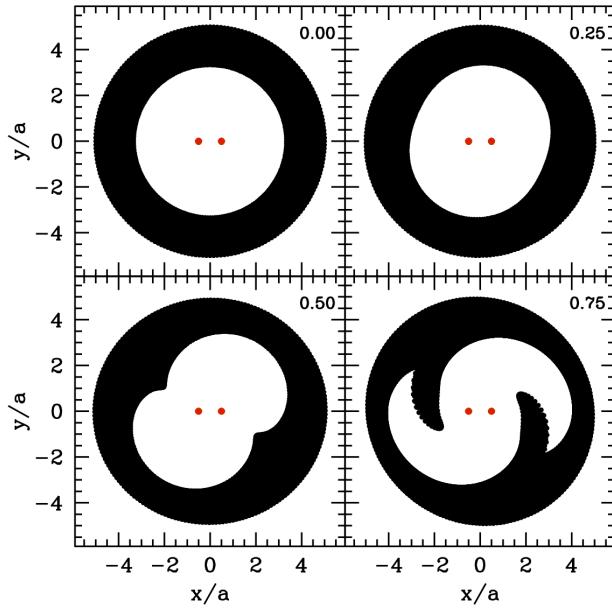


Figure 2.1: The distortion of a disc of massless test particles, initially in circular orbits around the center-of-mass of an equal-mass binary, with a central cavity. The panels show snapshots of the locations of the disc particles after 0.25, 0.5, and 0.75 binary orbits, as labeled. The orbits of the particles were followed by solving the restricted three-body problem, and are shown in a frame co-rotating with the binary. The binary point masses are marked by the two [red] dots at $x = \pm 0.5a$. The figure illustrates the tendency of the binary to create streams of particles entering the central cavity, due to gravity alone.

2.3 Results

2.3.1 Equal-Mass Binary

We begin by describing the results of our equal-mass binary runs in some detail. Although these are very similar to those of MM08, this will serve as a useful point of comparison for our unequal-mass runs.

2.3.1.1 A Toy Model with Massless Particles

Before showing the results from our simulations, we consider a simple toy model, based on the orbits of non-interacting massless test particles around a binary. We populate a 2D disc with test particles, centered on the binary’s center of mass, and leave a central cavity. We assign initial velocities equal to the Keplerian velocities around a single point mass $M = M_p + M_s$. We then follow the orbit of each test particle in the rotating binary potential, by numerically solving the restricted three-body problem for each individual particle (for 10^5 particles in practice, using equations 3.16 and 3.17 in [Murray & Dermott 2000](#)).

The results of this simple exercise are displayed in Figure 2.1, which shows the locus of the test particles initially, as well as after 0.25, 0.5, and 0.75 binary orbits (in a frame co-rotating with the binary). As this figure demonstrates, there is a tendency for the binary to pull streams of particles into the cavity. This, of course, is purely a gravitational effect. As [Artymowicz & Lubow 1996](#) have pointed out, such mass flows occur near unstable co-rotation equilibrium points in the binary potential.

The toy-model can not be pushed much further in time, since after ~ 0.75 binary orbits, the trajectories of the test-particles cross – this necessitates a hydrodynamical treatment. Nevertheless, the figure does suggest that an empty cavity can not be maintained by an equal-mass binary, even in the absence of pressure or viscosity. Furthermore, as we will see below, the simulations show accretion streams with morphologies quite similar to those in the bottom right panel of Figure 2.1.

2.3.1.2 Hydrodynamical Evolution: Reaching Steady State

We next present the results from our equal-mass binary simulations. The disc evolves through two distinct stages. As explained above, the disc is set up to be out-of-equilibrium, and we observe an initial transient state, which lasts for ~ 2500 orbits. We expect the details of this state to depend

on the initial conditions. The disc then settles to a quasi-steady-state, which persists for the rest of the simulation. In this quasi-steady-state, the disc exhibits significant accretion, which varies periodically on two timescales, $(1/2)t_{\text{bin}}$ and $\sim (5 - 6)t_{\text{bin}}$. We expect these latter features to be robust and insensitive to initial conditions.

Transient State. Initially, pressure forces and viscous stresses act to move material inside of the initial density peak at $r \sim 8a$ inward while pressure gradients move material outside of the density peak outward. Once the inner disc material reaches $r \simeq 2a$, its surface density becomes strongly perturbed and highly non-axisymmetric. Reminiscent of the evolution of the toy model in §2.3.1.1, two narrow point-symmetric streams develop, in which the gas flow becomes nearly radial. About 3% of the material in these streams exits the integration domain at $r_{\min} = a$. The rest gains angular momentum from the faster moving black holes and is flung back toward the bulk of the disc material at $r \simeq 2a$. This process maintains a central cavity within the disc, with gas streams being pulled in and pushed out on a period of $\sim 1.5 \Omega_{\text{bin}}$.

As more disc matter flows in from the initial density peak to the inner $r \sim 2a$ region, the streams become more dense. When these streams are flung back out and hit the opposing cavity wall, they generate noticeable over-densities, and deform the circular shape of the cavity to become eccentric, though still point-symmetric. These over-densities then rotate at the disc's orbital velocity. They spread out and propagate into the disc as differential rotation causes them to wind up, creating a point-symmetric ($m = 2$), rotating spiral pattern.

Quasi-Steady State. After the initial ~ 2500 orbits, the point symmetry of the transient state breaks down as stream generation becomes preferentially stronger on one side of the cavity. This causes more mass, in the form of a stream, to be driven into the opposite cavity wall, pushing that side of the wall farther away from the binary.

This lopsided state can grow from a small initial asymmetry, through a genuine physical instability, as follows:

Initially, in the transient point-symmetric state, the central cavity has an elliptical (but still point-symmetric) shape, which rotates along with the disc, its inner edge completing a rotation once every 3 binary orbits. Streams are simultaneously pulled in from the two near sides of this elliptical cavity. After the streams form, they are flung across the cavity, and hit the region approximately diagonally across, close to the azimuth where the opposite stream formed. The above process makes the cavity more eccentric over time, since when the outward-going stream material hits the cavity wall, it pushes it outward, further away from the binary's center of mass.

Figure 2.2 shows the above, via three snapshots of the disc in the transient, point-symmetric state, in time order from left to right. The first panel shows the initial formation of the two opposing streams, demonstrating that the streams originate at locations where the cavity wall is closest to the binary's center of mass. The third panel shows the collision of the outward-going streams with the opposing cavity wall, demonstrating that the streams collide with the cavity wall approximately diagonally across their initiation point. The second panel shows the morphology of the streams half-way between these time steps, for the sake of completeness.

Next, imagine that due to numerical noise, one stream (say, from side "A" of the cavity) carries slightly more momentum than the spatially opposite stream emanating from side "B". This can happen due to a small initial lopsidedness in the shape of the cavity, with side "A" being closer to the binary than side "B" (or due to an asymmetry in the density or velocity field). In our simulations, this can only be due to small numerical noise, but in reality, discs will obviously not be perfectly symmetric, either. The stronger stream will hit side "B" of the cavity as before, but will push the cavity wall farther away from the binary than the comparatively smaller counterpart stream hitting side "A". It is easy to see that this can lead to a runaway behavior: side "A" of the cavity will have absorbed less momentum, and will now be even closer to the binary's center of mass, relative to side "B" - causing a larger asymmetry in the next pair of streams, which further increases the lopsidedness of the cavity, and so on

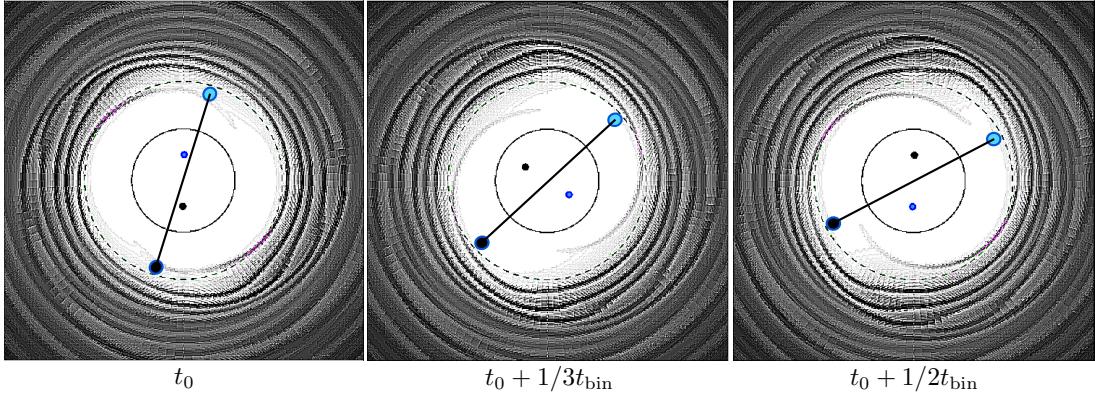


Figure 2.2: Snapshots of the surface density for the $q = 1$ disc in the point-symmetric transient state after ~ 1350 binary orbits. The snapshots are sharpened (masked to include only high Fourier frequencies in the image) in order to see the streams more clearly. The connected outer circles are drawn to guide the eye. These circles rotate with the disc structure at a period of $\approx 3t_{\text{bin}}$. Streams are generated on either side of the cavity, (shown in the left panel; the streams shown in this panel are moving inward, away from the cavity wall). These streams ultimately crash into the cavity wall on the diagonally opposite sides of the cavity, near the site where the opposing stream was generated (shown in the right panel; the streams shown in this panel are moving outward, toward the cavity wall). If a small asymmetry causes one stream to become stronger, then a runaway process would ensue pushing one side of the disc further from the binary and allowing the other side to come closer; this process could ultimately be responsible for the observed lopsided shape of the cavity.

This reenforcement-feedback process continues for a period of $\gtrsim 200$ -300 orbits, after which the weaker stream, and its effect on the cavity wall structure disappears entirely. The central cavity takes on a lopsided shape, with a near-side where streams are pulled from the cavity wall by each passage of the holes, and a far side where non-accreted material from the streams is flung back and crashes into the cavity wall. At the azimuthal locations of these crashes, the far-side of the wall develops very strong shocks, with Mach numbers up to $\mathcal{M} \sim 15$. (However, since our disc is locally isothermal, this is likely an upper limit for the shock strength). The lopsided cavity precesses in the frame at rest with respect to the binary center of mass, completing a rotation once every ~ 400 binary orbits. Excitation of a similar lopsided cavity is observed in the 3D MHD, as well as 2D hydrodynamical, simulations of ?, [Noble et al. 2012](#), and MM08. ? explore the possible generation mechanisms of this mode. They find its growth to be consistent with being caused by asymmetric stream impacts described above.

We show examples of the two-dimensional surface density distributions in Figure 2.3. The top row of this figure shows snapshots at ~ 1000 binary orbits (left) and at ~ 4000 binary orbits (right), of the inner 6% of the simulated disc (i.e. $\pm 6r/a$ are shown in both directions). The solid circle marks the inner boundary of the simulation domain at $r = r_{\min} = a$. The larger dotted circle at $r \simeq 2.08a$ is the position of the $(m, l) = (2, 1)$ outer Lindblad resonance (not present, but shown for reference). The left of these two panels illustrates the point-symmetric, transient state. The zoomed-in inset of this panel shows two weak point-symmetric streams reminiscent of the streams seen in the toy model of Figure 2.1. The right panel shows the disc after it has settled to its quasi-steady, lopsided state, with a single stream.

The top left panel of Figure 2.4 shows snapshots of the azimuthally averaged surface density profile of the equal-mass binary disc at three different times, after 0, 2000, and 4000 orbits. For comparison, the density profile is also shown for the single point-mass ($q = 0$) case, after 4000 orbits. As the figure shows, the inner circumbinary disc spreads inward with time. By 4000 orbits,

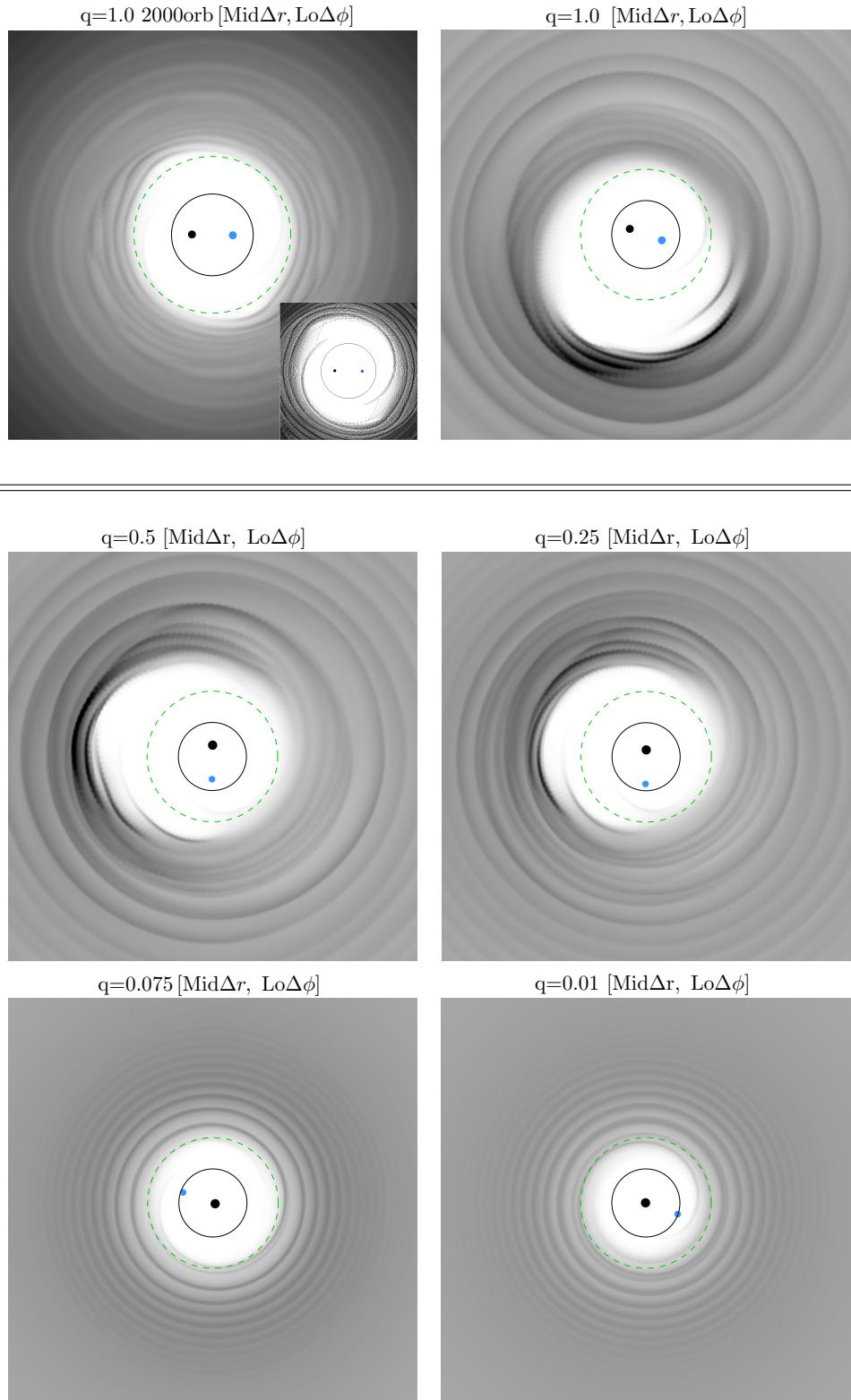


Figure 2.3: *Top row:* Surface density distributions for the equal-mass ratio ($q = 1.0$) binary during a transient, point-symmetric state after ~ 1000 binary orbits (left) and during the quasi-steady asymmetric state after ~ 4000 binary orbits (right). The inset in the top left panel zooms in to the inner $\pm 2.5r/a$ of the disc in order to show the stream morphology. *Bottom two rows:* snapshots at ~ 4000 binary orbits, during the quasi-steady-state phase, for mass ratios $q = 0.5, 0.1, 0.075$, and 0.01 .

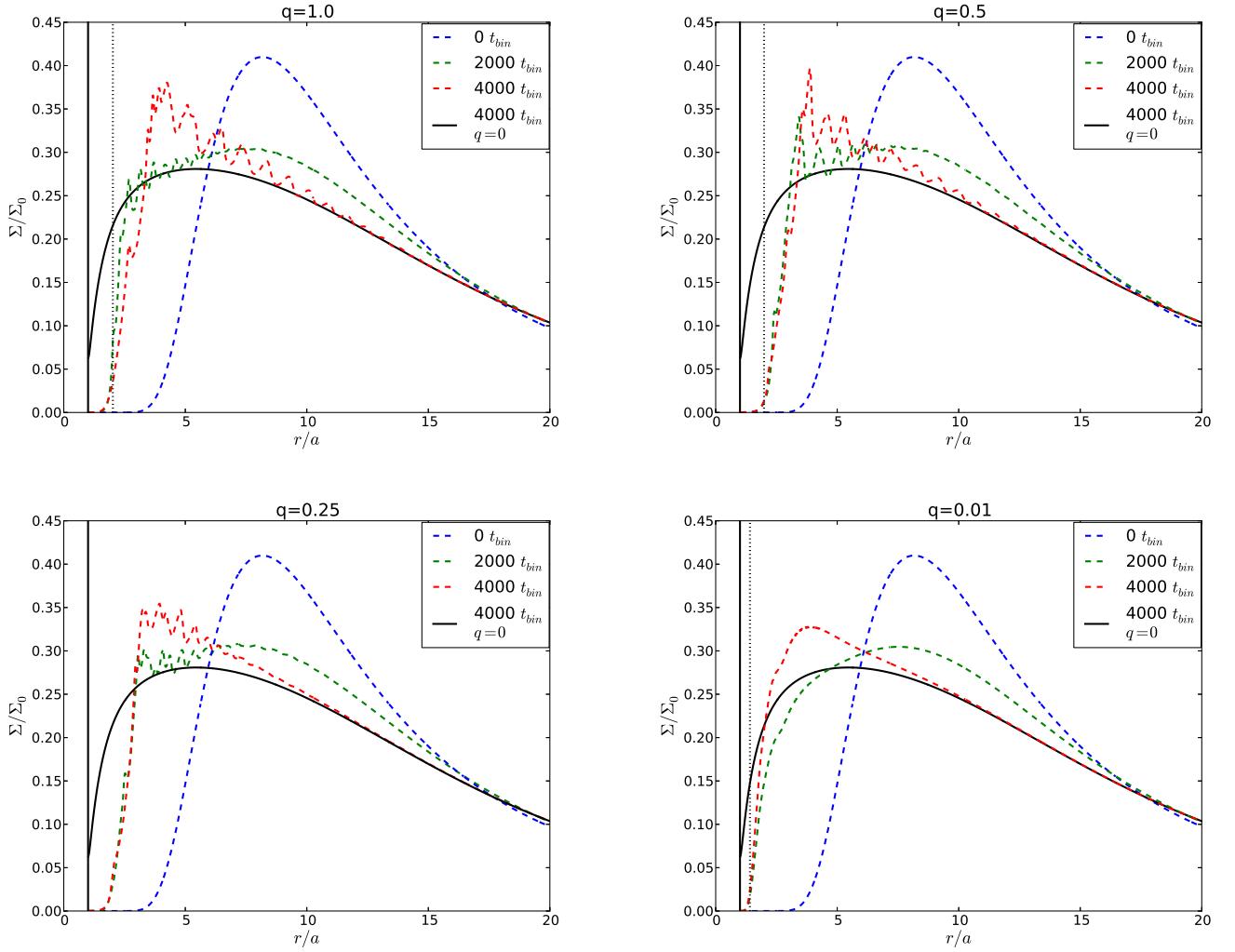


Figure 2.4: Snapshots of the azimuthally averaged disc surface density at different times (shown by different curves in each panel, from 0 to 4000 orbits, as labeled), and for different mass ratios (shown in different panels, from $q = 1.0$ to 0.01 , as labeled). In each panel, the solid [black] curve shows, for reference, the density profile in the point-mass ($q = 0$) case after 4000 orbits. The vertical dotted lines mark the radius where binary and viscous torques balance (from Figure 2.5); these lie close to the observed cavity edges. The vertical solid lines mark the inner edge of the integration domain ($r = r_{\min} = a$). In each case, the inner circumbinary disc spreads inward with time, but the density profile remains sharply truncated, with a low-density central cavity inside $r \lesssim 2a$.

the disc structure at $r/a \gtrsim 5$, where the effect of the binary is relatively small, closely follows the $q = 0$ case. However, the density profile remains sharply truncated inside $r \lesssim 2a$ (i.e. below the point-mass case, even at 4000 orbits). As a result of the initial density profile, the peak density first decreases as matter drains inward towards the holes and also outward towards the outer boundary. With time, despite the leakage of streams to the binary, inward viscous diffusion causes a gas pileup behind the cavity wall. The figure also shows that the position of the cavity wall moves slightly outward between the $t=2000$ and $t=4000$ snap-shots. This is because as the disc settles to its lopsided quasi-steady-state, the cavity size grows in the azimuthally averaged sense.

2.3.1.3 Torque Balance and the Size of the Central Cavity

As (the top left panel of) Figure 2.4 shows, the central cavity around the equal-mass binary extends to $r \sim 2a$. Here we take the cavity edge r_{ce} to be the radius where the negative viscous torque density matches the binary torque density (in an azimuthally averaged sense) ,

$$\left[\left(\frac{dT}{dr} \right)_{\text{bin}} + \left(\frac{dT}{dr} \right)_{\text{visc}} \right]_{r_{ce}} = 0. \quad (2.10)$$

To gain insight into the transport of angular momentum and the clearing of the central cavity, we therefore compute the time- and azimuthally-averaged torque densities from the binary potential,

$$\left(\frac{dT}{dr} \right)_{\text{bin}} = \left\langle \frac{1}{2\pi} \int_0^{2\pi} \Sigma(r, \phi) \frac{d\Phi}{d\phi}(r, \phi) r d\phi \right\rangle_t, \quad (2.11)$$

and from viscous stresses,

$$\left(\frac{dT}{dr} \right)_{\text{visc}} = 2\pi \left\langle \frac{d}{dr} \left[r^3 \nu \left\langle \Sigma \frac{\partial \Omega}{\partial r} \right\rangle_\phi \right] \right\rangle_t. \quad (2.12)$$

The outer derivative in equation (2.12) is taken numerically, and all of the above values are measured directly from the simulation outputs, except for the binary potential derivative in equation (2.11) which is given analytically. The time averages are taken over 25 binary orbits at a sample rate of 20 per orbit.

The top row of Figure 2.5 shows the binary and viscous torque densities for $q = 1$ over the inner $5r/a$ of the disc, during both the initial transient state (left panel) and the subsequent quasi-steady-state (right panel). There is indeed a well-defined central region, where the binary torques exceed the viscous torques and can be expected to clear a cavity. The transition (computed via eq. 2.10) is located at $r_{ce} \simeq 1.85$ and $r = 2.02a$ in the transient and quasi-steady-state, respectively, and is marked in both panels by a vertical dotted line. These vertical lines are also shown in Figure 2.4 and indeed lie very close to radii where the disc surface densities remain truncated.

The small outward drift of the average position of the cavity wall from the transient to the quasi-steady-state is also visible in Figure 2.4. As the disc transitions to the quasi-steady-state, the binary torque-density wavelength increases in r/a , while keeping approximately the same amplitude. These effects can be attributed to the increasingly elongated and lopsided shape of the inner cavity; when azimuthally averaged, this results in a larger cavity size.

2.3.1.4 Accretion Rates

The most interesting consequence of the lopsided cavity shape is on the accretion rate. In the top left pair of panels in Figure 2.6, we show the accretion rate, measured across the inner boundary of the simulation ($r_{\min} = a$) during the quasi-steady-state, after 4000 binary orbits. The upper panel shows the accretion rate as a function of time for ~ 16 binary orbits; the solid horizontal [blue] line shows the time-averaged accretion rate during this time, and, for reference, the horizontal dashed [green] line shows the accretion rate over the same orbits in the $q = 0$ reference simulation. The average accretion rate onto the binary is approximately $2/3$ of the accretion measured for the $q = 0$

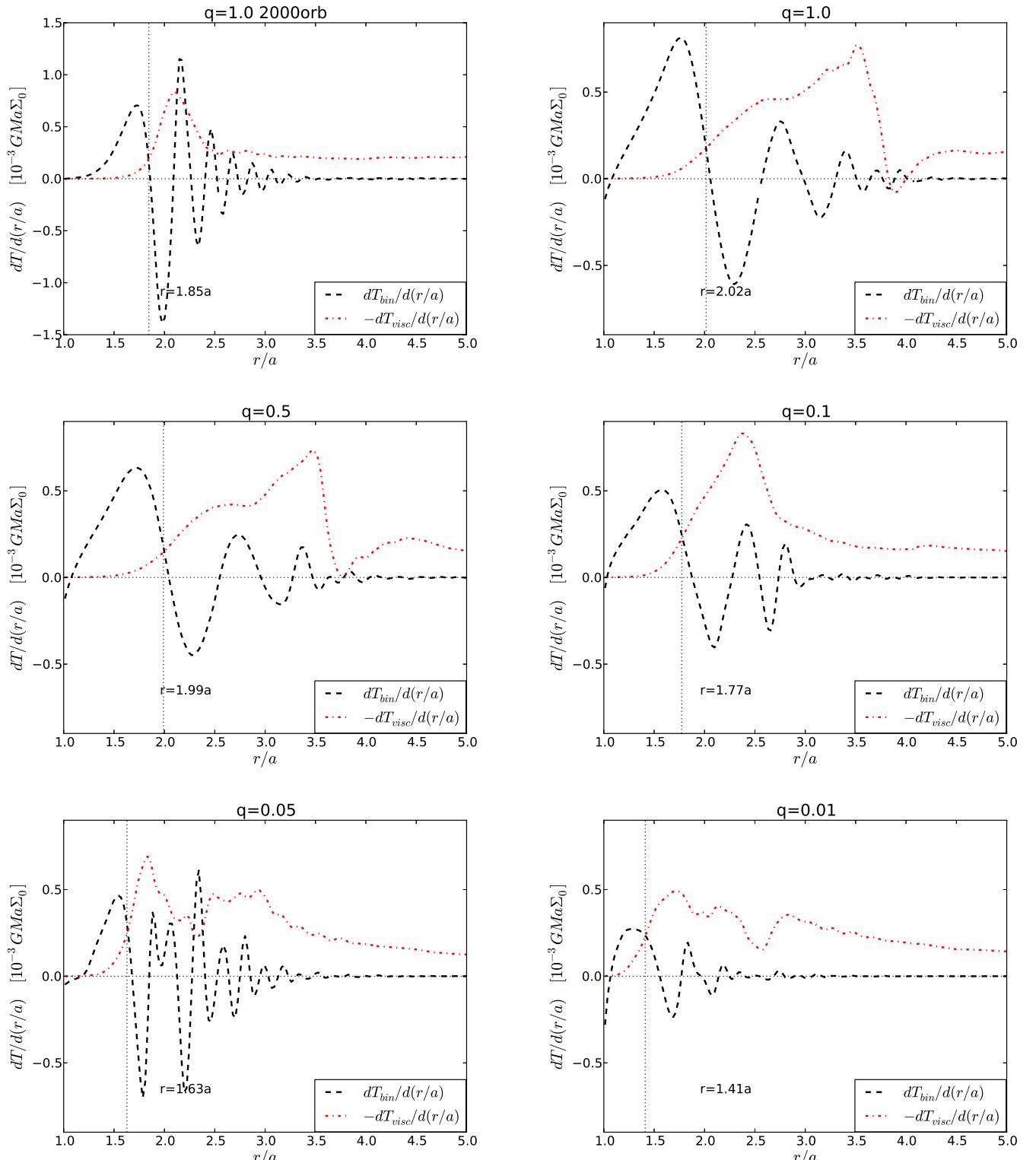
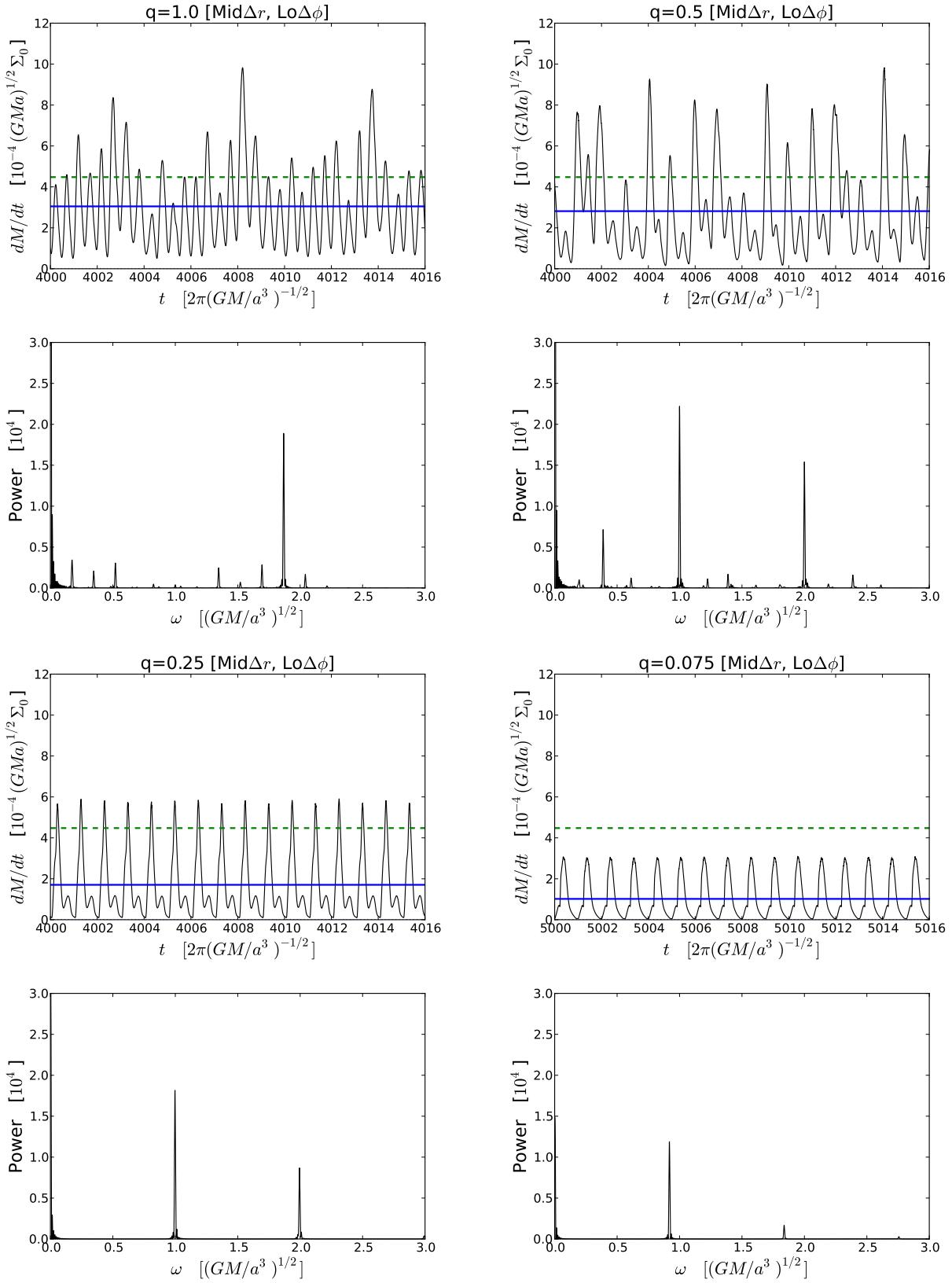


Figure 2.5: Azimuthally- and time-averaged torque density profiles in the inner disc for the equal-mass binary (top two panels) and for unequal-mass binaries (other panels, with different mass ratios q as labeled). The top left panel corresponds to the point-symmetric transient stage (after ~ 2000 orbits) and the top right panel to the asymmetric quasi-steady state (after ~ 4000 orbits). Only the quasi-steady state is shown for the $q < 1$ cases. In each panel, the dashed [black] curves show the gravitational torques from the binary, and the red [dot-dashed] curves show the *negative*



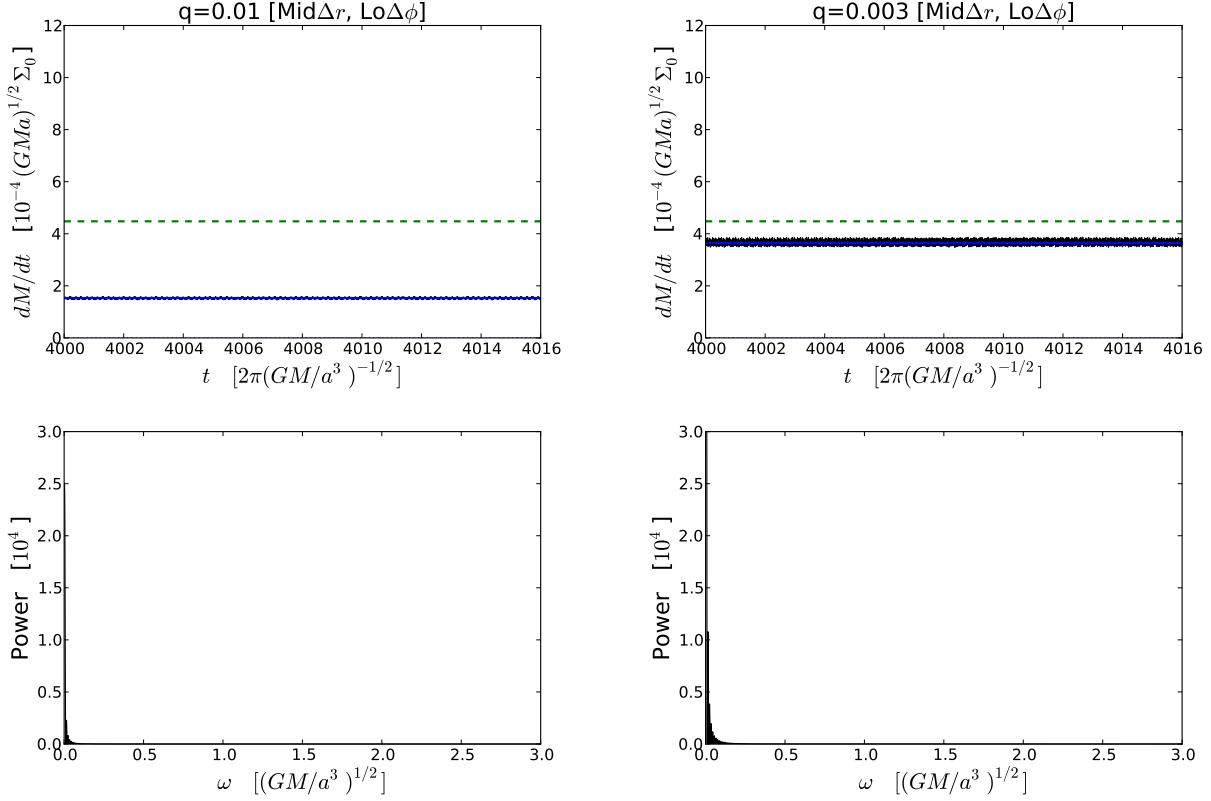


Figure 2.6: The time variable accretion rate across the inner boundary of the simulation measured at $r = r_{\min} = a$ (top of each pair of panels) and the corresponding Lomb-Scargle periodogram (bottom of each pair) computed over 100 orbits. Panels are displayed in order of decreasing binary mass ratio, starting from $q = 1.0$ at top left (on the previous page) to $q = 0.003$ at the bottom right (on this page). The average accretion rate in each panel is denoted by the solid [blue] horizontal line. The dashed [green] horizontal line in each plot shows the average accretion rate for the point mass ($q = 0$) case for reference. For $q > 0.05$, the accretion rate is strongly modulated by the binary, with either one, two, or three distinct periods present simultaneously, depending on the value of q (see text for detailed explanations). For $q \lesssim 0.05$, the binary still reduces the mean accretion rate noticeably, but does not imprint strong periodicity; the $q = 0.003$ binary is nearly indistinguishable from a single BH.

run at the same resolution. Note that this ratio stays constant over the course of the quasi-steady-state. The lower panel shows the corresponding Lomb-Scargle periodogram, measured over 100 binary orbits at a sampling rate of once per simulation time step (~ 500 per orbit).

As mentioned above, once the quasi-steady-state is reached, the accretion rate increases from that in the transient state by an order of magnitude. As Figure 2.6 shows, it also begins to exhibit strong (factor of ~ 3 above average) variability. Although this figure samples the accretion only between 4000 and 4016 orbits, the pattern is remarkably steady, and repeats itself until the end of the simulation (However, see bullet 6 in §2.4.2).

The accretion is clearly periodic, and displays two prominent periods, $(1/2)t_{\text{bin}}$ and $\sim 5.7t_{\text{bin}}$. The stronger variability at one half the orbital-time is due to the passage of each black hole by the near side of the lopsided disc and the corresponding stripping of gas streams from the cavity wall. These streams are then driven into the opposite side of the cavity (as seen in the top right panel in Figure 2.3; approximately 135 degrees from the generation point). The second, longer timescale corresponds to the orbital period at the cavity wall. As mentioned above, when the non-accreted material from the streams hits the far-side of the cavity, it creates an over-density which orbits at the disc's orbital period there which ranges from $\sim 2\pi(2.0a)^{3/2}(GM)^{-1/2} \sim 2.9t_{\text{bin}}$ out to $\sim 2\pi(3.3a)^{3/2}(GM)^{-1/2} \sim 6.0t_{\text{bin}}$. The larger streams pulled from the lump in turn create a new lump and the cycle repeats once every $\sim 5.7t_{\text{bin}}$. Similar over-dense lumps have also been found and described in the 3D MHD simulations of ?; more recently, ?? have also mentioned the contributions of such lumps to fluctuations in the accretion rate.

2.3.2 Unequal-Mass Binaries

We next turn to the main new results of this paper, and examine the disc behaviour as a function of the mass ratio. We start with a qualitative description of how the accretion pattern changes as we decrease q .

Table 2.2: The mean accretion rate \dot{M}_{bin} , averaged over 1000 orbits in the quasi-steady-state, for binaries with different mass ratios. The rates are shown in units of the corresponding rate \dot{M}_{q0} found in a single-BH ($q = 0$) simulation. This ratio is computed as an average from 3500 to 4500 orbits unless the quasi-steady state isn't reached until after (or for large α much before) 3500 orbits; in this case the value in the table is denoted by *. The first four rows show results for different combinations of radial and azimuthal resolutions. The first row is our fiducial resolution. The last three rows are for runs at the fiducial resolution but different magnitudes of the viscosity parameter α .

q	1.0	0.75	0.5	0.25	0.1	0.075	0.05	0.025	0.01	0.00
$\dot{M}_{\text{bin}}/\dot{M}_{q0}$ [Mid Δr , Lo $\Delta\phi$]	0.671	0.700	0.655	0.382	0.355*	0.228*	0.025	0.100	0.341	0.81
$\dot{M}_{\text{bin}}/\dot{M}_{q0}$ [Lo Δr , Lo $\Delta\phi$]	0.544	0.518	0.499	0.406	0.021	0.027	0.055	0.147	0.381	
$\dot{M}_{\text{bin}}/\dot{M}_{q0}$ [Mid Δr , Mid $\Delta\phi$]	0.821				0.426		0.028			
$\dot{M}_{\text{bin}}/\dot{M}_{q0}$ [Hi Δr , Hi $\Delta\phi$]	0.930				0.724		0.172*			
($\alpha = 0.02$)	0.899									
($\alpha = 0.04$)	0.921									
($\alpha = 0.1$)	1.015*									

2.3.2.1 Three-Timescale Regime: $0.25 < q < 1$.

The behaviour of systems with mass ratio in the range $0.25 \lesssim q < 1$ are illustrated in Figures 2.3, 2.4 and 2.6. These show snapshots of the 2D surface density in the quasi-steady-state, the evolution of the azimuthally averaged density profile, and the time-dependent accretion rates, respectively. Additionally, in Table 2.2, and in the corresponding Figure 2.7, we show the time-averaged accretion rate as a function of q . In each case, the accretion rate is averaged over 1000 binary orbits in the quasi-steady state (unless noted, from 3500-4500 binary orbits), and is quoted in units of the corresponding rate for a single-BH ($q = 0$) disc. This ratio changes very little over the course of the quasi-steady-state regime.

These figures and table illustrate several trends as q is lowered from $q = 1 \rightarrow 0.75 \rightarrow 0.5 \rightarrow 0.25$:

1. The cavity becomes more compact, and less lopsided, as one naively expects when the binary torques are reduced. These effects are clearly visible in the middle row of Figure 2.3, and

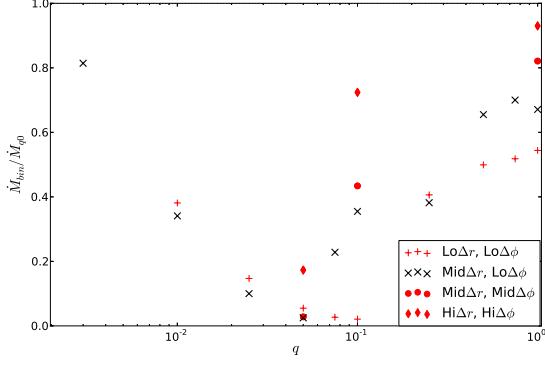


Figure 2.7: The average mass accretion rate as a function of q , and for simulations with different spatial resolution. This is the same information as in Table 2.2, except here shown graphically and with only the simulations at the fiducial magnitude of viscosity.

also in the corresponding azimuthally averaged density profiles in Figure 2.4: the profiles look remarkably similar for $q = 0.5$ and 0.25 , except the cavity for $q = 0.25$ is smaller, and its wall is visibly sharper (as a result of azimuthally averaging over a less lopsided 2D distribution).

2. The secondary (primary) moves closer to (farther from) the cavity wall as q is reduced. Thus occurs for two reasons. First, the position of the secondary (primary) moves away from (towards) the binary's center of mass,

$$r_s(q) = a(1 + q)^{-1} \quad r_p(q) = a(1 + 1/q)^{-1}. \quad (2.13)$$

Second, as mentioned in (i), the size of the central cavity decreases. (Fig. 2.4) shows the locations of the cavity edge r_{ce} expected from balancing the azimuthally averaged gravitational and viscous torques (Fig. 2.5), which agree well with the observed cavity sizes. In Figure 2.8, we explicitly show r_{ce} as a function of q . The points in the figure have been obtained by balancing the azimuthally averaged viscous and gravitational torques measured in

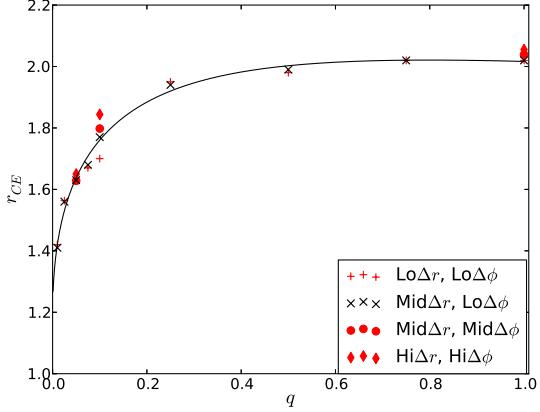


Figure 2.8: The position of the cavity wall as a function of q , in runs with different radial and azimuthal resolutions, as labeled. The points mark the radii r_{ce} at which the azimuthally averaged viscous and gravitational torques balance (equation 2.10). The black line is an empirical fit to the fiducial resolution data points (equation 2.14).

the simulation (equation 2.10). The black line is an empirical fit to the data points at fiducial resolution, given by

$$r_{ce}(q) \simeq A + B \ln(q^{1/2} + 1) + C \ln(q + 1)$$

$$A = 1.191 \quad B = 2.541 \quad C = -1.350 \quad (2.14)$$

3. The dense lump, created by the shocks due to the “regurgitated” stream-material thrown back out by the binary, is still present for $q = 0.5$ (see the corresponding panel in Fig. 2.3, around 9 o’clock at $r \sim 2a$), but is much less discernible for $q = 0.25$. Again, this trend is unsurprising - as the torques diminish, one expects weaker shocks and smaller over-densities in any resulting lump.
4. The accretion streams in the $q = 0.25$ panel of Figure 2.3, and the corresponding ripples in the azimuthally averaged density profiles of Figure 2.4, become noticeably weaker. How-

ever, the average accretion rate, shown in Figures 2.6 and 2.7 and in Table 2.2, stays at approximately $\simeq 0.7\dot{M}_{q0}$ over the range $0.5 \lesssim q \leq 1$. For $q \lesssim 0.5$, the average accretion rate drops more rapidly, falling by nearly a factor of two to $\simeq 0.36\dot{M}_{q0}$ by $q = 0.1$.

5. As the average accretion rate decreases, so does the maximum accretion rate (i.e., the amplitude of the spikes in Fig. 2.6) keeping an approximately constant enhancement factor of ~ 3 as the mass ratio is decreased to $q \sim 0.1$.
6. The percentage of a stream which leaves the domain at $r_{\min} = a$ as opposed to being flung back out also decreases by a factor of ~ 2 as q decreases from 1.0 to 0.1. We measure this percentage from the simulations by computing the ratio $\dot{M}(r_{95})/\dot{M}(r_{\min})$ averaged over 25 orbits in the quasi-steady-state. Here $r_{95} = 0.95r_{ce}$ (with r_{ce} given by equation (2.14)) is chosen to be just inside the cavity wall where the accretion rate is dominated by the streams. The percentage drops by approximately a factor of two, from $\sim 3.3\%$ at $q = 1$ to $\sim 1.8\%$ at $q = 0.1$ suggesting that the drop in average accretion rate in the three-timescale regime is due largely to the amount of stream material which can penetrate beyond the binary torque barrier at small r .
7. Perhaps the most interesting result is shown by the Lomb-Scargle periodograms in Figure 2.6. As q decreases, power is traded from both the $(1/2)t_{\text{bin}}$ and the $5.7t_{\text{bin}}$ variability timescales into the t_{bin} timescale. This is because of the increased proximity between the secondary and the cavity wall, and a corresponding larger distance between the primary and the cavity wall noted above. As a result, as q is decreased, the secondary begins to dominate the variability, pulling accretion streams off of the cavity wall once per binary orbit.

Focusing on the last finding: in the $0.25 \lesssim q < 1$ case, we find that the time-dependent accretion rate displays *three distinct and sharply defined periods*, with well-defined ratios at 0.5, 1, and $5.7t_{\text{bin}}$. While the last of these reflects the orbit of the dense lump at the elongated cavity wall

and could depend on details of the disc properties, the first two periods are fixed by the binary alone and are independent of the disc. *The 1:2 period ratio is therefore a robust prediction; if observed, it could serve as a smoking gun signature of a binary.*

2.3.2.2 Single-Orbital-Timescale Regime: $0.05 \lesssim q \lesssim 0.25$.

As q is further decreased, the overall distortions to the disc become less pronounced, and approach a nearly axisymmetric, tightly wound spiral pattern (see the $q = 0.075$ panel in Figure 2.3). The distance between the cavity wall and the secondary further shortens, and the accretion variability becomes dominated entirely by the streams created by the secondary’s passage once per orbit. As Figure 2.6 shows, for $q = 0.075$ the accretion rate displays a nearly sinusoidal variation, with the corresponding Lomb-Scargle periodogram showing a single spike just offset from the orbital timescale of the binary. For $q = 0.075$, the fluctuations are still large (factor of ~ 3), but by $q = 0.05$, the fluctuations disappear. Interestingly, we find that the average accretion rate is most strongly suppressed among all of our runs in a narrow mass ratio range $q \sim 0.05 \pm 0.0025$; by up to a factor of ~ 40 compared to the single-BH case. However, for our highest resolution run at $q = 0.05$, we no longer observe this extreme dip in average accretion rates (see §2.4.3).

As mentioned above, the disappearance of the $(1/2)t_{\text{bin}}$ variability timescale is easy to understand qualitatively: once the primary BH remains very close to the center-of-mass, its compact orbital motion no longer impacts the disc far away. The disappearance of the $5.7t_{\text{bin}}$ variability timescale is also clearly attributable to the lack of any dense lump near the cavity wall for $q \lesssim 0.25$. However, the reason this lump disappears is less obvious, and warrants some discussion.

1. As q decreases, the cavity becomes less lopsided, and the accretion rate spikes become weaker. This suggests that when these weaker accretion streams are flung back to the cavity wall, they create less over-dense lumps. For $q \leq 0.25$, the lump may not survive shear stresses and pressure forces, and may dissolve in less than an orbital time.

- As can be eyeballed from Figure 2.3, the stream impact zone (dense region outside the cavity) extends over an azimuth of $\Delta\phi \sim 100 - 120^\circ$. The orbital period at the cavity edge is $\leq 6t_{\text{bin}}$ (the orbital period at the furthest edge of the cavity), implying that multiple streams can hit parts of the same lump if streams are generated more than once per binary orbit.

To test whether the strength or the frequency of the streams is more important for lump generation, we repeat our simulation for an equal-mass binary, but we placed one of the BHs artificially at what would be the real binary’s center of mass. The second hole still orbits at $r = a/2$ as usual. In this setup, the cavity wall is perturbed by streams with a similar strength as in the real $q = 1$ simulation (however much less of the streams reach the inner edge of the simulation domain in this one-armed perturber case), but now only once, rather than twice per orbit. We found that, while this “one-armed” binary does generate significant stream impacts at the cavity edge, it does *not* create an orbiting over-density at the cavity edge, nor does it excite a significant elongation of the cavity. We therefore conclude that multiple, overlapping streams are required to generate a strong lump that survives for an orbital time. However, as the $q = 0.25$ accretion rate and periodogram show in Figure 2.6, simply generating two streams is not a sufficient condition for generating a cavity wall lump. Both streams must also be sufficiently large. This explains the disappearance of the cavity wall frequency for $q \lesssim 0.25$; as the mass ratio decreases, the primary generates less significant streams and the overlap of two large streams crashing into the cavity wall can no longer occur to generate an over-dense lump there. This result is also consistent with the cavity becoming less lopsided as q decreases.

2.3.2.3 Steady-Accretion Regime: $q \lesssim 0.05$.

As we continue to decrease q from 0.075 through 0.05 to 0.01, we find yet another distinct regime. The overall morphology of the snapshot of a $q = 0.01$ and $q = 0.05$ (not shown) disc in Figure 2.3 looks similar to the $q = 0.075$ case, except the nearly-concentric perturbations are even weaker,

and the cavity still smaller. However, the similarity is quite deceptive, with the movie versions of these figures showing a striking difference.^e In the $q = 0.075$ case, accretion streams form and disappear periodically, but in the $q = 0.05$ case, the disc pattern becomes constant and unchanging (in the frame co-rotating with the binary). There is still a visible accretion stream, hitting the inner boundary of the simulation just ahead of the secondary's orbit, but the stream steadily co-rotates with the binary.

As Figure 2.6 shows (see also Fig. 2.7 and Table 2.2), the average accretion rate has reached its minimum at $q = 0.05$. For $q \leq 0.05$, the accretion rate becomes steady, with no fluctuations, and its value *increases* back towards the $q = 0$ rate (dashed horizontal [green] line in Figure 2.6). For such a light secondary, the system begins to resemble a disc with a single BH. Although a cavity is still clearly present (moving from $r \simeq 1.6a$ at $q = 0.05$ to $r \simeq 1.4a$ at $q = 0.01$), it is being refilled, as for a single BH. Indeed, after 4000 orbits, the $q = 0.01$ azimuthally averaged density profile is approaching the $q = 0$ profile (see Fig. 2.4).

Although the secondary still excites small but visible ripples in the disc (Fig. 2.3), by $q \leq 0.05$ it can no longer exert a large enough torque to pull in large streams and drive them back out to produce a lopsidedness in the circumbinary disc. The ripples are in the linear regime, and they resemble the tightly-wound spiral density waves launched in protoplanetary discs (e.g. Goldreich & Tremaine 1980; Dong et al. 2011c; ?), except here our background discs have a pre-imposed central cavity. In the linear regime, the waves are excited by resonant interactions with the disc, and non-linear coupling to an $m = 1$ mode is no longer possible. Note the disappearance of any ripples in the azimuthally averaged surface density for $q = 0.01$ (Fig. 2.4).

^eMovie versions of the snapshots in Figure 2.3 are available at <http://www.astro.columbia.edu/~dorazio/moviespage>

2.4 Summary and Discussion

In summary, we find that the behavior of the accretion rate across the circumbinary cavity as a function of q can be categorized into four distinct regimes:

1. *Two-timescale regime*; $q = 1$. Confirming previous results, an equal-mass binary maintains a central low-density cavity of size $r \sim 2a$ and the time-averaged accretion rate is $\sim 2/3$ of that for a point-mass case. There are up to factor of ~ 3 fluctuations around the average on two prominent time-scales, $(1/2)t_{\text{bin}}$ and $\sim 5.7t_{\text{bin}}$.
2. *Three-timescale regime*; $0.25 < q < 1$. The time-averaged accretion rate drops by a factor of ~ 1.8 by $q = 0.25$; however, the maximum fluctuations continue to occur with amplitude ~ 3 times the average rate. There are three time-scales present, $(1/2)t_{\text{bin}}$, t_{bin} , and $\sim 5.7t_{\text{bin}}$.
3. *Single-orbital-timescale regime*; $0.05 \lesssim q < 0.25$. In this regime, the average accretion rate and its fluctuations continue to drop with decreasing q . Variability is dominated by the secondary, and is nearly sinusoidal on the binary period t_{bin} though accompanied by small accretion spikes due to the primary with maxima below the average rate.
4. *Steady-Accretion regime*; $q \lesssim 0.05$. The accretion becomes steady while reaching its lowest rate at $q \sim 0.05$. By $q = 0.003$, the accretion rate rises again to ~ 0.8 of the $q = 0$ case. The system overall resembles a cavity-filling single-BH disc, with small perturbations due to the secondary in the linear regime.

2.4.1 Comparison with MM08

Our main qualitative conclusions for the equal-mass binary case, including the morphology of the disc, and the accretion rate, are in good agreement with MM08. Nevertheless we do find a small discrepancy in the time averaged accretion rates.

Comparing MM08's Figures 7 and 8 with the top left panel of our Figure 2.9, which is run at the same resolution as the highest resolution used in the MM08 study,^f we see that the magnitudes of the time-averaged accretion rates, as well as the periodogram frequencies, agree. However, the detailed variability is not identical.

The primary difference is that there is more power at low frequencies in the MM08 periodograms. In-between the largest accretion spikes, the MM08 accretion rate drops closer to zero and becomes less uniform where as our accretion rate is a steady modulation of spikes occurring at the twice the binary orbital period. The accretion rate and periodogram observed in MM08 could be realized if viscous stresses are less efficient at breaking up the over-dense lump responsible for creating the $(5 - 6)t_{\text{bin}}$ modulation. Then the lump will be more centralized and there will be a greater disparity between streams which are generated from the lump and streams which are not. Thus we conjecture that this small difference may be due to different treatments of viscosity and grid setup, or differences between FLASH2 (used by MM08) and FLASH3 (used here).

2.4.2 Viscosity Study

To evaluate the sensitivity of our results to the magnitude of viscosity, we run three additional equal-mass binary simulations for $\alpha = 0.02, 0.04$, and 0.1 . Figure 2.10 plots snapshots of the 2D surface densities and Figure 2.11 plots the accretion rates with corresponding periodograms for each of these runs once they have reached the quasi-steady-state regime. Table 2.2 records the average accretion rates as a fraction of the reference $q = 0$ simulations (a new reference simulation is created for each α). Our findings can be summarized as follows:

1. As α is increased, the near side of cavity spreads closer to the binary, as a result, larger portions of streams leave the simulation domain at $r = a$. This results in, not only a larger

^fThis refers to the highest resolution used by MM08 at the inner region of the disc. MM08 use a lower resolution far away from the binary where this study uses a uniform resolution throughout.

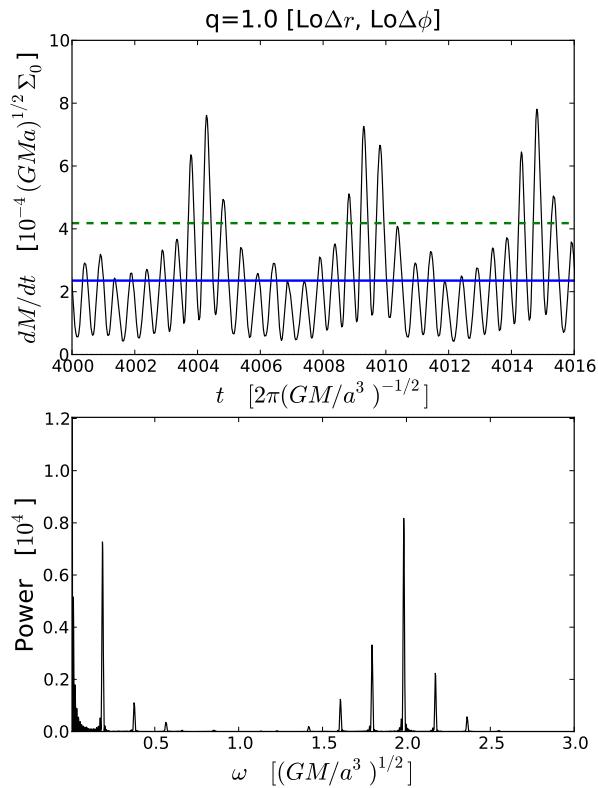


Figure 2.9: Accretion rates at the inner boundary $r_{\min} = a$ for the equal-mass binary as in the top left panels of Figure 2.6, except for the lowest resolution runs used in this study which matches the highest resolution used in the disc simulated by MM08. The solid horizontal [blue] line in the top panel is the average accretion rate, and the bottom panel shows the Lomb-Scargle periodogram computed over 100 binary orbits.

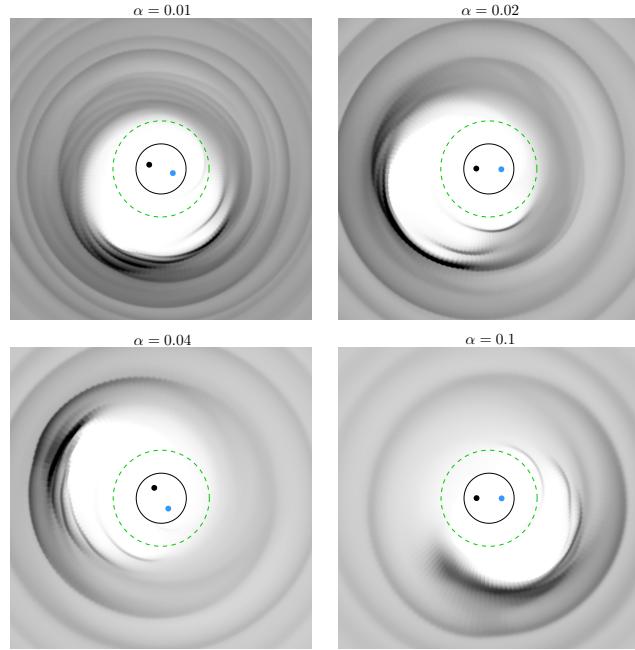


Figure 2.10: Two-dimensional surface density distributions during the quasi steady-state phase, as in Figure 2.3, except for the single mass ratio $q = 1.0$, and for four different values of the viscosity parameter α , as labeled. Increasing α causes ripples created by streams impacting the cavity wall to smear out more quickly causing the surface density snapshots to appear smoother. For all values of α shown here, the over dense lump still survives long enough to create the $\sim (5 - 6)t_{\text{bin}}$ modulation of the accretion rate. Also for larger α , the near side of the disc extends in closer to the binary causing a larger fraction of streams to exit the integration domain at $r = a$. This results in higher measured accretion rates relative to the point mass values for the same α 's.

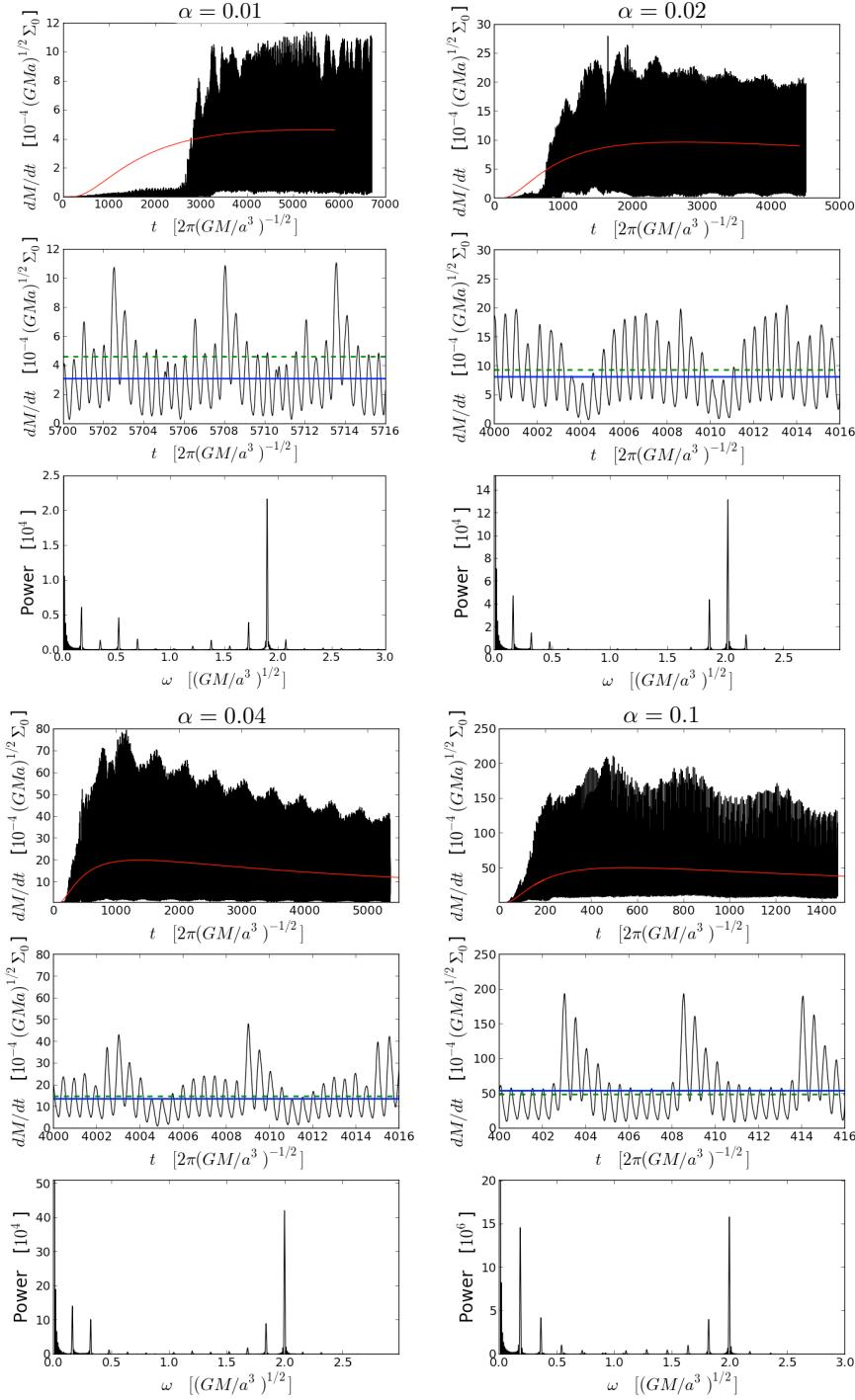


Figure 2.11: The time-dependent accretion rates as in Figure 2.6, except for the single mass ratio $q = 1.0$ and four different values of the viscosity parameter α , as shown in Figure 2.10. Additionally the top panel of each figure shows the entire accretion rate history for the $q = 1$ (black) and $q = 0$ (red) cases. Notice that long-term (once per $\sim 400t_{\text{bin}}$) variability appears in the larger α runs and is coincident with the period at which the elongated cavity precesses..

absolute accretion rate, but also an increased rate measured relative to the $q = 0$ rate with the same α (See Table 2.2).

2. Despite the increase in average accretion rate, the ratio of maximum accretion rate spikes to the average accretion rate stays constant at ~ 3 .
3. A transition from a symmetric state to an elongated quasi-steady-state still occurs resulting in discs with similar morphology including cavities which are of the same size and elongation. As can be seen from Figure 2.10, a difference is that density ripples observed in the fiducial $\alpha = 0.01$ disc are progressively more smeared out for the higher α runs. This is due to larger viscous shearing forces more quickly diffusing over-densities due to stream impacts.
4. Although there are larger shearing forces smearing out the small scale density structures seen in the fiducial case, the over-dense lump still survives for at least the necessary $\sim (1/2)t_{\text{bin}}$ needed to modulate the accretion rate at the cavity wall orbital period.
5. As a result of the above two points, the periodograms in Figure 2.11 stay largely the same as α increases. If there is a trend, it is that the timescale associated with the orbital period of the over-dense lump at the cavity wall is more prominent for larger α . However, it is possible that, for even larger α , the over-dense lump could break up before it can seed another lump via stream generation.
6. For the larger α runs, the top panels of Figure 2.11 show the appearance of a longer variability timescale with the same period as the lopsided cavity precession - once per ~ 400 orbits. This variability manifests itself in a modulation of the maximum accretion rate achieved by the largest streams pulled from the cavity edge lump; every ~ 400 orbits the largest accretion rate spikes reach 30% higher above the average than they do ~ 200 orbits later. Note that in the fiducial α case, there is a similar long-term variation in the strength of the $5.7t_{\text{bin}}$

modulation, but it occurs more erratically and with approximately half of the total variation.

7. Finally, the top panels of Figure 2.11 also show that the quasi-steady, lop-sided mode occurs much earlier for larger α . For the fiducial case the transition takes place after $\sim 2500t_{\text{bin}}$, for $\alpha = 0.02$ after $\sim 1000t_{\text{bin}}$ and for larger $\alpha = 0.04, 0.1$ after less than a few 100 orbits which is set largely by the time for fluid to diffuse to the inner regions of the disc.

A more detailed investigation of the effects of viscosity should be carried out in future studies and understood self-consistently from simulations which generate turbulent viscosity via the magnetorotational instability (MRI) (See ?Noble et al. 2012).

2.4.3 Resolution Study

Up to now we have discussed the results from our fiducial set of medium-radial, low-azimuthal resolution runs ($[\text{Mid}\Delta r, \text{Lo}\Delta\phi]$ in Table 2.1). Ideally, we would repeat these runs at increasingly high radial and azimuthal resolutions, until the results converge. Unfortunately, this is computationally prohibitive, and we instead choose the following approach.

1. For the $q = 1.0$ case, we perform two higher-resolution runs ($[\text{Mid}\Delta r, \text{Mid}\Delta\phi]$ and $[\text{Hi}\Delta r, \text{Hi}\Delta\phi]$ in Table 2.1) and one lower resolution run ($[\text{Lo}\Delta r, \text{Lo}\Delta\phi]$ in Table 2.1) to look for signs of convergence.
2. We then explore the resolution sensitivity of the boundaries between each accretion regime: The $q = 0.1$ case is at the cusp of the three-timescale and single-timescale regimes, where we expect the accretion behaviour to be particularly sensitive to q . The $q = 0.05$ case is at the cusp of the single-timescale and steady accretion regimes, where the minimum accretion rates are achieved. Thus we also run the $q = 0.1$ and $q = 0.05$ cases for the same set of resolutions as the $q = 1$ case.

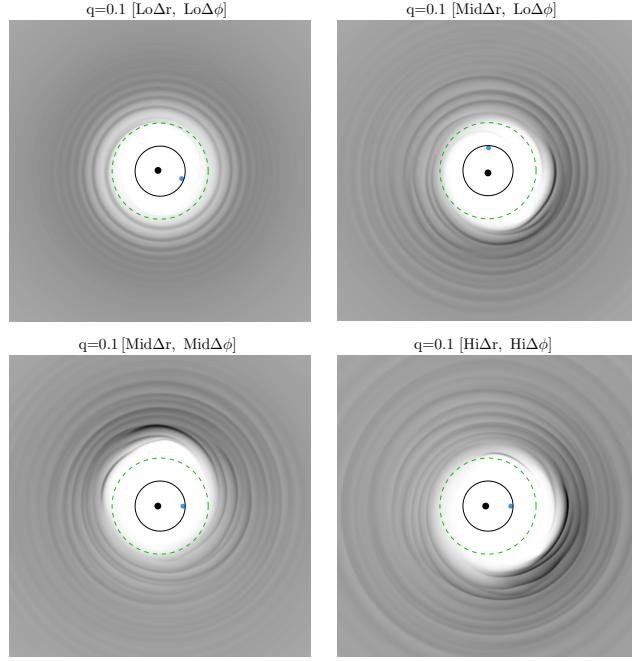


Figure 2.12: Two-dimensional surface density distributions during the quasi steady-state phase, as in Figure 2.3, except for the single mass ratio $q = 0.1$, and for four different combinations of high/low radial and azimuthal resolutions, as labeled. Increasing the spatial resolution decreases numerical diffusion, leads to sharper features, and allows stronger accretion streams. The stronger streams lead to more over-dense lumps where the regurgitated streams hit the cavity wall. As a result, the cavity becomes larger and more lopsided as the resolution is increased. In the lowest resolution case, the cavity never becomes lopsided.

3. We repeated each of our runs at the lowest resolution. Redoing the entire set of runs allows us to assess whether different mass ratios are affected by resolution differently.

Figure 2.12 gives a visual impression of the surface density distribution at the four different combinations of resolutions for $q = 0.1$. They show a clear trend: as the resolution is increased, the lumps near the cavity wall become sharper and more over-dense, and the cavity becomes larger and more lopsided.^g This trend can be attributed to numerical dissipation which is most prominent at shocks, *i.e.* where regurgitated streams impact the cavity wall. Increasing the resolution implies

^gFigure 2.8 shows that the *average* position of the cavity wall, found from the azimuthally averaged torques in equation (2.10), is much less affected over the range of resolutions studied here.

weaker numerical diffusion, stronger accretion streams, more momentum carried by these streams into the disc, and an overall more efficient driving of the $m = 1$ mode. This is further evidenced by an earlier onset of the elongated mode for the higher resolution $q = 0.1$ simulations. The $q = 0.1$ simulations develop an elongated cavity after ~ 1500 (highest resolution), ~ 2500 (medium resolution), and ~ 3500 (fiducial resolution) binary orbits. The lowest resolution run never develops an elongated cavity even after ~ 7000 binary orbits.

The 2D surface density profiles for the $q = 1.0$ runs at different resolutions remain qualitatively the same as the fiducial resolution counterpart. We do find that as the azimuthal resolution is increased the cavity becomes slightly more elongated likely due again to more efficient stream impacts.

The 2D surface density profiles for the $q = 0.05$ runs at different resolutions remain qualitatively the same for all resolutions except the highest resolution. For the highest resolution $q = 0.05$ run, the disc transitions in to the single timescale regime only after ~ 4000 binary orbits and resembles the $q = 0.075$ disc at the fiducial resolution.

Figure 2.13 shows the corresponding time-dependent accretion rates at the inner boundary of the $q = 0.1$ simulations at different resolutions. The accretion patterns look visibly different for the lowest resolution run which, for reasons stated above never excites the lopsided cavity mode. However, encouragingly, the difference between our fiducial [Mid Δr ,Lo $\Delta\phi$] and the highest-resolution [Hi Δr , Hi $\Delta\phi$] cases is modest with a primary trend of increasing accretion rate relative to the $q = 0$ rate, with increasing resolution (see Fig. 2.7 and Table 2.2).

The accretion rates for the $q = 1.0$ runs at different resolutions remain qualitatively the same as their fiducial resolution counterparts. One difference is that the higher and lower (Figure 2.9) resolution $q = 1.0$ runs have more power at the cavity wall periodicity than the fiducial run; the highest resolution run having more power at the cavity wall frequency than at the $2t_{\text{bin}}$ frequency. Also, the cavity wall period becomes slightly longer as the resolution increases, from $5.3t_{\text{bin}}$ to $5.7t_{\text{bin}}$

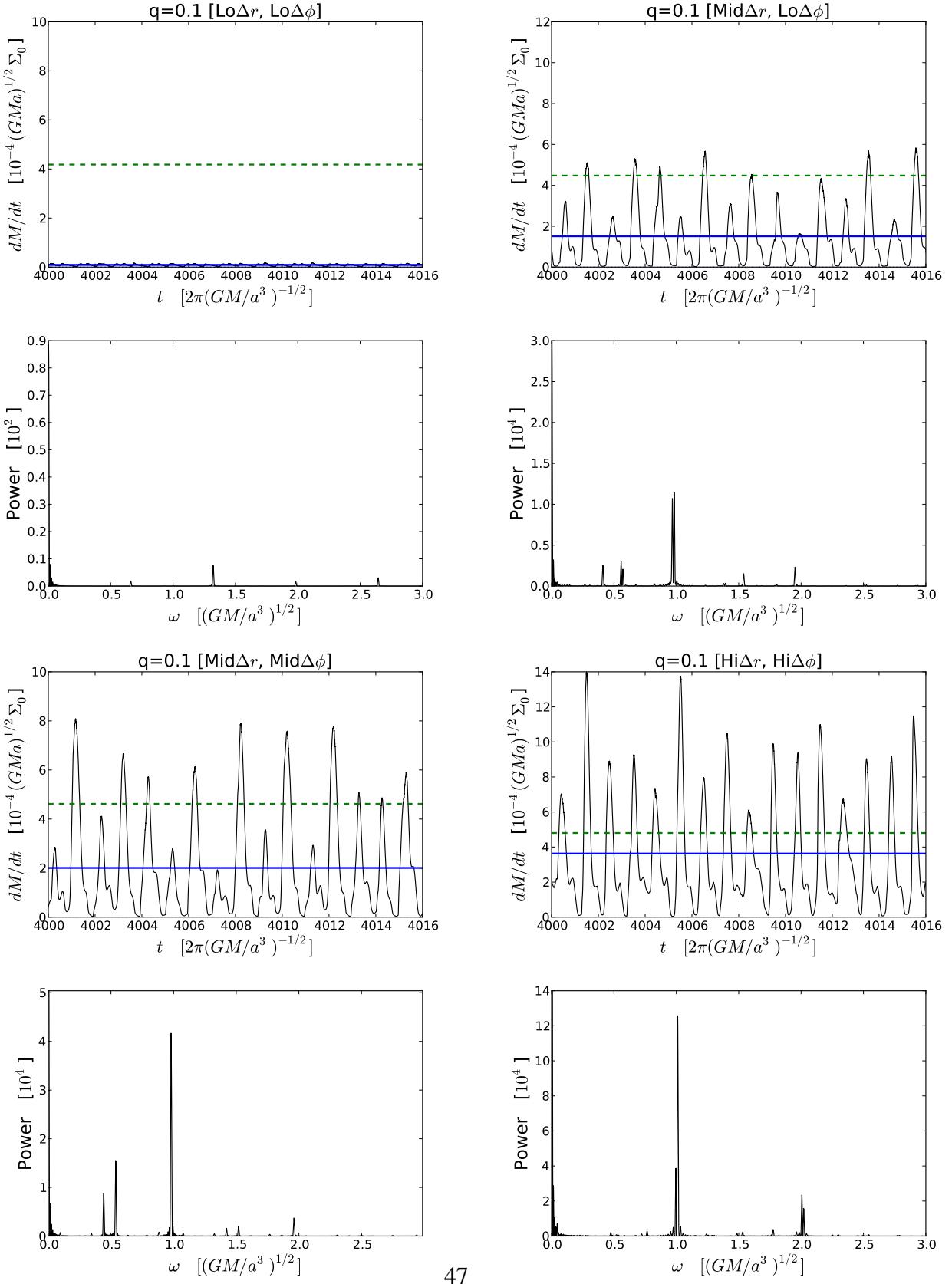


Figure 2.13: The time-dependent accretion rates as in Figure 2.6, except for the single mass ratio $q = 0.1$ and the four different resolutions also shown in Figure 2.12. The $q = 0.1$ binary is at the

to $5.9t_{\text{bin}}$, to $6.4t_{\text{bin}}$ for the [Lo Δr ,Lo $\Delta\phi$], [Mid Δr ,Lo $\Delta\phi$], [Mid Δr ,Mid $\Delta\phi$], and [Hi Δr ,Hi $\Delta\phi$] runs respectively. A larger cavity again indicates that higher resolution allows more efficient elongation of the cavity. This is further evidenced by an earlier onset of the elongated mode for the higher resolution $q = 1$ simulations. The two highest resolution $q = 1$ simulations develop an elongated cavity at ~ 1500 binary orbits where as the two lower resolution runs develop the elongated cavity only after ~ 2500 binary orbits.

The accretion rates for the $q = 0.05$ runs at different resolutions again remain nearly identical for all resolutions except the highest resolution. For the highest resolution $q = 0.05$ run, the disc transitions in to the single, orbital-timescale regime after ~ 4000 binary orbits and exhibits modulation of the accretion rate at the orbital frequency, mimicking the $q = 0.075$ accretion rates at the fiducial resolution.

Table 2.2 and the corresponding Figure 2.7 shows the time-averaged accretion rates as a function of q at different resolutions. In all cases, at the same fixed q , we find that increasing resolution produces a higher accretion rate. This is consistent with the interpretation above that higher resolution allows stronger accretion streams. Interestingly, we find a strong correlation between the values of $\dot{M}_{\text{bin}}/\dot{M}_{q0}$ listed in Table 2.2 and the accretion patterns seen in Figure 2.13: runs at different resolutions but with similar values of $\dot{M}_{\text{bin}}/\dot{M}_{q0}$ have very similar accretion patterns (including the variability and the values of the maxima). The result of increasing [decreasing] the resolution can therefore be interpreted as a shift of the accretion behavior to lower [higher] mass ratios.

Comparing the full set of mass ratio runs for the lowest resolution to the fiducial resolution runs, we observe the same progression of the accretion rate through each of the accretion variability regimes discussed above; a difference being that, as discussed in the previous paragraph, the boundaries between each regime are delineated at larger mass ratios in the lowest resolution runs. We also notice that in the three-timescale regime, there is more power in the periodogram peak associated with the cavity wall orbital period (*e.g.* compare the top left of Figure 2.6 with

Figure 2.9). The cavity wall peak still disappears for $q = 0.25$ when the stream due to the primary becomes much smaller than the stream due to the secondary and the overlapping of large streams at the cavity wall no longer generates an over-dense lump.

Encouragingly, the two higher resolution runs at $q = 1.0$, also plotted in Figure 2.7, lie closer to each other than the two lower resolution runs. Since the resolution steps are evenly spaced, we consider this evidence that the simulations are converging monotonically with resolution. Since the $q = 0.1$ and $q = 0.05$ discs are positioned at the boundary of different accretion regimes we find a large dependence of disc response and accretion rate on resolution, but with a clear trend of increasing resolution moving the boundaries between the accretion regimes described in this study to slightly lower values of the binary mass ratio.

2.4.4 Physical Regime: Black Hole Binary Parameters

The simulations presented above can be scaled, in principle, to any black hole mass and orbital separation. In this section, we discuss the physical scales for which our simulations could be relevant (i.e. physically viable and observationally interesting). The shaded region in Figure 2.14 plots this relevant portion of parameter space by imposing the following restrictions.

1. $10^5 M_\odot \lesssim M_p \lesssim 10^9 M_\odot$. It is not clear whether smaller BHs exist in galactic nuclei, and, in any case, the radiation from such a low-mass BHs would likely be too faint to detect. Likewise, much more massive BHs are known to be rare.
2. $10^{-2} \lesssim q \leq 1$. As we have shown, for the set-up we study (i.e. with a cavity inserted by hand into the disc), the accretion pattern converges as we decrease the mass ratio to $q = 0.01$ and below. In practice, a physical lower limit of $q \sim 0.01$ may arise from the fact that bound binary BHs can be created only in relatively major mergers. In a minor merger, the smaller satellite galaxy may be tidally stripped during the early stages of the merger, decreasing the

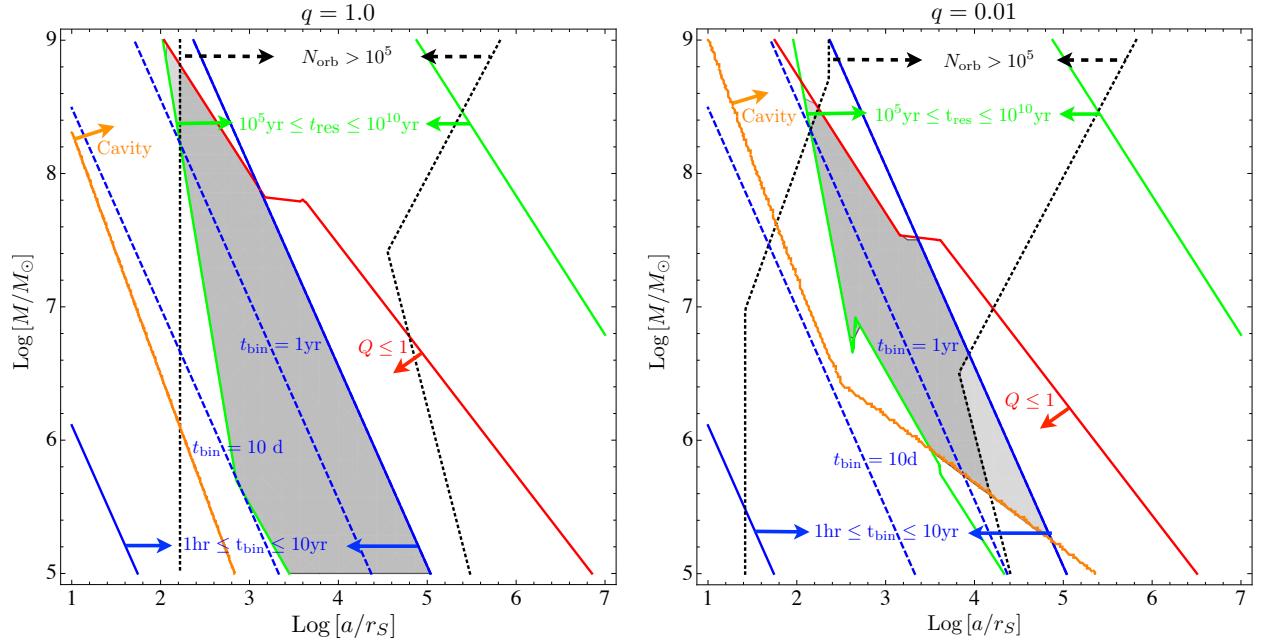


Figure 2.14: The shaded regions in each panel denote the values of binary total mass M and separation a for which a binary + disc system would be physically viable and observationally interesting as well as meet simulation specific constraints. Blue lines denote contours of binary orbital time (a characteristic variability timescale). Green lines denote contours of binary residence time $t_{\text{res}} \equiv -a(da/dt)^{-1}$ computed from Haiman et al. 2009 for migration of the secondary through a gaseous disc as well as gravitational radiation. Dashed black lines denote the simulation specific constraint that the binary separation not change appreciably over the course of a simulation time. Red lines denote the boundary between gravitationally stable and unstable disc regions via the Toomre Q parameter. Orange lines denote the boundary between binaries which can maintain a cavity and those which will not (computed via the steady-state solutions given by Kocsis et al. 2012a). The left panel is for an equal-mass ratio binary and the right panel is for a binary with mass ratio $q = 0.01$. For both plots we use $\alpha = 0.1$. Note that for all mass ratios, the most massive binaries do not fit into a gravitationally stable disc. However, this is determined for an undisturbed α -disc surrounding the primary; perturbations due to a large secondary would increase the stability of the disc out to larger radii (Haiman et al. 2009). In the $q = 0.01$ case, less massive, close binaries do not maintain cavities and do not represent systems which are consistent with the initial conditions adopted in this study.

efficiency of dynamical friction, and aborting subsequent binary formation (Callegari et al. 2009). Coupled with the well-established correlations between the mass of a MBH and its host galaxy, this suggests that the q -distribution may not extend to values significantly below $q \sim 0.01$. ^h Figure 2.14 plots the restriction of a, M parameter space for these limiting mass ratios $q = 1$ (left) and $q = 0.01$ (right).

3. *The binary is embedded in a thin gaseous disc.* Following a merger of two MBH-harboring galaxies, the MBHs sink to the bottom of the new galactic potential via dynamical friction in approximately a galactic dynamical timescale (Begelman et al. 1980). In addition to stellar interactions (e.g. Preto et al. 2011), many studies have shown that gas in the vicinity of the binary could aid in hardening the binary down to \ll pc separations (e.g. Escala et al. 2005; Dotti et al. 2007; Mayer et al. 2007; Lodato et al. 2009; Cuadra et al. 2009; Nixon et al. 2011; Chapon et al. 2013). We have assumed that such gas in the vicinity of the binary cools efficiently and forms a rotationally supported thin disc. For a given M_p and M_s , we are still free to choose a physical distance for the orbital radius a , which could correspond to a snapshot of the binary anywhere along its orbital decay. The assumption that the binary is embedded in a thin disc allows us to make the following additional constraints on a given M :

- (a) *The accretion disc is gravitationally stable.* Accretion discs become self-gravitating, and unstable to fragmentation, beyond a radius of order $\gtrsim 10^4(M/10^7M_\odot)^{-1}r_S$ (where r_S is the Schwarzschild radius; see, e.g., Goodman 2003; Haiman et al. 2009 for the formulae used to generate the $Q \leq 1$ criteria in Figure 2.14). Since the binary has to fit inside a gravitationally stable disc, this puts an upper limit on the orbital separation denoted by the red lines in Figure 2.14.

^hIn principle, less massive BHs may grow from the accretion discs around the primary (McKernan et al. 2012); the long-term evolution of such systems would be worthy of further study.

- (b) *Variability occurs on an observable timescale.* The binary orbital time is given by

$$t_{\text{bin}} = \frac{2\pi}{\Omega} = 0.88 \left(\frac{M}{10^7 M_\odot} \right) \left(\frac{a}{10^3 r_S} \right)^{3/2} \text{yr.} \quad (2.15)$$

As we have shown, the accretion rate shows periodicity on a timescale of $\sim t_{\text{bin}}$. In a realistic survey, it will be feasible to look for periodic variations between $0.1 \text{ hr} \lesssim t_{\text{bin}} \lesssim \text{few yr}$ denoted by the solid blue lines in Figure 2.14. Here the lower limit comes from the integration time required to measure the flux variations for MBHBs in the above mass range (for a survey instrument with a sensitivity similar to LSST; [Haiman et al. 2009](#)), and the upper limit comes from the duration of proposed time-domain surveys. As a guide, the dashed blue lines in Figure 2.14 are contours of constant orbital times drawn at 10 days and 1 year.

- (c) *The binary spends a long time at a given separation.* Assuming that the binary is embedded in a thin disc, [Haiman et al. 2009](#) compute residence times, $t_{\text{res}} \equiv -a(da/dt)^{-1}$, as a function of the binary separation a , due to migration of the secondary through the disc and due to gravitational wave decay at small enough binary separations. In Figure 2.14 the green lines denote the requirement that $10^5 \leq t_{\text{res}} \leq 10^{10}$ years. A residence time of greater than 10^{10} years does not on its own exclude a binary system from observation. Nevertheless, we include this limit in order to show which binaries will not merge (due to migration through a gaseous disc) in a Hubble time. Note also that there is a trade-off: a longer residence time is desirable since it increases the probability of finding such a system; however, longer residence times occur at larger separations and longer orbital times, which will make it more difficult to verify any periodic behavior.
- (d) *A cavity is maintained.* For consistency with the initial conditions adopted here, we require that the binary + disc systems will indeed form a cavity during earlier stages of

their evolution. The region of parameter space for which a cavity may be maintained is denoted by the orange lines in Figure 2.14 and is calculated using the *steady-state* disc solutions detailed in Kocsis et al. (2012a).

- (e) *The orbital separation is fixed.* Throughout our simulations, we fix the binary separation; we therefore require that the orbital decay should be slow enough for the binary’s orbit not to change significantly over a few thousand orbits. This is denoted by the dashed black lines in Figure 2.14 which are drawn where $N_{\text{orb}} = t_{\text{res}}/t_{\text{bin}} = 10^5$.
4. Though not expressed in Figure 2.14, for our simulations to be self-consistent, we also require $a \gtrsim 100r_S$, since our Newtonian treatment ignores general relativity. Furthermore, at approximately the same binary separation, the orbital decay of the binary due to gravitational wave emission becomes more rapid than the viscous time at the edge of the cavity. As a result, the disc decouples from the binary and is ‘left behind’, rendering our initial conditions inconsistent in this regime (e.g. Milosavljević & Phinney 2005; although see Farris et al. 2012 and Noble et al. 2012 whose MHD simulations suggest that the gas can follow the binary down to smaller separations).

2.4.5 Caveats

However instructive, the simulations presented here are still of course simplified models of a real binary disc system, and it is worth listing some major caveats.

1. Our simulations are two dimensional – we expect that the 3D vertical structure could modify the structure of the accretion streams, including their q -dependence. However, note that ? find similar features in the accretion rate periodograms measured from 3D simulations.
2. Our discs are assumed to have angular momentum co-aligned with that of the binary (prograde discs). In principle, a random accretion event onto the binary could result in a mis-

aligned disc which will eventually be torqued into co- or counter-alignment with the binary angular momentum ([Nixon et al. 2011](#)).

3. Our simple α -viscosity prescription may be inaccurate, especially in the nearly radial accretion streams. Recent MHD simulations in the $q = 1$ case find a larger effective α than the fiducial value adopted here ([?Noble et al. 2012](#)). It is interesting, however, that our highest $\alpha = 0.1$ simulation for an equal mass binary exhibits the same variability as our fiducial case but a larger accretion rate, in agreement with the above mentioned 3D MHD simulations.
4. We assumed a locally isothermal equation of state; more realistic equations of state could have an especially large impact on the strength and dissipation of shocks at the cavity wall.
5. Our initial conditions correspond to an unperturbed, near-Keplerian, circular disc, and circular binary orbit, with a significant pile-up of gas. In reality, accretion onto the binary could produce significant binary (as well as disc) eccentricity (*e.g.* [Cuadra et al. 2009](#); [Lodato et al. 2009](#); [?](#)). In this case, the variability we find would most likely have a more complex structure (*e.g.* [Hayasaki et al. 2007](#)) due to the plethora of resonances available in an eccentric binary potential (*e.g.* [Artymowicz & Lubow 1994](#)). In a study of circumbinary discs around eccentric binaries (with mass ratio 1/3), [?](#) find similar accretion rate periodograms as in this study; periodogram peaks exist at the orbital frequency, twice the orbital frequency and the cavity wall orbital frequency. They find that an increasingly eccentric binary: (a) increases the size of the cavity and thus decreases the overall magnitude of the accretion rate, (b) enhances power at harmonics of the orbital frequency, and (c) increases power in a peak located at the beat frequency of the orbital frequency and the cavity wall frequency.
6. As previously stressed, these simulations do not allow the binary orbit to evolve in response to forces exerted by the gas disc. When this assumption of a massless disc is lifted, in addition to changes in binary eccentricity and semi-major axis, the binary center of mass

could oscillate around the disc center of mass due to the orbiting eccentric disc. For massive discs, the above effects could alter the description of disc evolution, and hence accretion, presented in this zero disc-mass study.

7. Although we begin with an “empty” cavity, this cavity may overflow already at a large radius ([Kocsis et al. 2012a](#)). Future studies should construct a self-consistent initial density profile, by evolving the binary’s orbit from large radius through gap clearing, and thus determining whether a true pile-up occurs.
8. We have ignored the radiation from the gas accreted onto the BHs. Given that we find high accretion rates – comparable to those for a single BH – the secondary BH can be fed at super-Eddington rates, and the flow dynamics can then be strongly affected by the radiation.
9. We have not allowed accretion onto the BHs, and have excised the inner region $r < a$ from the simulation domain. This could have an impact on the dynamics of the streams that are flung back towards the cavity wall, and therefore on the formation of the dense lump, the lopsidedness of the cavity, and the variable accretion patterns.

These caveats should all be pursued in future work, to assess the robustness of our results. We expect that our main conclusions, namely that the accretion rate is strongly modulated by the binary, and that the power-spectrum of the accretion shows distinct periods, corresponding to the orbital periods of the binary and the gas near the location of the cavity wall, will be robust to all of the above caveats. However, the numerical values, such as the mean accretion rate, and the critical value of q for the transition between variable and steady accretion, will likely be affected.

2.5 Conclusions

We have investigated the response of an accretion disc to an enclosed binary via two-dimensional, Newtonian, hydrodynamical simulations. As previous work has shown ([Artymowicz & Lubow 1994](#); [Hayasaki et al. 2007](#); MM08; [Cuadra et al. 2009](#); [?](#); [?](#)), for non-extreme mass ratios, the binary carves out a cavity in the disc, but gas still penetrates the cavity in streams which possibly accrete onto the binary components. Here we have followed up on the work of MM08 by investigating the nature of this inflow across the circumbinary cavity, as a function of binary mass ratio q . We have simulated 10 different mass ratios in the range $0.003 \leq q \leq 1$. This corresponds to the expected range of q -values for massive BH binaries produced in galaxy-galaxy mergers.

We find that while the binary ‘propellers’ are effective at maintaining a low-density cavity at the center of the disc, they can not efficiently suppress accretion across the cavity. For $q = 1$, the average accretion rate is on order of 2/3 that of a single BH with accretion spikes of ~ 3 times larger. As long as the circumbinary disc is fueled at a near-Eddington rate from large radius, these binaries could therefore have quasar-like luminosities. This should facilitate finding counterparts to GW events ([Kocsis et al. 2006](#)), and should also allow their detection in electromagnetic surveys ([Haiman et al. 2009](#)).

We have found that the accretion is not only strong, but can be strongly variable (by a factor of ~ 3), with a characteristic q -dependent frequency pattern. While the accretion for $q < 0.05$ is steady, for $q \gtrsim 0.05$ there is a strong modulation by the binary, and a clear dependence on q of both the variability pattern, and the magnitude of the time-averaged accretion rate. For an equal-mass binary, the accretion rate is modulated at twice the orbital frequency and $\sim 1/6$ the orbital frequency. As the mass ratio is lowered, the power in the $1/2t_{\text{bin}}$ and $(5 - 6)t_{\text{bin}}$ variability timescales is reduced, and traded for a third variability timescale at t_{bin} . In the range $0.05 \lesssim q \lesssim 0.25$, the single t_{bin} timescale is dominant.

Increasing the magnitude of viscous forces has little effect on the above findings except to increase the magnitude of the accretion rate (both absolute and relative to $q = 0$) and to bring out a long-term accretion variability timescale with a periodicity of $400t_{\text{bin}}$. However, accretion discs with even larger viscous forces could quench the $(5 - 6)t_{\text{bin}}$ variability timescale if the over-dense lump responsible for its generation can be broken up before it repeats \sim an orbit at the cavity wall. Hence, further investigation into the effects of viscosity are warranted.

Strong and highly variable accretion, with characteristic frequencies, should aid in identifying massive BH binaries in galactic nuclei. The presence of two frequencies, in the ratio 1:2 for unequal-mass binaries ($0.1 \leq q < 1$), is an especially robust prediction that is *independent of disc properties*, and could serve as a ‘smoking gun’ evidence for the presence of a binary. Our results suggest that the ratio of the power at these two frequencies could probe the mass ratio q , while other features of the periodogram could probe properties of the disc, such as its viscosity.

The variability time-scales are on the order of the orbital period, and we have argued that the most promising candidates in a blind electromagnetic search would be those with total mass and separation contained in the shaded regions of Figure 2.14; $10^{6-7}M_{\odot}$ binaries, preferably with orbital periods of days to weeks. The time-variable accretion to the central regions could produce corresponding variability in broad-band luminosities, allowing a search in a large time-domain survey, such as LSST, without spectroscopy. Additionally, the emission lines could exhibit periodic shifts in both amplitude and frequency; kinematic effects from the binary’s orbit could be distinguished from those due to the fluctuating accretion rate, whenever the latter contains multiples of the binary period.

A few percent of the accretion streams generated periodically fuel the BHs, but the majority of the stream material is flung back and hits the accretion disc farther out. The shocks produced at these impact sites are prominent for $q \gtrsim 0.1$, and can provide additional observable signatures. In particular, radiation from these shocks should be temporally correlated with the luminosity

modulations arising near the secondary and/or primary BH, with a delay time on the order of a binary orbit.

GW observatories, such as eLISA, and Pulsar Timing Arrays will be able to constrain the mass ratios of in-spiraling MBHB’s at the centers of galactic nuclei ab-initio, providing a template for the expected variability pattern. This should be helpful in identifying the unique EM counterpart among the many candidates (with luminosity variations) in the eLISA/PTA error box, as the source with a matching period.

In summary, our results imply that massive BH binaries can be both bright and exhibit strongly luminosity variations, at the factor of several level. This raises the hopes that they can be identified in a future, suitably designed electromagnetic survey, based on their periodic variability. Although encouraging, these conclusions are drawn from simplified 2D hydrodynamical models of a real binary disc system, and should be confirmed in future work.

Acknowledgments

We thank Brian Farris and Bence Kocsis for useful discussions. We thank Paul Duffell for useful discussions, for verifying some of our results in independent runs with the code DISCO, and also for suggesting the Cartesian shear flow test of our viscosity implementation. We also thank the anonymous referee for constructive comments which have improved this manuscript. We acknowledge support from NASA grant NNX11AE05G (to ZH and AM). DJD acknowledges support by a National Science Foundation Graduate Research Fellowship under Grant No. DGE1144155.

2.6 Viscous implementation in polar coordinates

2.6.1 Viscosity and the Momentum Equation

We may write the momentum equation in component form and with respect to an arbitrary basis as

$$\partial_t(\rho v_i) = -\nabla_j \Pi_i^j \quad (2.16)$$

where ∇ represents the covariant derivative,

$$\Pi_{ij} = Pg_{ij} + \rho v_i v_j - \sigma_{ij} \quad (2.17)$$

are the components of the momentum flux density vector, P is the mechanical pressure, and g is the metric tensor. The first two terms represent a reversible momentum flux due to pressure forces and mechanical transport of the fluid. The last term expresses a non-reversible momentum flux due to viscous forces via the *viscous stress tensor* σ . Writing out (2.16) with (2.17),

$$\partial_t(\rho v_i) = \nabla_j(Pg_i^j) + \nabla_j(\rho v_i v^j - \sigma_i^j)$$

or in vector notation

$$\partial_t(\rho \mathbf{v}) = -\nabla P - \nabla \cdot (\rho \mathbf{v} \mathbf{v} - \boldsymbol{\sigma}) \quad (2.18)$$

Thus we see that the effects of viscosity can be incorporated by computing viscous momentum fluxes from $\boldsymbol{\sigma}$ and subtracting them from the mechanical transport term $\rho \mathbf{v} \mathbf{v}$. This is what FLASH does currently to incorporate the effects of viscosity in Cartesian coordinates. However, in non-Cartesian coordinates, there will also be geometric source terms from taking the divergence of the rank-two-tensors $\mathbf{v} \mathbf{v}$ and $\boldsymbol{\sigma}$. Thus we must compute not only the components of these terms, but

also the divergence in order to identify geometric source terms.

2.6.2 The Form of the Viscous Stress Tensor

The viscous stress tensor has components ([Landau & Lifschitz 1959](#))

$$\sigma_{ij} = \rho \left[\nu (\nabla_i v_j + \nabla_j v_i) + \left(\zeta - \frac{2}{D} \nu \right) g_{ij} \nabla_l v^l \right] \quad (2.19)$$

where ν is the kinematic coefficient of viscosity, ζ is the bulk coefficient of viscosity, and D is the number of spatial dimensions. The $2/D$ factor is chosen so that only the bulk viscosity term survives upon taking the trace of σ .

2.6.3 Components of σ in Polar Coordinates

To compute the viscous stress tensor components we work in a coordinate basis to evaluate the covariant derivatives in terms of Christoffel symbols,

$$\nabla_j T_i = \partial_j T_i - \Gamma_{ij}^k T_k$$

Working in 2D polar coordinates, (r, ϕ) , the non-zero Christoffel symbols are

$$\Gamma_{\phi\phi}^r = -r \quad \Gamma_{r\phi}^\phi = \Gamma_{\phi r}^\phi = \frac{1}{r}$$

In the polar coordinate basis Eqs. (2.19) become

$$\begin{aligned}\sigma_{rr} &= \Sigma [2\nu\partial_r v_r + (\zeta - 1)\nabla \cdot v] \\ \sigma_{r\phi} &= \Sigma \left[\nu \left(\partial_r v_\phi + \partial_\phi v_r - \frac{2}{r} v_\phi \right) \right] \\ \sigma_{\phi\phi} &= \Sigma [2\nu(\partial_\phi v_\phi + rv_r) + (\zeta - 1)r^2\nabla \cdot v]\end{aligned}$$

where Σ is the height integrated 2D surface density, $v^\phi = \Omega$ and $v_\phi = r^2\Omega$, Ω being the angular frequency. Transforming to an orthonormal basis (used in FLASH)ⁱ these components become,

$$\begin{aligned}\sigma_{\hat{\mathbf{r}}\hat{\mathbf{r}}} &= \Sigma [2\nu\partial_r v_{\hat{\mathbf{r}}} + (\zeta - 1)\nabla \cdot v] \\ r\sigma_{\hat{\mathbf{r}}\hat{\phi}} &= \Sigma \left[\nu \left(\partial_r (rv_{\hat{\phi}}) + \partial_\phi (rv_{\hat{\mathbf{r}}}) - \frac{2}{r} (rv_{\hat{\phi}}) \right) \right] \\ r^2\sigma_{\hat{\phi}\hat{\phi}} &= \Sigma [2\nu(\partial_\phi(rv_{\hat{\phi}}) + rv_{\hat{\mathbf{r}}}) + (\zeta - 1)r^2\nabla \cdot v]\end{aligned}$$

where $v^{\hat{\phi}} = v_{\hat{\phi}} = r\Omega$. Since the value of the bulk viscosity coefficient ζ is somewhat arbitrary, we set it to 0. Then simplifying the above using $\nabla \cdot v = \partial_r v_r + \frac{1}{r}\partial_\phi v_\phi + \frac{v_r}{r}$,

$$\begin{aligned}\sigma_{\hat{\mathbf{r}}\hat{\mathbf{r}}} &= \Sigma \nu \left[\partial_r v_{\hat{\mathbf{r}}} - \frac{1}{r} \partial_\phi v_{\hat{\phi}} - \frac{v_{\hat{\mathbf{r}}}}{r} \right] \\ \sigma_{\hat{\mathbf{r}}\hat{\phi}} &= \Sigma \nu \left[\partial_r v_{\hat{\phi}} + \frac{1}{r} \partial_\phi v_{\hat{\mathbf{r}}} - \frac{v_{\hat{\phi}}}{r} \right] \\ \sigma_{\hat{\phi}\hat{\phi}} &= \Sigma \nu \left[\frac{1}{r} \partial_\phi v_{\hat{\phi}} - \partial_r v_{\hat{\mathbf{r}}} + \frac{v_{\hat{\mathbf{r}}}}{r} \right]\end{aligned}\tag{2.20}$$

ⁱFor this conversion one needs to contract the coordinate tensor components with the orthonormal components of the coordinate basis vectors. This simply amounts to multiplying each ϕ -up component by r and each ϕ -down component by $1/r$ (e.g. $v^{\hat{\phi}} = rv^\phi = v_{\hat{\phi}} = v_\phi/r = r\Omega$).

2.6.4 Divergence of σ in Polar Coordinates

To compute the geometric source terms which will modify the 2D polar momentum equation we compute the divergence of the second rank tensor σ . Starting again in a coordinate bases we may write,

$$(\nabla \cdot \sigma)^i = \partial_j \sigma^{ji} + \Gamma^j_{kj} \sigma^{ki} + \Gamma^i_{kj} \sigma^{kj} \quad (2.21)$$

giving us the components of the viscous force

$$\begin{aligned} (\nabla \cdot \sigma)^r &= \partial_r \sigma^{rr} + \partial_\phi \sigma^{\phi r} + \frac{1}{r} \sigma^{rr} - r \sigma^{\phi\phi} \\ (\nabla \cdot \sigma)^\phi &= \partial_r \sigma^{r\phi} + \partial_\phi \sigma^{\phi\phi} + \frac{3}{r} \sigma^{r\phi} \end{aligned}$$

Transforming again to an orthonormal basis for implementation in FLASH,

$$\begin{aligned} (\nabla \cdot \sigma)^{\hat{r}} &= \partial_r \sigma^{\hat{r}\hat{r}} + \partial_\phi \frac{1}{r} \sigma^{\hat{\phi}\hat{r}} + \frac{1}{r} \sigma^{\hat{r}\hat{r}} - r \left(\frac{1}{r^2} \sigma^{\hat{\phi}\hat{\phi}} \right) \\ \frac{1}{r} (\nabla \cdot \sigma)^{\hat{\phi}} &= \partial_r \left(\frac{1}{r} \sigma^{\hat{r}\hat{\phi}} \right) + \partial_\phi \frac{1}{r^2} \sigma^{\hat{\phi}\hat{\phi}} + \frac{3}{r} \left(\frac{1}{r} \sigma^{\hat{r}\hat{\phi}} \right). \end{aligned}$$

Simplifying,

$$\begin{aligned} (\nabla \cdot \sigma)^{\hat{r}} &= \frac{1}{r} \partial_r (r \sigma^{\hat{r}\hat{r}}) + \frac{1}{r} \partial_\phi (\sigma^{\hat{\phi}\hat{r}}) - \frac{1}{r} \sigma^{\hat{\phi}\hat{\phi}} \\ (\nabla \cdot \sigma)^{\hat{\phi}} &= \frac{1}{r} \partial_r (r \sigma^{\hat{r}\hat{\phi}}) + \frac{1}{r} \partial_\phi (\sigma^{\hat{\phi}\hat{\phi}}) + \frac{1}{r} \sigma^{\hat{r}\hat{\phi}}. \end{aligned} \quad (2.22)$$

Plugging in the values of the components from (2.20) we have,

$$\begin{aligned}
(\nabla \cdot \sigma)_{\hat{r}} &= \frac{1}{r} \partial_r \left[r \Sigma \nu \left(\partial_r v_{\hat{r}} - \frac{1}{r} \partial_\phi v_{\hat{\phi}} - \frac{v_{\hat{r}}}{r} \right) \right] \\
&\quad + \frac{1}{r} \partial_\phi \left[\Sigma \nu \left(\partial_r v_{\hat{\phi}} + \frac{1}{r} \partial_\phi v_{\hat{r}} - \frac{v_{\hat{\phi}}}{r} \right) \right] \\
&\quad - \frac{\Sigma \nu}{r} \left[\frac{1}{r} \partial_\phi v_{\hat{\phi}} - \partial_r v_{\hat{r}} + \frac{v_{\hat{r}}}{r} \right] \\
(\nabla \cdot \sigma)_{\hat{\phi}} &= \frac{1}{r} \partial_r \left[r \Sigma \nu \left(\partial_r v_{\hat{\phi}} + \frac{1}{r} \partial_\phi v_{\hat{r}} - \frac{v_{\hat{\phi}}}{r} \right) \right] \\
&\quad + \frac{1}{r} \partial_\phi \left[\Sigma \nu \left(\frac{1}{r} \partial_\phi v_{\hat{\phi}} - \partial_r v_{\hat{r}} + \frac{v_{\hat{r}}}{r} \right) \right] \\
&\quad + \frac{\Sigma \nu}{r} \left[\left(\partial_r v_{\hat{\phi}} + \frac{1}{r} \partial_\phi v_{\hat{r}} - \frac{v_{\hat{\phi}}}{r} \right) \right]
\end{aligned} \tag{2.23}$$

The first two terms in each of the above look like a normal divergence of a rank-one-tensor and the third terms are the geometric source terms that we must add to the momentum equation. They are akin to the centrifugal and Coriolis terms which arise from a similar exercise performed on the ρvv term of (2.18). Note that each term has the units of force per volume while the components of the stress tensor have units of density times velocity squared which matches the ρvv term in the hydro equations.

2.6.5 Implementation in FLASH

In FLASH 3.2, the components (with respect to the orthonormal basis) of the viscous stress tensor are computed in the routine Diffuse_visc.F90. This routine then subtracts the flux indicated by the viscous stress tensor from the Σvv flux. Note that since FLASH computes fluxes on cell boundaries, the stress tensor components in Diffuse_visc.F90 must be computed on the lower face of the current sweep direction (See *e.g.* ([Edgar 2006](#))). We compute the viscous source terms along with the centrifugal and Coriolis source terms in the routine hy_ppm_force.F90.

We test the above implementation in FLASH with two tests, the viscously spreading ring of [Pringle \(1981\)](#) and a Cartesian shear flow.

2.6.6 Viscously Spreading Ring

The viscously spreading ring test begins with a delta function initial density distribution which spreads solely due to viscous forces. This test assumes axis-symmetry $\partial_\phi \equiv 0$ and exercises only the terms with $\sigma^{\hat{r}\hat{\phi}}$ in [\(2.23\)](#). The analytic solution assuming a constant coefficient of kinematic viscosity is well known and given by [Pringle \(1981\)](#).

To implement the viscously spreading ring test we choose $\nu = \text{cst.} = \alpha/\mathcal{M}^2$ with dimensionless parameter $\alpha = 0.1$, and mach number of the disc $\mathcal{M} = 100$. This choice of Mach number mitigates pressure effects and also allows time for a small initial transient to pass through the simulation domain without greatly affecting the evolution of the solution (wiggles in the earliest blue-dashed radial velocity curve of Figure [2.15](#)). As required by the analytic solution, we also turn off all terms in [\(2.23\)](#) which do not include $\sigma^{\hat{r}\hat{\phi}}$. The outer and inner boundaries are at $r_{\min} = 0.2$ and $r_{\max} = 2.2$ where boundary values are set by the time-dependent analytic values. We use a spatial resolution of 128 radial cells by 64 azimuthal cells and we start with initial conditions corresponding to the dimensionless initial time parameter $\tau = 12\nu t/r_0^2 = 0.032$, where r_0 is the initial position of the delta function ring.

Figure [2.15](#) shows the result of the test. Besides some expected deviation at the inner boundary, the numerical solution (dashed lines) agree well with the analytic solution (solid lines).

2.6.7 Cartesian Shear Flow

The viscously spreading ring test confirms that the most important terms for thin disc accretion (those containing $\sigma^{\hat{r}\hat{\phi}}$) are implemented properly. The Cartesian shear flow tests all of the terms in [\(2.23\)](#). The idea is to choose a problem which is analytic in Cartesian coordinates and use the

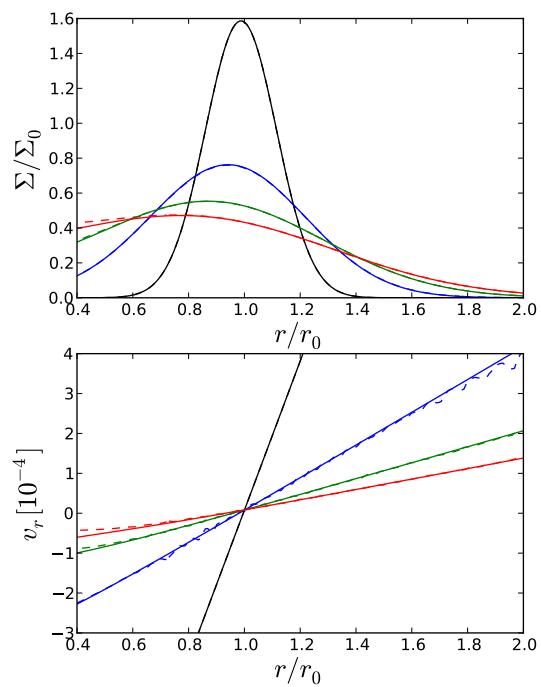


Figure 2.15: Viscously spreading ring test with constant coefficient of kinematic viscosity. See text for details.

computer to solve it in polar coordinates thereby exercising all of the polar derivatives to make up the simple Cartesian derivatives.

To set-up the Cartesian shear flow problem, we start with the Navier-Stokes equation for an incompressible fluid with constant coefficient of viscosity ν ,

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{1}{\Sigma} \nabla P + \nu \nabla^2 \mathbf{v} \quad (2.24)$$

Choosing constant pressure and $\mathbf{v} = v^x(y, t)\mathbf{e}_x$ reduces (2.24) to a simple 1D diffusion equation in $v_x(y, t)$

$$\partial_t v_x(y, t) = \nu \partial_y^2 v_x(y, t) \quad (2.25)$$

A solution is

$$v_x(y, t) = \frac{v_0}{\sqrt{2\pi\nu t}} \exp\left[\frac{-(y - y_0)^2}{4\nu t}\right] \quad (2.26)$$

while v_y is 0 for all time.

In practice we implement the Cartesian shear flow test with initial conditions,

$$\begin{aligned} \Sigma(r, \phi, t_0) &= 1.0 \\ v_r(r, \phi, t_0) &= v_x(y, t_0) \cos \phi \\ v_\phi(r, \phi, t_0) &= -v_x(y, t_0) \sin \phi \end{aligned} \quad (2.27)$$

with $y_0 = 1.0$, $v_0 = 1.0$, $\nu = 0.1$, $t_0 = 0.5$. Choice of outer and inner boundaries of $r_{\min} = 0.5$, $r_{\max} = 5.0$ allow the solution to not be greatly affected by the boundaries while supplying reasonable resolution requirements. Figure 2.16 plots the results of this test set up in FLASH for a

number of different resolutions and cell aspect ratios.

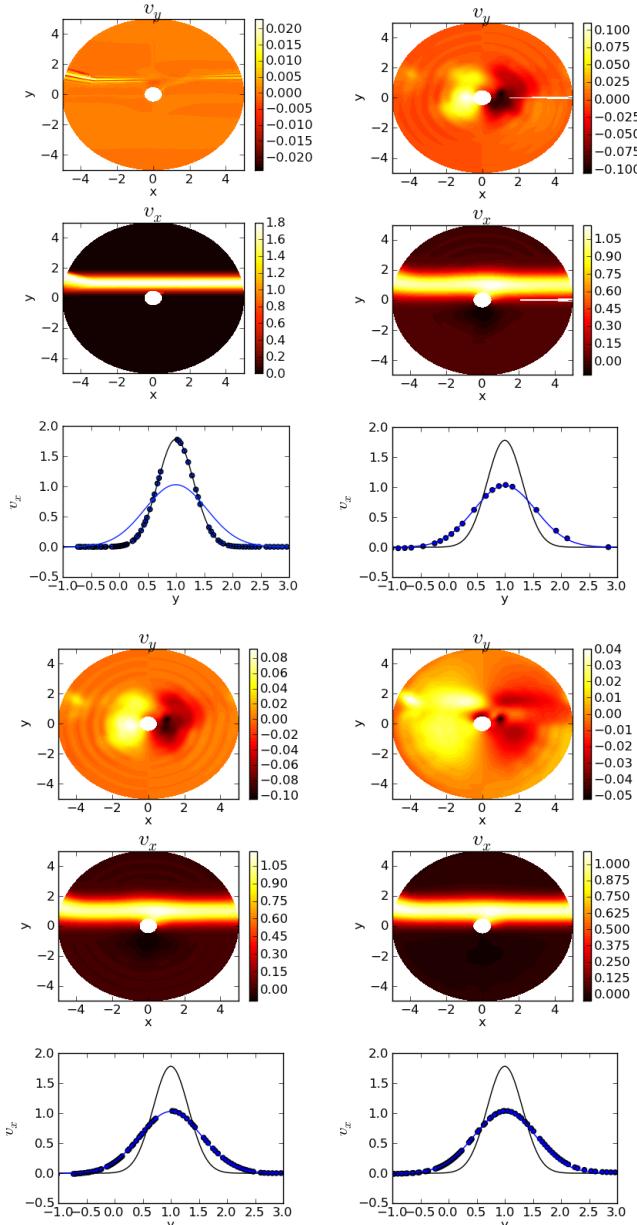


Figure 2.16: Results of the Cartesian shear flow test. Each panel is a snapshot taken at $t = 1.0$ ($t_0 = 0.5$) of components of velocity in the y -direction (top image in the panel) and components of velocity in the x -direction (middle image in the panel). The bottom image in each panel shows the initial x -velocity (black solid line) and the x -velocity at time $t=1.0$ (blue solid line) given by (2.26). The blue dots plotted in the bottom panels are the simulation values of the x -velocity sampled along the line $x = -2$. The top left panel is run with viscosity turned off at a spatial resolution of 64 radial by 512 azimuthal cells (making cells square at $r=1.0$). The other three panels have viscosity turned on at different spatial resolutions. The top right panel has 16 radial by 128 azimuthal cells (making cells square at $r=1.0$), the bottom left panel has 64 radial by 512 azimuthal cells (making cells square at $r=1.0$), and the bottom right panel has 24 radial by 512 azimuthal cells (making cells square at $r=2.5$). We see that the numerical solution follows well the analytic solution (2.26) for the evolution of the velocity. Also, the non-zero components of the y -velocity (which should stay zero), decreases with higher resolution and better chosen cell aspect ratio.

Chapter 3

A transition in circumbinary accretion discs at a binary mass ratio of 1:25

3.1 Introduction

Binaries embedded in gas discs are ubiquitous astrophysical systems. They are realized in the proto-planetary nebulae surrounding young stars and their growing planets ([Kley & Nelson 2012](#)) and possibly in young binary star systems as evidenced by circumbinary planets (*e.g.* [Orosz et al. 2012](#)). They also arise at the centers of galactic nuclei to which gas can be funneled to accompany an inspiraling massive black hole binary (MBHB) ([Barnes & Hernquist 1996](#), and see recent reviews by [Dotti et al. \(2012\)](#); [Mayer \(2013\)](#)).

Understanding the long-term evolution of the binary+disc system is complicated by the coupled nature of mass, angular momentum, and energy conservation for the total binary+disc system. The binary affects the structure of the disc, and the disc alters the orbital parameters of the binary. For planets and stars enveloped by a gas disc, the binary+disc interaction determines the migration and growth of the planets, dictating the post-disc-configuration of the planetary system. For a

MBHB+disc system, gas torques can alter the inspiral rate of the binary. The effect is important for deciphering the final parsec problem and predicting the rate of gravitational wave events due to MBHB mergers (Begelman et al. 1980; Gould & Rix 2000; Armitage & Natarajan 2002, 2005), and possibly even affecting the gravitational wave signal from inspiral (D’Orazio et al., *in preparation*; Yunes et al. 2011; Kocsis et al. 2011).

Additionally, interaction of the binary and disc can lead to periodic accretion (Hayasaki et al. 2007; MacFadyen & Milosavljević 2008; Cuadra et al. 2009; Roedig et al. 2011; Noble et al. 2012; Shi et al. 2012; Roedig et al. 2012a; D’Orazio et al. 2013; Farris et al. 2014; Dunhill et al. 2015; Shi & Krolik 2015) which can aid in identifying MBHB candidates in electromagnetic (EM) surveys (Haiman et al. 2009). As has been recently been made clear by the discovery of multiple MBHB candidates in EM time-domain surveys (Graham et al. 2015a,b; Liu et al. 2015), the interpretation of variability in EM surveys will rely heavily on our knowledge of how accretion variability depends on system parameters such as the binary mass ratio (see *e.g.* D’Orazio et al. 2015a,b).

Although the disc and binary are coupled, a useful first step in determining their mutual evolution is to determine the perturbation to the disc surface density by a binary on a fixed orbit. From this exercise three distinct regimes arise as a function of binary mass ratio ($q \equiv M_s/M_p$) and disc hydrodynamic parameters. Small-mass-ratio binaries, $q_{\text{lin}} \lesssim \mathcal{M}^{-5/2}\alpha^{1/2}$ (Duffell & MacFadyen 2013) excite linear spiral density waves in the disc. Here \mathcal{M} is the disc Mach number near the binary’s orbit and α is the alpha-law viscosity parameter (Shakura & Sunyaev 1973). The torque on the binary from the spiral density wave perturbation causes the binary’s orbit to shrink on the so-called Type I migration timescale (Goldreich & Tremaine 1979, 1980; Ward 1997).

Larger-mass-ratio binaries ($q \gtrsim \mathcal{M}^{-5/2}\alpha^{1/2}$) open a low surface density, annular gap in the disc, altering the binary migration rate. Often it is assumed that the migration in this regime is equal to the viscous drift rate in the disc (Lin & Papaloizou 1986; Nelson et al. 2000; D’Angelo & Lubow 2008), though recent numerical works have called this into question (Edgar 2008; Duffell

et al. 2014; Dürmann & Kley 2015).

The critical mass ratio for gap opening depends on the Mach number and disc viscosity. In the small mass ratio regime, this criterion has been explored analytically (*e.g.*, Goldreich & Tremaine 1979, 1980; Papaloizou & Pringle 1977) and numerically (*e.g.*, Bryden et al. 1999; Nelson et al. 2000; Papaloizou et al. 2004; Zhu et al. 2013). It has been thought that a necessary condition for a gap is that the secondary’s Hill radius be larger than the disc scale height (Lin & Papaloizou 1993; Crida et al. 2006). Goodman & Rafikov (2001) first proposed that this was not a necessary condition, but that low-mass perturbers could open gaps in low-viscosity discs, on much longer timescales. This has so far been validated in 2D simulations (Dong et al. 2011b; Duffell & MacFadyen 2012, 2013; Fung et al. 2014), though it has not yet been validated in 3D, due to the computational expense and long timescales necessary to capture gap opening in the low-mass-ratio regime.

For binary mass ratios near unity, hydrodynamical simulations find a central, low-density cavity cleared in the disc (*e.g.*, Artymowicz & Lubow 1994; Artymowicz & Lubow 1996; Farris et al. 2014). A variety of methods have been employed to determine the properties of circumbinary discs (CBDs) with near-equal binary masses. Cavity sizes have been estimated by calculating the truncation radius of the inner edge of CBDs, determined by orbit intersection and stability in the restricted three-body problem (R3Bp) (Rudak & Paczynski 1981) and through resonant torque calculations (Artymowicz & Lubow 1994; Artymowicz & Lubow 1996). Studies by del Valle & Escala (2012) and del Valle & Escala (2014) have examined cavity opening/closing conditions for $q > 0.1$ binaries with massive discs by calculating non-resonant torques due to a non-axisymmetric disc structure. Roedig et al. (2012a) have used 3D, smoothed-particle hydrodynamics to analyze the gas and binary torques acting on a massive disc and a near-unity, $q = 1/3$, mass ratio binary. They find binary orbital decay and binary eccentricity growth in the presence of a central cavity fed by gas streams. None of these studies, however, has asked what conditions are required to form

a central cavity rather than an annular gap, or even if an important distinction exists between the two regimes.

While the transition from the small mass ratio, weakly-perturbed regime to the larger mass ratio, gapped regime is well defined in the literature, the transition to the near-unity mass ratio state is not. Here we utilize the circular R3Bp as well as 2D viscous hydrodynamical simulations to show that there is a dynamically important transition from the gapped regime to the near-unity mass ratio regime which is marked by

1. A transition in surface density structure from a low-density annular gap towards a lower-density central cavity.
2. A transition from steady-state to strongly-fluctuating disc dynamics.
3. The development of strong asymmetry (i.e. a lopsided shape) of the central cavity and the slow precession of this cavity.

These characteristics of large binary mass ratio systems begin to appear above a mass ratio of $q \sim 0.04$. This is the same mass ratio above which stable orbits cease to exist in the co-rotation region of the R3Bp (Murray & Dermott 1999). Here we show that the above transition occurs over a wide range of hydrodynamical parameters and provide evidence that the transition is linked to R3B-orbital-stability criteria.

This study proceeds as follows. In §3.2.1 we use the integral of motion of the R3Bp to infer the structure of density gaps and cavities in a CBD. In §3.2.2 we integrate the equations of the circular R3Bp to elaborate on these findings. §3.2.3 analytically considers the effects of pressure and viscosity which are present for an astrophysical CBD. In §3.3.1, we present viscous hydrodynamical simulations to compare with the findings of the R3Bp analysis. In §3.3.2 we conduct a parameter study over disc viscosity and pressure, providing evidence that the high mass ratio, CBD transition is generated by the loss of stable particle orbits in the binary co-orbital region.

3.2 Restricted 3-body Analysis

Gap/cavity clearing and morphology are governed by the gravitational interaction between disc particles and the binary, as well as viscous and pressure forces. We conduct a purely gravitational study by ignoring hydrodynamical effects, treating the disc as a collection of test-particles obeying the circular R3Bp equations of motion (See *e.g.* Murray & Dermott (1999)). This gravitational study lends remarkable insight into the full hydrodynamical problem.

3.2.1 Restrictions on Orbits from Conserved Integrals of Motion

Circumbinary discs are characterized by global, low-density gaps and cavities. These structures are created when gas/particles cannot stably exist in a region. We begin by searching for restricted regions in the binary-disc plane. We seek to find which particles in the binary orbital plane are restricted from crossing into/out of the binary's orbit, and which are free to be expelled in the formation of a gap or cavity. To do this we utilize the Jacobi constant

$$C_J = 2U - v^2, \quad (3.1)$$

where U is the negative of the Roche potential of the binary depicted in Figure 3.1, v is the velocity of a test particle orbiting the binary, and all quantities are functions of the coordinates. As the only integral of motion in the R3Bp, the Jacobi constant is conserved along a test-particle orbit. Since C_J is conserved, a particle with Jacobi constant C_J^p is restricted to regions of the binary orbital plane where $C_J^p \leq 2U$, else the particle would have a complex velocity. We use this property to draw zero-velocity curves (ZVCs), level curves of the Roche potential (see Figure 3.1), defined by the equation $C_J^* = 2U$. Particles with $C_J > C_J^*$ cannot enter the closed region delineated by the ZVCs.

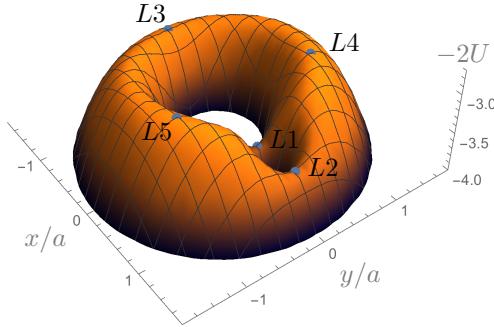


Figure 3.1: A three-dimensional representation of the effective binary potential in the co-rotating frame for a binary with mass ratio $q = 0.1$. Here we have plotted twice the Roche potential, $-2U$, the negative of the Jacobi constant for a particle with zero velocity (see Eq. 3.1). The x and y coordinates in the binary plane are measured in units of the binary separation a . The primary and secondary are located at $(x_p, y_p) = (-a/(1 + 1/q), 0)$ and $(x_s, y_s) = (a/(1 + q), 0)$ respectively. The five Lagrange points are labeled for reference.

For the purpose of studying gap/cavity structure, we examine the ZVCs which connect, *i.e* separate the binary plane into ≥ 2 distinct regions, an inner disc and an outer disc. We define a critical ZVC corresponding to the smallest value of the Jacobi constant for which ZVCs connect, C_J^{crit} . This is the Jacobi constant with ZVC which passes through the second Lagrange point, L2. Particles in the disc for which $C_J > C_J^{\text{crit}}$ cannot pass between the inner and outer disc. ZVCs for values of C_J at and around the critical value are illustrated in Figure 3.2 for a binary with mass ratio $q \equiv M_s/M_p = 0.1$.

To determine the structure of the R3B disc, we ask if particles in the inner disc have $C_J < C_J^{\text{crit}}$; if so, they may be evacuated by the binary to form a central cavity. If the majority of particles in the inner (outer) disc have $C_J > C_J^{\text{crit}}$, then they are trapped in the inner (outer) disc, and an annular gap will define the density structure. This analysis does not determine whether the un-trapped orbits actually cross the ZVCs or not - this depends on the initial magnitude and orientation of the velocity vector of each particle that is not trapped by the critical ZVC (green regions of Figure 3.3 below). We determine the fate of the un-trapped particles in §3.2.2.

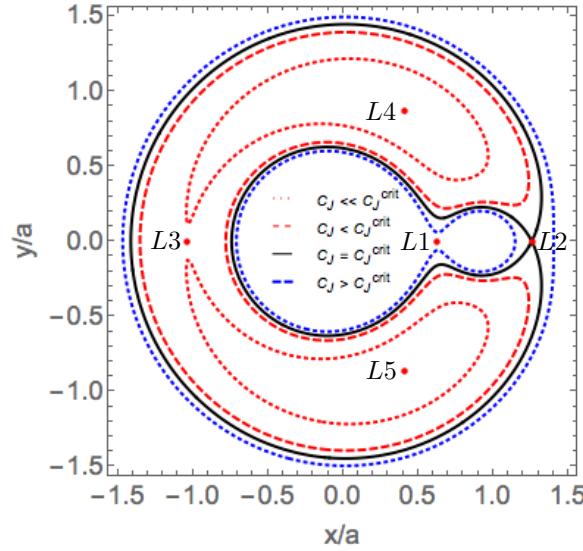


Figure 3.2: Zero-velocity curves for four different values of the Jacobi constant for a binary with mass ratio $q = M_s/M_p = 0.1$ (level curves of the potential plotted in Figure 3.1). For $C_J \geq C^{crit}$ (blue thick dashed and solid black), the zero-velocity curves connect, separating the binary plane into distinct inner and outer regions (as well as a third region around the secondary for large enough values of the Jacobi Constant). For $C_J < C^{crit}$ (red lines), the zero-velocity curves open at L2 and then at L3 for even smaller C_J . The critical zero-velocity curve (black) passes through the Lagrange point L2. The five Lagrange points are labeled for reference.

Since the value of C_J depends on a particle's position as well as its velocity, we must prescribe a velocity profile in the disc. We choose a prescription given by the virial theorem which approaches the Keplerian value for the binary at large $r \equiv \sqrt{x^2 + y^2}$ and the Keplerian value for each BH at $r \rightarrow r_p$ and $r \rightarrow r_s$

$$v_\phi = \sqrt{\frac{GM_s}{r_s} + \frac{GM_p}{r_p}} - r\Omega_{\text{bin}}. \quad (3.2)$$

where r_p and r_s are the ϕ dependent distances from the primary and secondary, and we subtract the angular frequency Ω_{bin} because we are working in the rotating frame of the binary.

Given the velocity profile in Eq. (3.2), Figure 3.3 displays the morphology of restricted regions in the binary orbital plane. Particles trapped in the outer (inner) disc have Jacobi constant greater than the critical value and are painted blue (red). Disc particles which are not restricted to either region are painted green. The enclosed area of the critical ZVC, which also consists of un-trapped particles, is painted dark-green.

We emphasize that the shape of the restricted regions depends on the choice of velocity profile, while the ZVCs are independent of this choice. An extreme choice of $|v| = 0$ in the co-rotating frame causes the blue and red regions to extend to the boundary of the dark green region; all particles outside of the critical ZVC are forbidden to cross into the area enclosed by the ZVC. Conversely, if the velocities in the co-rotating frame are large, the red region can vanish and the blue region can recede far from the binary. This, however, is just a result of filling the system with initially unbound particles and has little meaning for an accretion disc. Choosing a purely Keplerian velocity distribution can also cause issues. The Keplerian velocity approaches infinity at the origin regardless of primary position. The result is an artificial depletion of trapped inner-disc particles for larger mass ratios. Thus, to create a representation that realistically describes the restricted regions in a quasi-steady-state CBD, one must choose a velocity distribution near an equilibrium state (though one may not exist for larger mass ratios) and without artificial singularities. This

leads us to Eq. (3.2).

Because the green regions in Figure 3.3 consists of particles which are free to cross the binary orbit, we identify these regions with a putative gap/cavity. We examine this claim in §3.2.2 and find that their locations provide an adequate tracer for the gap/cavity size and shape. Furthermore, the locations of the outer gap/cavity edge identified in this manner, and those found in Figure 3.5 below, agree with the locations of circumbinary disc truncation computed from the stability and intersection of periodic R3Bp orbits (Rudak & Paczynski 1981). This suggests a correspondence between locations in the outer binary potential where periodic orbits exist, and locations where particles with approximately Keplerian velocity are trapped outside the binary orbital barrier.

We note that the meaning of our Jacobi constant analysis in regions where periodic orbits cannot exist is somewhat ambiguous. In these regions, we must assign velocities to particles which do not correspond to fluid velocities in a stable disc. Rudak & Paczynski (1981) have delineated these regions explicitly, and they turn out to coincide with the green, untrapped particles of our analysis. As long as this is true, our analysis is consistent with particle stability considerations.

As the binary mass ratio is increased, Figure 3.3 shows that the putative gap grows in size and also morphs in shape from a small horseshoe, or annulus, in the orbit of the secondary, to a cavernous shape which includes the inner region. Additionally, as the mass ratio is increased, the fraction of particles trapped in a disc around the primary decreases while the number of particles trapped in a disc around the secondary increases. The following picture appears: for small mass ratios, the system consists of an inner disc (red) and outer disc (blue) separated by a low-density annulus. For larger mass ratios, Figure 3.3 depicts circum-primary and circum-secondary (mini-)discs (red) surrounded by a CBD (blue). The change in disc morphology is the most stark between the $q = 0.01$ and $q = 0.1$ panels of Figure 3.3.

Within the analysis so far, the transition between an inner and outer disc, vs. mini-discs+CBD, is defined only by terminology. There is however an important dynamical transition which occurs

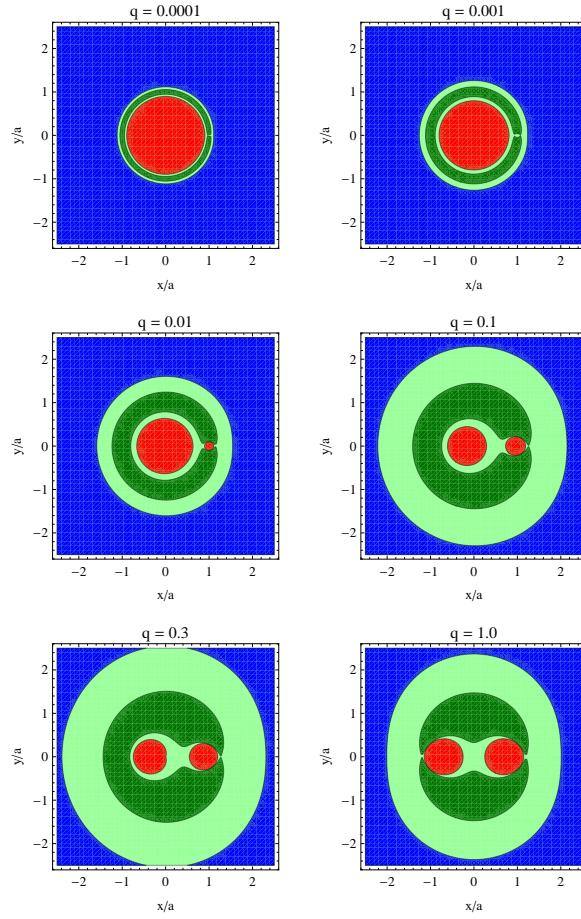


Figure 3.3: The dark green regions are bounded by the zero-velocity curve which passes through L2, delineating the smallest restricted regions which connect and separating the binary plane into distinct inner and outer regions. Particles trapped outside (inside) of the dark green region are labeled blue (red). Depending on their velocity vectors, light- and dark-green particles are free to move from inner to outer regions.

within the R3Bp for mass ratios $q > 0.04$, namely the loss of stable orbits around the L4/L5 equilibrium points in the binary co-orbital region. In what follows we provide evidence for a well-defined and physically meaningful critical mass ratio, related to this stability criterion, which divides the two regimes described above.

3.2.2 Restrictions on Orbits from Equations of Motion

We now elaborate on the picture painted in Figure 3.3 by integrating the circular R3Bp equations of motion for a disc of 256^2 test particles with initial velocity given by Eq. (3.2). We paint each of the particles the colour corresponding to its initial location in Figure 3.3. Using an adaptive step, Dormand Prince, 5th order Runge Kutta method (Press et al. 2007), we evolve the orbit of each particle for 100 binary orbits, conserving the Jacobi Constant to fractional order better than 10^{-6} (the majority of orbits conserve C_J to machine precision). Note that, for large-mass-ratio binaries, a small fraction of the green particle orbits conserve C_J to only the 10^{-2} level. This occurs for green particles which undergo large accelerations and move to large distances early in the evolution and does not affect the fate of the disc.

Figure 3.4 shows the location of each particle after only one binary orbit. During the first binary orbit, for mass ratios $q \gtrsim 0.04$, the green regions funnel towards the L2 and L3 points into streams reminiscent of those seen in hydrodynamical simulations and also in the R3Bp study of D’Orazio et al. (2013). Recall that the green particles have Jacobi Constant corresponding to ZVCs which are not connected, but which still delineate a restricted region (red-dashed curve in Figure 3.2). The ZVCs of the green particles allow transfer of the green particles between inner and outer disc via the lowest barriers in the Roche potential, the L2 and L3 points respectively (the lowest point in the potential, L1, allows transfer between primary and secondary - see Figures 3.1 and 3.2). Without additional forces due to viscosity, pressure, or particle self gravity, the streams do not persist after a few orbits. If, however, dissipative forces refill the green region, or particles can

interact, the streams continue to form as is seen in hydrodynamical simulations (see Figure 3.16 of the Appendix).

For smaller mass ratios $q \lesssim 0.04$, the L3 equilibrium point is much higher than the L2 point and particles stream in horseshoe orbits past L2 only. The particle density structure, in the $q = 0.0001$ and $q = 0.001$ cases (Figure 3.4), is reminiscent of low-mass-ratio hydrodynamical simulations where a spiral density wave propagates from the location of the secondary (see the top left panel of Figure 3.9 below).

Figure 3.5 shows the final location of each particle, after 100 binary orbits. We find that the red and blue particles do indeed remain trapped on their respective sides of the binary orbit. For $q \leq 0.001$ the green particles move along horseshoe and tadpole (horse-pole) orbits. As we approach $q = 0.01$, we find green particles librating around the L4 or L5 points. Figure 3.6 zooms in on mass ratios between $q = 0.02$ and $q = 0.08$; for $q \lesssim 0.04$, some green particles remain in the co-orbital region. For mass ratios $q \gtrsim 0.04$, green particles are cleared from the binary orbit. This is due to the lack of stable L4/L5 orbits in the R3Bp for $q > 0.04$. The widening of the annulus and loss of particles on horse-pole orbits at $q \simeq 0.04$ marks a dynamically significant transition caused by a change in orbital stability. This loss of orbital stability will be especially important in the hydrodynamical case, where particles can interact.

3.2.3 Hydrodynamical Effects

Before numerically solving the viscous hydrodynamical equations, we attempt to estimate the effects of pressure and viscosity by making simple extensions to the standard R3Bp. Pressure and viscosity provide additional forces in the R3Bp equations of motion which destroy the conservation of C_J and modify orbital stability in the standard R3Bp.

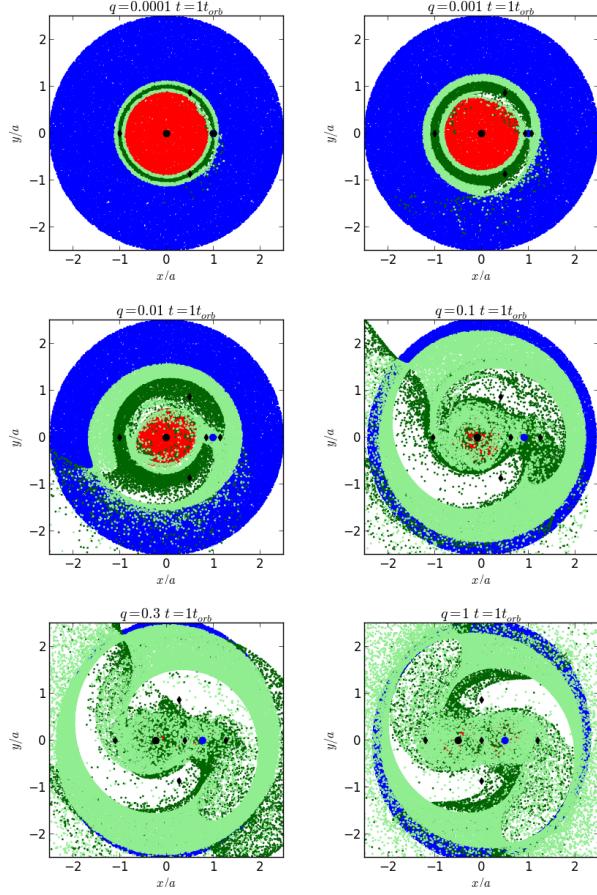


Figure 3.4: Each panel is the result of evolving an initially spatially random distribution of particles, within radius $r < 2.5a$, via the R3Bp equations, for one binary orbital period. The colouring of particles refers to the initial placement of a particle as in Figure 3.3. The black diamonds mark the Lagrange points (see Figures 3.1 and 3.2). These snapshots, after only one binary orbit, show the formation of streams acting to deplete green particles.

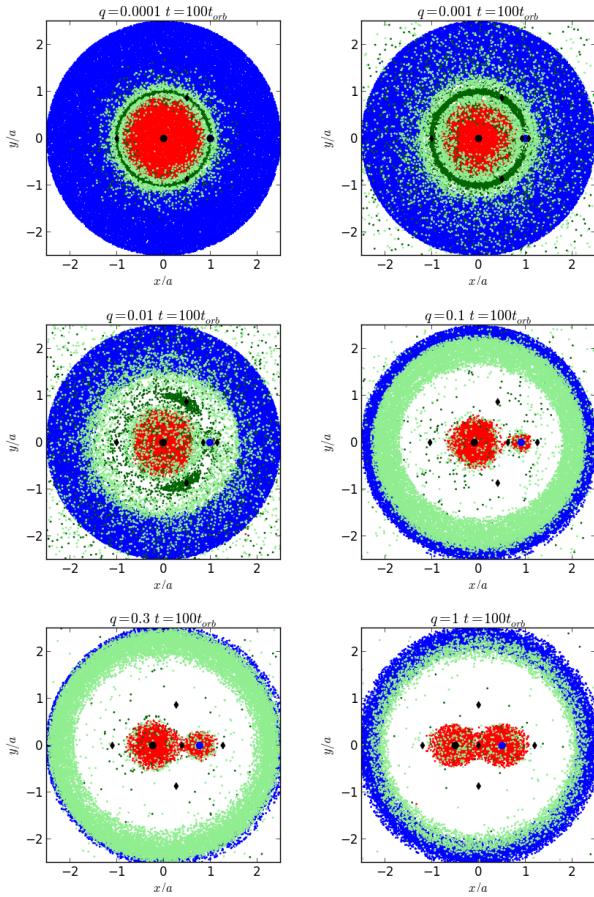


Figure 3.5: Same as Figure 3.4 except after 100 binary orbital periods.

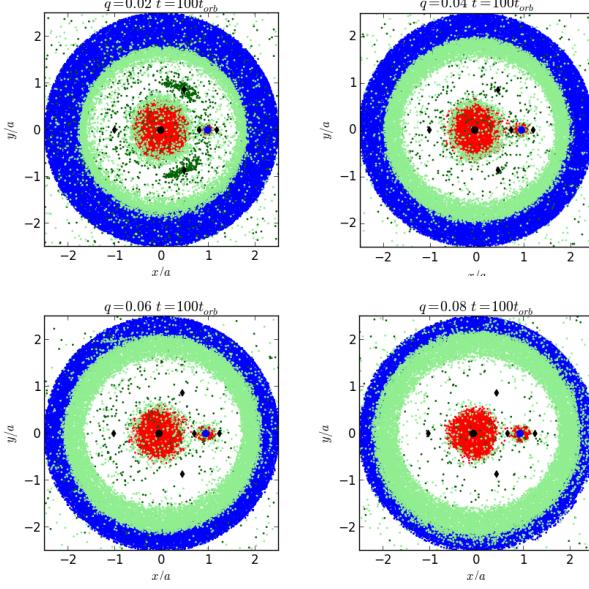


Figure 3.6: The same as Figure 3.5 except zooming in on the mass-ratio range $0.02 \leq q \leq 0.08$.

3.2.3.1 Pressure

Hydrodynamic pressure can be thought of as altering the effective potential of the binary at any point along a particle’s trajectory thus altering the ZVCs and hence the boundaries of the restricted regions in the previous section. We estimate the magnitude of pressure below which the R3B analysis may still be relevant by comparing the Jacobi constant to the analogue for a pressurized flow, Bernoulli’s constant

$$C_B = 2U - v^2 - 2 \int \frac{dP}{\rho}, \quad (3.3)$$

where the integral is along the trajectory of a fluid element from a reference point to the point of evaluation. Subtracting Eqs. (3.1) and (3.3) we find $2 \int dP/\rho = C_J - C_B$. By setting the difference in integrals of motion equal to the change in Jacobi constant across the binary orbit, we estimate the level of disc pressure necessary to overflow the previously described restricted regions of the purely gravitational problem (see Paczynski & Rudak 1980; Rudak & Paczynski 1981, who

perform a similar calculation). For an ideal gas,

$$\int \frac{dP}{\rho} = \frac{(c_s^{\text{ad}})^2}{\gamma - 1} = \frac{\gamma(c_s^{\text{iso}})^2}{\gamma - 1} \quad \text{Adiabatic Flow}$$

$$\int \frac{dP}{\rho} = (c_s^{\text{iso}})^2 \ln \frac{\rho}{\rho_0} \quad \text{Isothermal Flow,} \quad (3.4)$$

where the isothermal sound speed c_s^{iso} is related to the adiabatic sound speed c_s^{ad} by a factor of the adiabatic index γ . In the last line, the ratio ρ/ρ_0 comes from integrating from reference position outside of the binary orbit at density ρ_0 , to a point inside the putative gap/cavity at density ρ .

The condition for vertical hydrostatic equilibrium in a thin Keplerian disc is $P/\rho = (GM/r)(H/r)^2$. This allows us to write the sound speed $\gamma(c_s^{\text{iso}})^2 = (c_s^{\text{ad}})^2 = \gamma\Omega_K^2 H^2$ in terms of the Keplerian angular frequency of the disc Ω_K and disc height H . Then the disc orbital Mach number can be expressed as $\mathcal{M} \equiv v_k/c_s^{\text{iso}} = (H/r)^{-1}$, encoding the temperature and pressure forces for a disc in vertical hydrostatic equilibrium. From vertical hydrostatic balance, we place a condition on the disc aspect ratio, at the location of the secondary, for which pressure forces can overcome the binary gravitational barrier,

$$\left(\frac{H}{r_s}\right)_{\text{ad}}^{-1} \lesssim \sqrt{2 \frac{\gamma}{\gamma - 1} \frac{GM_{\text{bin}}}{\Delta C_J^{\text{gap}} a} (1 + q)}$$

$$\left(\frac{H}{r_s}\right)_{\text{iso}}^{-1} \lesssim \sqrt{2 \left| \ln \left(\frac{\rho}{\rho_0} \right) \right| \frac{GM_{\text{bin}}}{\Delta C_J^{\text{gap}} a} (1 + q)}, \quad (3.5)$$

where ΔC_J^{gap} is the variation of C_J across the dark-green restricted regions of Figure 3.3. Operationally, we choose ΔC_J^{gap} to be the difference in C_J at L2 and L4 (or L5), as this is the largest ΔC spanning the dark-green restricted regions.

We emphasize that Eqs. (3.5) are not gap/cavity closing conditions; they are necessary conditions for the pressure to overcome the gravitational barrier of the binary. These conditions do

not take into account the direction of pressure forces, and they do not consider hydrodynamical shocks, which have been shown, in the small mass ratio regime ($q \leq 10^{-3}$), to be responsible for gap opening in competition with viscous forces (Dong et al. 2011d,b; Duffell & MacFadyen 2013; Fung et al. 2014; Duffell 2015). True gap/cavity closing conditions must incorporate a full hydrodynamical treatment, whereas the conditions (3.5) provide a necessary condition for pressure to dominate the flow dynamics.

In Figure 3.7 we plot critical disc aspect ratios (3.5) as a function of binary mass ratio. Overplotted dots mark the positions of hydrodynamical simulations run in this study (§3.3). For the density ratio ρ/ρ_0 across a gap/cavity, we use the empirical relation from Duffell & MacFadyen (2013). All of the primary simulations run in this study have disc aspect ratios below the critical limit. To probe this limit we run high aspect ratio, $q = 1$ hydrodynamical simulations for the isothermal and adiabatic cases (see §3.3). Figure 3.8 shows that the results of these high-aspect-ratio (equivalently, high pressure, thick disc, high temperature, or low-orbital-Mach number) simulations^a are in agreement with our Eqs. (3.5). The isothermal case shows a stark difference between an overflowed cavity at $H/r_s = 1/3$ and a cleared cavity at $H/r_s = 1/6$ while the adiabatic case is less extreme, with the $\mathcal{M} = 3$ adiabatic case exhibiting a marginal central cavity.

Duffell & MacFadyen (2013) have shown that in the small mass ratio, thin disc case, the clearing of a gap always occurs for inviscid discs. This is because, in the absence of viscous forces, density waves generated by the binary will shock and deposit angular momentum into the disc clearing a gap without competition. It is not clear whether this is always true for high mass ratio, highly-pressureized discs; the overflowed CBDs in Figure 3.8 may eventually clear a cavity due to shocks, overcoming pressure on a longer timescale than considered here.

In summary, a necessary condition for pressure forces to overflow the binary cavity is predicted by Eq. (3.5). Equal-mass binary simulations show this condition to be sufficient in the large mass

^aIn these hot discs, we often see a one-armed spiral structure similar to that reported by Shi & Krolik (2015), also for hotter discs.

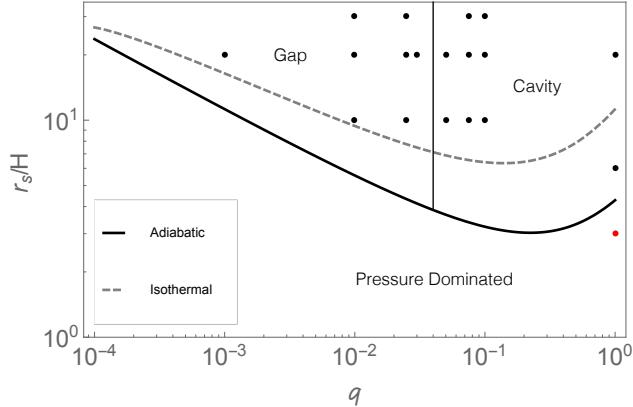


Figure 3.7: Delineation of different phases in a non-viscous circumbinary disc. The y-axis records the inverse disc aspect ratio, equivalent to the orbital Mach number; a smaller value signifies larger pressure forces; large pressure forces preclude reasoning based on a purely gravitational analysis. Points represent the parameters of hydrodynamical simulations run in this study (§3.3). Red denotes a simulation with a filled gap/cavity.

ratio regime, at least initially. Hence, having a high pressure in the disc does not impede the cavity formation, or whether it is lopsided, unless the pressure becomes so large that the disc is no longer thin. In this case, 3D effects can also become important, invalidating the 2D analysis above.

3.2.3.2 Viscosity

In order to more closely compare to viscous hydrodynamical simulations, and with the goal of linking orbital stability to the CBD phase transition, we follow Murray (1994) and Murray & Dermott (1999) to add an external viscous force $\mathbf{F}(x, y, \dot{x}, \dot{y})$ to the R3Bp equations of motion. With these viscous R3Bp equations, we conduct a linear-stability analysis and integrate the viscous R3Bp equations of motion for a disc of test particles. We refer the interested reader to the Appendix for details; here we only state a brief summary of the three main conclusions.

1. Upon integrating the viscous R3Bp equations for an initial disc of test particles, we find that viscosity indeed acts to overflow gaps in the R3Bp and continually generates streams which penetrate the binary orbit (compare Figure 3.16 to Figure 3.5). The time rate of change of

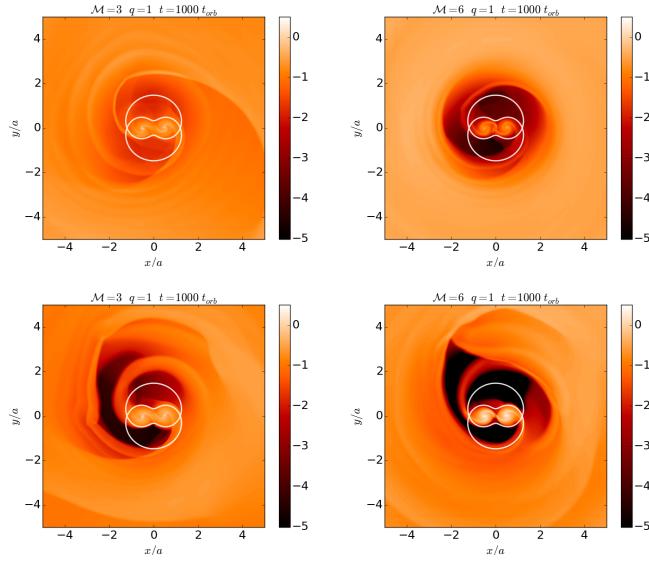


Figure 3.8: Snapshots of the surface density distribution (shown in units of the unperturbed value, with a logarithmic colour scheme) for an equal mass binary with disc aspect ratios ($r_s/H \equiv \mathcal{M}$) surrounding the pressure dominated condition Eq. (3.5). Here we set the viscosity to be very small (the coefficient of kinematic viscosity is $\nu = 10^{-6}a_0^2\Omega_{\text{bin}}$, where Ω_{bin} is the binary angular frequency) in order to examine the analytic R3Bp prediction (3.5). The top row is for an isothermal equation of state $P = (c_s^{\text{iso}})^2\Sigma$ and the bottom row is for an adiabatic equation of state $P = (c_s^{\text{ad}})^2\Sigma^{5/3}$.

the Jacobi constant due to viscosity is given for a simple Keplerian velocity prescription in the Appendix.

2. The addition of viscosity causes orbits around L4 and L5 to become formally unstable for all binary mass ratios (Figure 3.14). However, the instability timescale for orbits below $q = 0.04$ is of order a viscous time, while for $q > 0.04$ it drops to of order the binary orbital time. Hence, viscosity does not greatly change the mass ratio at which orbits around L4 and L5 become effectively unstable. If the phase transition at $q = 0.04$ is linked to the instability of orbits in the co-orbital region of the binary, then the level of viscosity should not greatly affect the mass ratio of the $q = 0.04$ phase transition - for thin discs for which Eqs. (3.5) hold. We verify this with a suite of hydrodynamical simulations in the next section.
3. Viscosity induces a difference in orbital instability timescales for particle orbits around the L4 and L5 points (Figure 3.15). This difference is proportional to the magnitude of viscosity and it is small for $q < 0.04$, becoming larger for $q > 0.04$. While this asymmetry between L4 and L5 may aid in seeding the instability to an asymmetric cavity, it cannot be the only mechanism which causes the cavity to be lopsided. For example, symmetry between L4 and L5 must be restored at $q = 1$, and a prominent asymmetry still appears in the disc morphology in this case. Future work will explore in more detail the relationship between the different orbital stability timescales and the cavity morphology.

3.3 Hydrodynamical simulations

3.3.1 Fiducial Simulations

In this section we show that the intuition gained from the R3Bp carries over to the hydrodynamical regime by running viscous 2D hydrodynamical simulations of the binary-disc system. We utilize

the moving mesh code DISCO (Duffell & MacFadyen 2011) to simulate a binary embedded in a locally isothermal, initially uniform surface density ($\Sigma = \text{cst.}$) disc. To enforce a locally isothermal equation of state, the vertically integrated pressure is set as $P = (v_{\text{eff}}/\mathcal{M})^2 \Sigma$, where

$$v_{\text{eff}} = \sqrt{\frac{GM_s}{r_s} + \frac{GM_p}{r_p}} \quad (3.6)$$

is the initial azimuthal velocity in the disc for an initially, spatially constant Mach number \mathcal{M} , and Σ is the disc surface density. The initial radial velocity is given by viscous diffusion

$$v_r = -\frac{3}{\Sigma r^{1/2}} \partial_r [\nu \Sigma r^{1/2}] , \quad (3.7)$$

where we choose a coefficient of kinematic viscosity which is constant in space and time $\nu = \alpha a_0^2 \Omega_0 / \mathcal{M}^2$, for fiducial values $\alpha = 0.01$ and $\mathcal{M} = 20$. These choices assure gap/cavity formation for $q \gtrsim 10^{-4}$. Here a_0 and Ω_0 represent the fixed separation and angular frequency of the binary. The simulation domain extends from the origin to $r_{\text{max}} = 12.0a$ employing a log grid (r_{max} can range from $8.0a$ to $100.0a$ for non-fiducial disc parameters as discussed in §3.3.2). The radial resolution is $\Delta r \sim 0.02a$ inside the binary orbit, similar to that of Farris et al. (2014), except we do not add an additional high-resolution region around each BH. We choose an outflow outer boundary condition. We do not apply an inner boundary condition and instead allow gas to flow through six cells at $r < 0.05a$. Around each binary component we employ a density sink of size r_{acc} which removes gas at the rate $3/2\Sigma r_n^{-2}\nu$ from each cell within the sink radius, where r_n is the distance from the n^{th} binary component. The total accretion rate onto each binary component is found by integrating over each cell located inside of the sink radius. Unless otherwise specified, we set $r_{\text{acc}} = \min \{R_{\text{Hill}}, 0.5a\}$ for the secondary and $r_{\text{acc}} = 0.5a$ for the primary, where R_{Hill} is the Hill radius of the secondary.

We evolve the viscous hydrodynamical equations for at least one viscous time at the location of

the binary. We plot the resulting 2D surface-density distribution for different mass ratios in Figure 3.9. We make three observations on the CBD-density structure:

- **An annular gap morphs into a central cavity:** Comparing the $q = 0.03$ and the $q = 0.05$ panels in Figure 3.9, we see that there is a transition occurring by $q = 0.03$ from an annular gap to a lopsided cavity (consistent with Farris et al. (2014) which used a hotter and more viscous disc) as well as a constant α (as opposed to constant ν) viscosity prescription. We explore the dependence of the critical mass ratio on disc Mach number and viscosity in §3.3.2 below.
- **The surface density decreases in the co-orbital region:** The surface density in the annular gap decreases with increasing mass ratio until the gap becomes nearly devoid of gas above the transition to a lopsided cavity. While there are multiple factors that determine gap depth (and width; Fung et al. 2014; Kanagawa et al. 2015; Duffell 2015), we note that the loss of stable orbits librating around L4/L5 and the disappearance of gas near the orbit of the secondary occur near the same mass ratio, suggesting that the existence of stable L4/L5 orbits may be necessary for long-lived gas in the gap. This behaviour helps to define the phase transition occurring at the same mass ratio.
- **A lopsided, precessing cavity appears:** The hydrodynamical study introduces a phenomenon accompanying the annular-gap to central-cavity transition which is not captured by the R3Bp alone, a lopsided, precessing cavity. This is illustrated by the surface density contours in the bottom panels of Figure 3.9. A lopsided cavity has been observed in previous studies which have used different numerical codes and physical assumptions (MacFadyen & Milosavljević 2008; Shi et al. 2012; Noble et al. 2012; D’Orazio et al. 2013; Farris et al. 2014, 2015b,a; Dunhill et al. 2015; Shi & Krolik 2015). A complete description of the growth of lopsided cavities, especially in the symmetric equal-mass binary case, does not yet exist. However,

the growth has been attributed to an initially small asymmetry in stream generation which causes a feedback loop between stream strength and cavity-edge location (Shi et al. 2012; D’Orazio et al. 2013). We observe that lopsided cavities are generated when stable orbits do not exist in the co-orbital region. Hence, the generation of a lopsided cavity may be intimately tied to orbital dynamics in the co-orbital region. When there are no stable co-orbital orbits, fluid passing into the co-orbital regime is flung into the outer disc rather than librating on stable orbits.

To further examine the connection between orbital stability and the transition to a time-dependent, lopsided cavity, we track the flow of gas in the co-orbital region. To do this we evolve separate conservation equations for two passive scalars. A passive scalar is a scalar quantity which obeys a conservation equation given an initial concentration and the fluid velocities from the hydrodynamic problem. We start one scalar outside of the critical ZVC and the other inside. The light green regions of Figures 3.3 and 3.5 show that, even in the non-hydrodynamic case, this setup should result in the passive scalars moving across the orbit of the binary, which is what we wish to track. Figure 3.10 plots the evolution of the passive scalars as well as fluid-velocity vectors for three different mass ratios. We colour the passive scalar which is initially inside (outside) of the critical ZVC red (blue).

The first row in Figure 3.10 tracks the fluid motion for a binary mass ratio well below the stability/phase transition, $q = 0.001$. It is clear that gas inside the orbit moves outwards across the position of the secondary at L2 and travels along the horseshoe orbits delineated by the critical ZVC. Gas initially outside of the orbit similarly moves across L2 and enters onto horseshoe orbits which eventually deposit the gas onto the inner disc.

The second row in Figure 3.10 tracks the fluid motion for a binary with $q = 0.01$, which is still below the R3Bp linear-stability mass ratio. From the velocity vectors it is clear that the mean motion of the fluid is along the horseshoe and librating L4/L5 orbits. In the $q = 0.01$ case, the

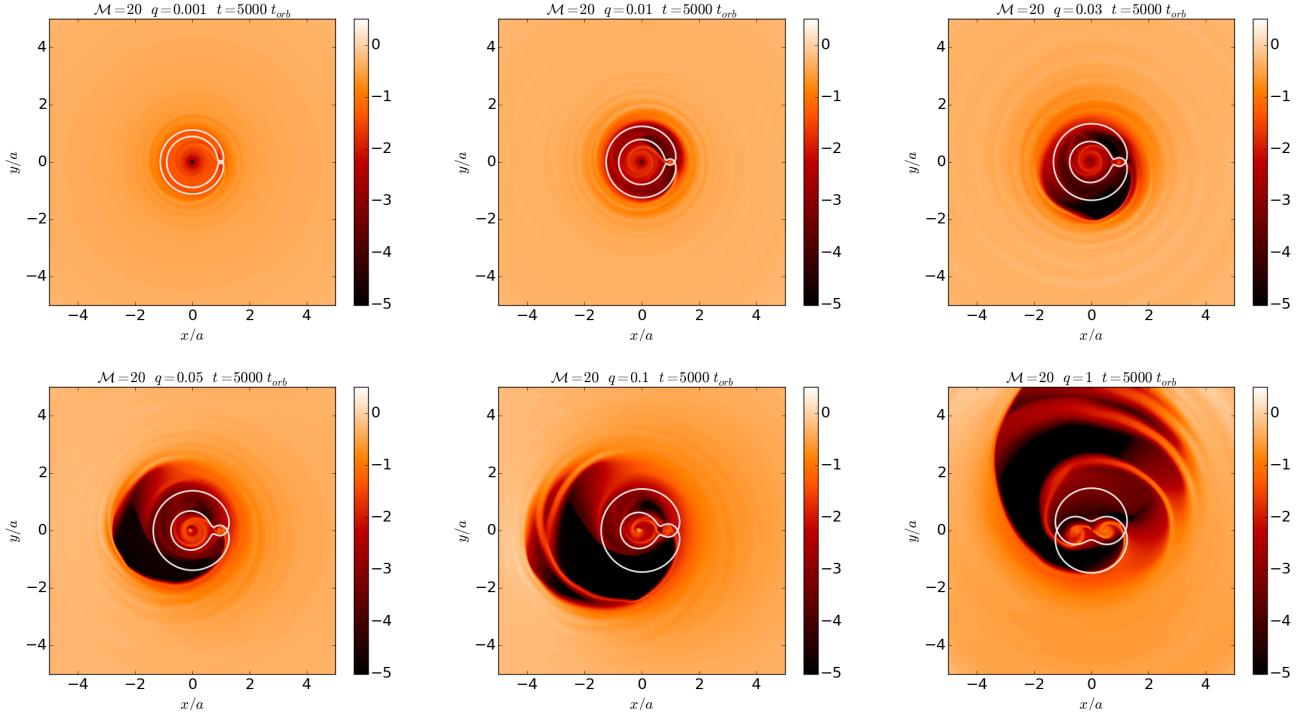


Figure 3.9: Snapshots of the surface density distribution (shown in units of the unperturbed value, with a logarithmic colour scheme) from hydrodynamical simulations for a disc with orbital Mach number $\mathcal{M} \equiv r/H \equiv v_{\text{eff}}/c_s = 20$, and constant coefficient of kinematic viscosity $\nu = 0.01a_0^2\Omega_{\text{bin}}/\mathcal{M}^2$. The binary mass ratio increases from left to right, top to bottom, as labeled. For small mass ratios, the system is in nearly steady-state and an annular gap is cleared in the orbit of the secondary black hole. For $q \gtrsim 0.03$, the gap morphs into an even lower density time-dependent, precessing central cavity. The critical zero-velocity curve, which passes through L2, is over-drawn in white. The relatively shallow annular gap in the $q = 0.001$ case is difficult to see on this scale because the accretion prescription and inner boundary cause the inner disc to drain onto the primary.

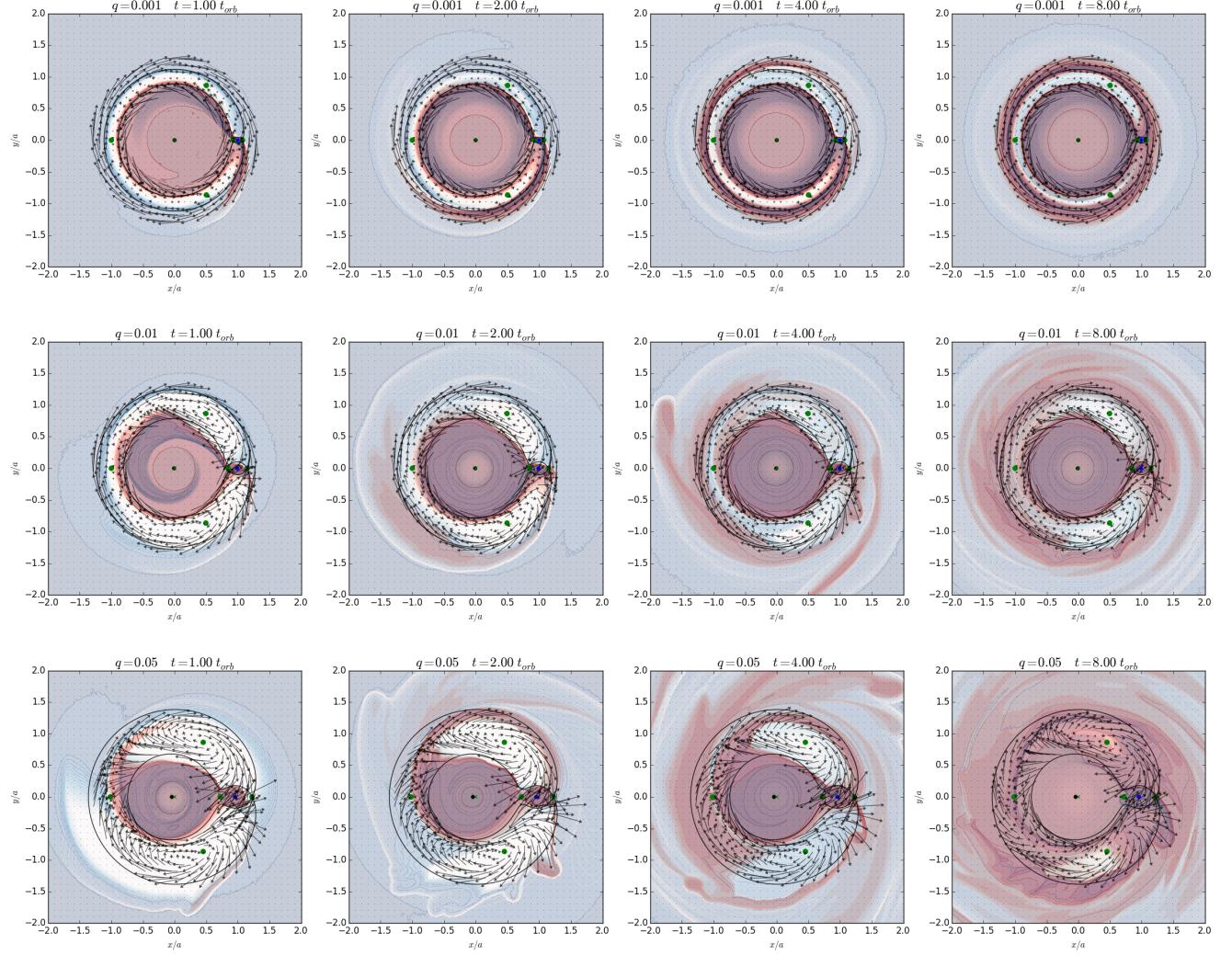


Figure 3.10: Evolution of two passive scalars for $q = 0.001$, $q = 0.01$, and $q = 0.05$ (rows) at different times (columns) during gap opening as well as velocity vectors showing fluid motion in the co-rotating frame. The red scalar starts inside of the critical ZVC (overlaid, black curve) and the blue scalar starts outside. The green dots denote the Lagrange points of the binary potential (See Figures 3.1 and 3.2). Published with this article are *movies* of the above three simulations with the red and blue passive scalars plotted on different panels and with the same initialization as Figure 3.3 (these movies and the corresponding movies of surface density evolution can also be found at <http://user.astro.columbia.edu/~dorazio/moviespage>).

velocity of the fluid is larger in the presence of larger binary forces. The result is a greater deviation of the gas from the critical ZVC curve, some of which begins to peel off into the outer disc. As the last panel for the $q = 0.01$ case hints, this behaviour is not sustained once the gap is cleared and a steady state ensues.

The third row in Figure 3.10 tracks the fluid motion for $q = 0.05$, which is above the mass ratio for which stable L4/L5 orbits exist. We now see that the red gas immediately flows out along L3 and L2. The gas leaving L3 connects back to the orbital flow around L4 while the gas leaving L2 is flung into the disc. A striking feature is the large velocity vectors pointing outward from between L2 and L5 into the disc rather than pointing back along the horseshoe orbit. Because of the vigorous ejection of fluid from the L5 point, an asymmetry builds between the L4 and L5 regions.

In addition to the ejection of co-orbital particles for $q \gtrsim 0.04$, we observe the beginnings of a second stream connecting to the primary through the L3 equilibrium point, the lowest point in the binary potential after L1 and L2. The bottom panels of Figure 3.9 show us that, for near equal mass binaries, two accretion streams feed the binary. Previous work ([Shi et al. 2012](#); [D’Orazio et al. 2013](#)) has argued that streams crashing into the surrounding CBD causes lopsided cavity growth. Specifically, ([D’Orazio et al. 2013](#)) conducted an experiment where a disc around an equal mass binary is simulated with one binary component placed artificially at the origin of coordinates. In this experiment, only one stream is generated and lopsided growth is inhibited. Hence, though it may not be necessary for lopsided growth, the generation of a second, strong stream for $q \gtrsim 0.04$ seems to facilitate such growth.

Because the new stream through L3 feeds the primary, the $q = 0.04$ transition also signifies an increase in the accretion rate of the primary relative to the secondary for $q \gtrsim 0.04$.

Finally, we note that, in order to show the dynamics of gap clearing, the flows depicted in Figure 3.10 are chosen at early times in the disc evolution. While initial conditions are chosen so

that initial transients are minimal, such transients may be present, and one should take this example only as suggestive of the properties of the flow at later times.

3.3.2 Hydrodynamic Parameter Study

Although we identified a clear transition in the disc behaviour near $q = 0.04$, it is natural to ask whether this critical value is universal, or if it depends on disc parameters. To determine this, we repeat our earlier hydrodynamical simulations but for two new values of the coefficient of kinematic viscosity $\nu_0 = 10^{-3}, 10^{-4}$, where $\nu_0 \equiv \nu a_0^{-2} \Omega_{\text{bin}}^{-1}$, and two new values of the Mach number $\mathcal{M} = 10, 30$, corresponding to a factor of 9 variation in disc pressure and temperature. For each of the four new pairings of Mach number and viscosity coefficient, we run four simulations at the binary mass ratios $q = 0.01, 0.025, 0.075, 0.1$.

Note that we increase the outer boundary of the simulation domain to $r_{\text{max}} = 16a$ for the $\nu_0 = 10^{-4}, \mathcal{M} = 10$ simulation, because in these simulations, spiral density waves have longer wavelengths and are less damped by viscosity, allowing them to reach the outer boundary of the simulation, with the results therefore potentially depending on the outer boundary condition. For all other simulations we choose $r_{\text{max}} = 8a$. Because we are using a log grid, this corresponds to only a minimal change in resolution. We have run higher resolution simulations for the $\nu_0 = 10^{-4}, \mathcal{M} = 10$ and $\nu_0 = 10^{-3}, \mathcal{M} = 30$ cases with outer boundary at $r_{\text{max}} = 100a$ finding minimal changes in the surface density distributions, accretion rates, and disc lopsidedness presented below.

From our set of 8 (fiducial) + 16 (parameter study) = 24 CBD simulations (Table 3.1), we use the following two diagnostics to track the onset of the $q \sim 0.04$ cavity transition.

- 1. Amplitude of accretion-rate variability:** To emphasize the change from steady-state to strongly-fluctuating solutions across the CBD phase transition, we compute the standard deviation from the mean accretion rate measured separately onto the individual BHs, as well

as the total accretion rate onto both BHs,

$$\begin{aligned}\delta\dot{M}^n &= \sqrt{\frac{1}{N-1} \sum_j^N \left(\dot{M}_j^n - \langle \dot{M}^n \rangle_t \right)^2} \\ \dot{M}^n(t) &= \sum_k \frac{3}{2} \Sigma_k(t) (r_k^n)^{-2} \nu \quad k \mid r_k^n \leq r_{\text{acc}}^n.\end{aligned}\quad (3.8)$$

In the top equation j denotes the j^{th} timestep out of N total timesteps and $\langle \cdot \rangle_t$ is the average over the entire time interval. In the bottom equation, k denotes the k^{th} cell within the sink radius and the summation is over all cells within the sink radius. In both equations n denotes the n^{th} binary component.

2. Disc Lopsidedness: To measure the lopsidedness of the cavity we measure the quantity

$$\epsilon = \left\langle \left\langle \frac{\left| \langle \Sigma v_r e^{i\phi} \rangle_\phi \right|}{\langle \Sigma v_\phi \rangle_\phi} \right\rangle_r \right\rangle_t, \quad (3.9)$$

where $\langle \cdot \rangle_\phi$ denotes an azimuthal average, $\langle \cdot \rangle_r$ denotes a radial average from $r = a$ to the edge of the simulation domain, $\langle \cdot \rangle_t$ denotes a time average over an integer number of binary orbits, and $|\cdot|$ is the magnitude of a complex number. The only non-zero contributions to Eq. (3.9) are from the components of Σv_r which are proportional to $A \cos(\phi - \phi_A) + B \sin(\phi - \phi_B)$, for arbitrary constants A , B , ϕ_A , and ϕ_B . Hence ϵ measures the lopsidedness of the disc. Note that ϵ is often referred to as the disc eccentricity (e.g., MacFadyen & Milosavljević 2008; Farris et al. 2014).

The above diagnostics are time-averaged over the final 100 orbits of the simulation, for which a quasi-steady state has been achieved.

Figure 3.11 plots the accretion rate variability onto the secondary $\delta\dot{M}^s$ (left), primary $\delta\dot{M}^p$

$q \equiv M_s/M_p$	\mathcal{M}	ν
0.01, 0.025, 0.075, 0.1	10	10^{-3}
0.01, 0.025, 0.075, 0.1	10	10^{-4}
0.001, 0.01, 0.025, 0.03, 0.05, 0.075, 0.1, 1.0	20	2.5×10^{-5}
0.01, 0.025, 0.075, 0.1	30	10^{-3}
0.01, 0.025, 0.075, 0.1	30	10^{-4}

Table 3.1: Parameters for the simulations run in this study. The columns, from left to right, are the binary mass ratio, the orbital Mach number, and the coefficient of kinematic viscosity.

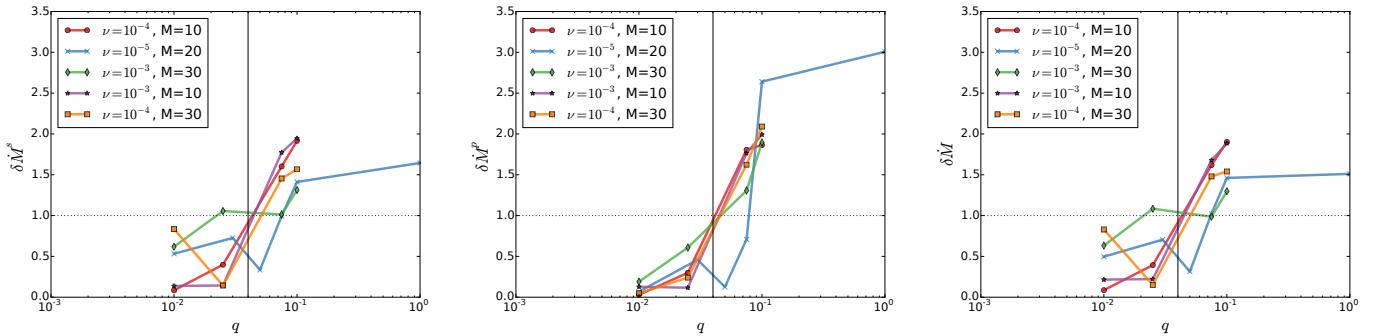


Figure 3.11: The standard deviation of \dot{M} onto the secondary (left), primary (center), and both (right) BHs, computed over the final 100 orbits of each simulation. The vertical solid line is drawn at the $q = 0.04$ transition, the dashed horizontal line is the mean of the standard deviation of each data set.

(central), and both \dot{M} (right) BHs, as a function of binary mass ratio for each set of disc parameters. Because there is a scatter in the magnitude of accretion variability across the range of different disc parameters (notably, the magnitude of variability varies with viscosity), we normalize each set of mass ratios for a given set of disc parameters by the average $\delta\dot{M}$ over all q (excluding the extreme values $q = 0.001$ and $q = 1$ for the fiducial case). First notice that the left and right panels of Figure 3.11, $\delta\dot{M}^s$ and $\delta\dot{M}$, are very similar. This is because over the range of mass ratios probed, the accretion rate onto the secondary, and any resulting variation, is larger than that onto the primary (in agreement with Farris et al. 2014). Both of these panels show a clear trend in increasing accretion variability across the $q = 0.04$ transition, for all sets of disc parameters.

While observationally interesting, relative accretion rates measured onto the secondary (and hence the total accretion rate onto both black holes) may be a less robust diagnostic for the $q = 0.04$ CBD transition than accretion rates measured onto the primary. The transition depicted by the central panel of Figure 3.11 - for the standard deviation of accretion rates onto the primary - looks sharper because the accretion rate variability settles down to small values below the $q = 0.04$ transition. The same measurements, for accretion rate variability onto the secondary (and hence the total accretion rate variability), do not settle to as tight a range of small values. This effect may be tied to how accretion rates are measured in the simulation; for small mass ratios, accretion is measured in the small region within the Hill sphere of the secondary, and rates measured in the simulations may be sensitive to this sink size and the resolution. For example, for fiducial disc parameters, the $q = 0.001$ and $q = 0.025$ cases exhibit anomalously large variability onto the secondary, which is omitted from Figure 3.11.^b Future work should examine the dependence of accretion rates on sink size, accretion prescription, and resolution around the secondary. These considerations are, however, less important for the primary.

The central panel of Figure 3.11, displaying accretion variability onto the primary, exhibits the most striking depiction of the transition. This panel shows a sharp increase in variability amplitude across the $q = 0.04$ transition for all disc parameters. An increased magnitude of accretion variability onto the primary is in agreement with the earlier observation that the stream feeding the primary BH becomes significant at the $q = 0.04$ transition. Because the second stream could be necessary for generating the lopsided cavity, $\delta\dot{M}^p$ acts as an excellent diagnostic for the $q = 0.04$ transition.

[MacFadyen & Milosavljević \(2008\)](#); [Shi et al. \(2012\)](#); [Noble et al. \(2012\)](#); [D’Orazio et al. \(2013\)](#); [Farris et al. \(2014, 2015b,a\)](#); [Shi & Krolik \(2015\)](#) have shown that an equal-mass binary can create a lopsided CBD (referred to as eccentric in [MacFadyen & Milosavljević \(2008\)](#) and

^bFor fiducial disc parameters the $q = 0.025$ disc also exhibits anomalously large accretion variability onto the primary and should be investigated in future studies.

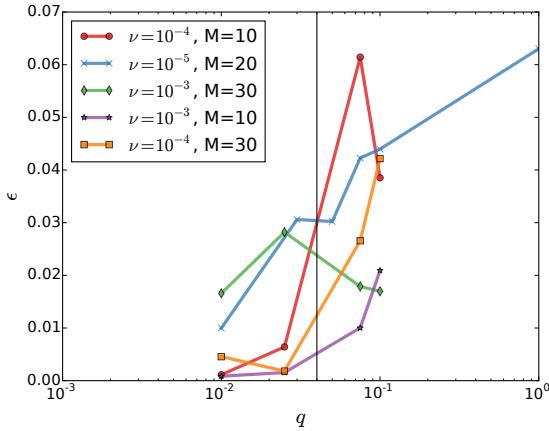


Figure 3.12: The lopsidedness, Eq. (3.9), of the circumbinary disc, spatially averaged over the region outside of the binary orbit, $r > a$, and time averaged over the final 100 orbits of each simulation. The vertical solid line denotes $q = 0.04$.

Farris et al. (2014)). Farris et al. (2014) measured the disc lopsidedness for a range of binary mass ratios, finding a sharp transition to lopsided discs for $q \gtrsim 0.05$ (see Figures 5 and 6 of Farris et al. 2014). Figure 3.12 plots the time averaged disc lopsidedness, Eq. (3.9), measured at radii $r > a$, as a function of binary mass ratio. We find a sharp increase in disc lopsidedness near $q = 0.04$ for all disc viscosities and pressures except for the high viscosity $\nu_0 = 10^{-3}$, low pressure case $\mathcal{M} = 30$.

We display the surface density of the $q = 0.01$, $\nu_0 = 10^{-3}$, $\mathcal{M} = 30$ disc in the left panel of Figure 3.13. In this case, strong shocks form intermittently as the spiral arm connecting to the secondary propagates outwards into the disc. These shocks lead to the formation of a slightly lopsided disc even for the $q = 0.01$ case. We note that this asymmetry could be caused by the difference in instability times derived for the viscous R3Bp in the Appendix (see Figure 3.15). However, we leave an investigation for future work

In addition to the possible viscous instability explained in the Appendix, there are other mechanisms which may make the CBD lopsided and induce variable accretion. Kley & Dirksen (2006) find that a spatial gap asymmetry can be induced even for mass ratios as low as $q = 0.003$. This

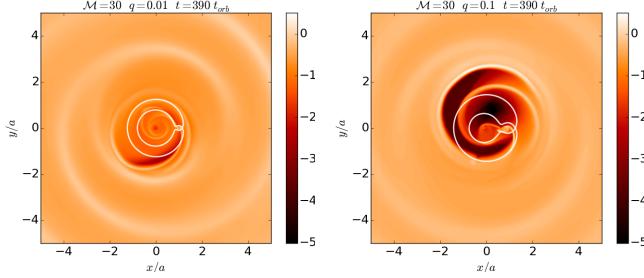


Figure 3.13: Snapshots of the surface density distribution (shown in units of the unperturbed value, with a logarithmic colour scheme) from hydrodynamical simulations for a disc with $\mathcal{M} = 30$, and constant coefficient of kinematic viscosity $\nu = 10^{-3}a_0^2\Omega_{\text{bin}}$. For these high-viscosity, low-pressure simulations, we find an asymmetric-disc shape even for small mass ratio binaries (left). The transition to a time-dependent, lopsided cavity at $q \sim 0.04$ still takes over for larger mass ratio binaries (right).

is only found to occur when viscosity is low enough for a deep gap to be cleared. The explanation of Kley & Dirksen (2006) uses the work of Lubow (1991a,b); a deep gap mitigates the eccentric co-rotation resonances in the disc which act to damp disc lopsidedness while eccentric Lindblad resonances (eLRs) act to grow disc lopsidedness. This asymmetry growth due to eLRs is not likely the mechanism causing the $q = 0.04$ phase transition. Figures 3.11 and 3.12 show that the transition to a lopsided cavity occurs around $q = 0.04$ even for large viscosities and that there is no trend of a decreasing critical mass ratio with decreasing viscosity as the Kley & Dirksen (2006) mechanism would predict. Additionally, Shi et al. (2012) show that the growth timescale of asymmetry for the $q = 1$ binary does not match the growth timescale expected from the argument of Kley & Dirksen (2006). Future work should measure the growth rate of lopsidedness for multiple mass ratios in order to disentangle the mechanism by which the low mass ratio systems of Kley & Dirksen (2006) become asymmetric (presumably eLRs) and the mechanism which causes the phase transition studied here (presumably orbital instability).

Regardless of hydrodynamic processes which may act to grow a lopsided, time-variable disc below the $q = 0.04$ transition, the present study shows that a robust transition in CBD structure

dominates for $q \gtrsim 0.04$ over a wide range of hydrodynamic disc parameters.

3.4 Discussion and Summary

Circumbinary accretion discs exhibit a phase transition above a binary mass ratio of $q \simeq 0.04$. The transition signifies a loss of spatial and time-translation symmetry in the CBD. The density structure changes from an annular low-density gap near the orbit of the secondary ($q \lesssim 0.04$), to a lopsided cavity ($q \gtrsim 0.04$). It also marks a transition from steady-state to strongly-fluctuating behaviour. We conjecture that this transition is closely tied to the loss of stable orbits in the binary co-orbital region of the restricted three-body problem (R3Bp) for $q > 0.04$. When stable co-orbital orbits connect the outer, circumbinary disc to the inner, circumprimary disc, steady-state gapped solutions are realized. When no such stable co-orbital orbits exist, accretion streams violently impact the inner mini-discs and the outer circumbinary disc, leading to fluctuating lopsided cavity solutions. We employ the R3Bp as well as 2D viscous hydrodynamical simulations to investigate the CBD transition.

We show that the change in density morphology, from annular gaps to cavities, in circumbinary accretion discs can be largely captured by the spatial restriction of test particles imposed by conservation of the Jacobi constant in the R3Bp. To quantify the limitations of the R3Bp and to more closely compare to hydrodynamical simulations, we extend the R3Bp analysis to include the effects of pressure and viscosity. To estimate the effects of pressure in the disc, we compare the Jacobi constant with the closely related Bernoulli constant and derive a maximum disc aspect ratio (minimum Mach number) necessary for pressure forces to overcome the gravitational barrier of the binary. For binaries with $q > 0.04$ this only occurs for discs which are no longer thin.

To study the effects of viscosity on the R3Bp and investigate the relation of the CBD phase transition to orbit stability, we add a viscous force to the R3Bp equations and perform a linear-

stability analysis on orbits around the L4 and L5 equilibrium points in the presence of this force. We find that the viscous R3Bp shares a similar orbital stability transition which occurs very near the classical stability mass ratio of $q = 0.04$ over multiple magnitudes of viscosity (see the Appendix), consistent with our hydrodynamical parameter study. We also find that viscous forces break a symmetry between the stability of the leading and trailing triangular Lagrange points (L4/L5) of the classical R3Bp. This may be related to the growth of asymmetry in CBDs, though we save such a study for future work.

The effects of both viscosity and pressure on the CBD transition are studied via 2D viscous hydrodynamical simulations. These show that changing the disc viscosity by a factor of 40 and the disc pressure by a factor of 9 leaves the critical mass ratio largely unaffected. The hydrodynamical simulations also provides further evidence for a mechanism by which orbital instability seeds the transition. For $q \lesssim 0.04$ disc particles in the binary co-orbital region stably oscillate on horseshoe-like orbits, while for $q \gtrsim 0.04$, particles are flung out of the co-orbital region into the outer disc (Figure 3.10).

We note that, in addition to loss of L4/L5 orbital stability, there is another dynamical transition which occurs near $q = 0.04$ in the R3Bp. With the goal of determining the truncation of CBDs, Rudak & Paczynski (1981) studied the intersection and radial stability of test particle orbits around a circular binary with arbitrary mass ratio. They posit that the innermost stable orbits of a CBD are set either by orbit intersection, or instability to radial perturbations of a Keplerian orbit around the binary. They find that, for binaries with $q \lesssim 0.01$, orbit intersection truncates the CBD, for $0.01 \leq q \leq 0.05$ the inner edge of the disc becomes marginally unstable before orbit intersection becomes important, and for $q > 0.05$, violent instability is responsible for disc truncation. In this paper, we have shown evidence that links the L4/L5 orbital instability to the observed CBD transition. The additional radial instability found by Rudak & Paczynski (1981) in the particle limit, near the inner edge of the CBD, also occurs near $q=0.05$. However, our hydrodynamical

simulations do not show strong fluctuations in the inner edge of the CBD for $q \gtrsim 0.05$. Rather, the variability in the accretion rates we observe appears to track the unstable behavior of the gas near L4/L5 (see Figure 3.10 and corresponding movies).

In accordance with the simplicity of the circular R3Bp and our numerical simulations, we have restricted our analysis to systems consisting of isothermal and adiabatic thin discs surrounding a binary on a fixed circular orbit. A binary on an eccentric orbit exhibits loss of co-orbital stability at mass ratios smaller than $q = 0.04$ of the circular case (Kovács 2013). Supersonic gas dynamics in the vicinity of the binary will likely invalidate the assumption of an isothermal gas disc. A massive disc will move the binary; for mass ratios above the transition, the binary may become eccentric (Cuadra et al. 2009; Roedig et al. 2011, 2012a) and migrate differently than the often assumed Type II rate (*e.g.* Haiman et al. 2009; Dotti et al. 2015). Future work should address the change of binary orbital parameters due to the change in CBD structure. Other important physics not included in this work may also impact the $q = 0.04$ transition. Including the vertical disc dimension, magnetic fields and the magneto-rotational instability, and radiative feedback from accretion could all affect the onset and behaviour of the transition described here.

Future work on CBD structure may find insight into more general binary+disc systems from extensions of the circular R3Bp for a binary with time-dependent separation (Schnittman 2010), a mis-aligned disc (Erwin & Sparke 1999), or non-zero eccentricity (Pichardo et al. 2005).

Within the limitations of this study, we have identified a dramatic transition occurring in circumbinary discs and offered an intriguingly simple origin; further work should clarify whether it survives the additional physical effects mentioned above.

The $q = 0.04$ CBD transition is relevant to MBHB+disc systems. Accreting MBHBs below the critical mass ratio will prove difficult to detect in time-domain electromagnetic surveys due to accretion variability alone. For such low mass, steadily-accreting MBHBs, other mechanism could cause variability, such as a precessing jet, disc instabilities, or relativistic Doppler boost in the case

of compact binaries (D’Orazio et al. 2015b). Because of the drastic change in disc lopsidedness across the $q = 0.04$ transition, there may be an equally drastic change in the binary orbital parameters. This would result in a connection of the binary mass ratio with binary eccentricity and migration rate. This could affect gravitational wave detection rates as well as waveforms.

The $q = 0.04$ CBD transition is also relevant for proto-planetary systems. The formation of planets around a binary system may progress differently for a proto-planetary disc around a brown dwarf and main sequence pair (*e.g.*, De Rosa et al. 2014; Curé et al. 2015; Hinkley et al. 2015) than for a near equal mass binary. Though theoretically disfavoured (Payne & Lodato 2007), there would be important consequences for the formation of planetary systems around brown dwarfs which contain a large ($\gtrsim M_J$) planet. It will be interesting to look for differences in planetary populations around binaries above and below $q \sim 0.04$ if they are discovered in upcoming searches (*e.g.* Triaud et al. 2013; Ricker et al. 2014).

Acknowledgments

The authors thank Jeffery J. Andrews and Adrian Price-Whelan for useful discussion of the two- and three-body problems, and Ganesh Ravichandran for writing an initial version of a code used for analysis of the hydrodynamical parameter study. The authors also thank Roman Rafikov, Jeno Sokolowski, and Scott Tremaine for useful discussions. We acknowledge support from a National Science Foundation Graduate Research Fellowship under Grant No. DGE1144155 (DJD) and NASA grant NNX11AE05G (ZH and AM).

3.5 Viscous, restricted three-body problem

We follow [Murray \(1994\)](#) and [Murray & Dermott \(1999\)](#) to add an external viscous force $\mathbf{F}(x, y, \dot{x}, \dot{y})$ to the R3Bp equations of motion. In the rotating frame,

$$\ddot{x} - 2\Omega\dot{y} = \frac{\partial U}{\partial x} + F_x$$

$$\ddot{y} + 2\Omega\dot{x} = \frac{\partial U}{\partial y} + F_y,$$

where Ω is the frequency of the rotating frame and $U(x, y)$ is the Roche potential with sign convention taken to be consistent with [Murray & Dermott \(1999\)](#).

For the external force we use the force associated with the $r - \phi$ component of the viscous stress tensor in a Keplerian flow. In the inertial frame this force (per unit mass) is

$$F_{\hat{\phi}} = \frac{\partial_r \sigma^{\hat{\phi}\hat{\phi}}}{\rho} = -\partial_r [\nu r \partial_r \Omega_{\text{disc}}] = -\frac{9}{4} \nu \frac{v_\phi}{r^2},$$

where we assume that the angular velocity of the flow is Keplerian around the binary mass M , and that the kinematic coefficient of viscosity ν and density ρ are constants. In the last line we relabel $v_\phi \equiv \sqrt{GM/r}$. By transforming to Cartesian coordinates and then moving into the rotating frame we find

$$F_x = -F_{\hat{\phi}} \frac{y}{r}$$

$$F_y = F_{\hat{\phi}} \frac{x}{r}.$$

Writing these in terms of the rotating frame coordinates and velocities,

$$F_x = -\frac{9}{4} \frac{\nu}{x^2+y^2} (\dot{x} - y\Omega)$$

$$F_y = -\frac{9}{4} \frac{\nu}{x^2+y^2} (\dot{y} + x\Omega),$$

which follows from our choice of azimuthal velocity profile $v_\phi(r)$ in the inertial frame of the binary so that $\dot{x} = -(v_\phi - r\Omega)y/r$ and $\dot{y} = (v_\phi - r\Omega)x/r$.

As with pressure forces (§3.2.3.1), viscous forces also destroy conservation of the Jacobi constant. C_J changes at a rate

$$\dot{C}_J = \frac{9}{2} \frac{\nu}{x^2+y^2} [\dot{x}^2 + \dot{y}^2 + \Omega(x\dot{y} - y\dot{x})]. \quad (3.10)$$

Comparing the change in C_J due to viscosity to the change due to pressure over N_{orb} orbital times gives,

$$|\Delta C_J|_\nu = \alpha N_{\text{orb}} \frac{5}{9\pi} |\Delta C_J|_P, \quad (3.11)$$

where we have used Eq. (3.3) and the first of Eqs. (3.4) for the RHS and assumed a constant coefficient of kinematic viscosity $\nu = \alpha/\mathcal{M}^2 a^2 \Omega$ with $\alpha < 1$. In the case that N_{orb} is a gap opening time, which we approximate as the viscous time across a gap of size equal to the disc scale height, we find

$$|\Delta C_J|_\nu = \frac{5}{27\pi^2} |\Delta C_J|_P \sim 0.02 |\Delta C_J|_P. \quad (3.12)$$

Hence for short timescales relevant for local density distribution such as gaps, the pressure forces dominate in the decay of the Jacobi constant. For longer timescales $N_{\text{orb}} \gg 1/\alpha$ it is viscosity which dominates the deviation from the purely gravitational problem.

To investigate the consequences of a viscous force for orbital stability, we perform a linear-

stability analysis for particles perturbed from the analogue of the L4/L5 points in the viscous R3Bp (3.10), with external force (3.10). The location of the new equilibrium points in the presence of the viscous force are found by setting $\ddot{x} = \ddot{y} = \dot{x} = \dot{y} = 0$ in the equations of motion (3.10) and solving for the coordinates (x^\dagger, y^\dagger) of the new equilibrium points. Murray (1994) finds solutions by Taylor expanding the equations of motion around the $\mathbf{F} = 0$ classical equilibrium points (x_0, y_0) and solving for the deviation (\bar{x}, \bar{y}) from (x_0, y_0) . To linear order in \bar{x}, \bar{y} the new equilibrium points are,

$$x^\dagger = x_0 - \frac{(1+q)}{q} \frac{F_x^*}{3} \pm \frac{(1+q)}{q} \frac{F_y^*}{3\sqrt{3}}$$

$$y^\dagger = y_0 \pm \frac{(1+q)}{q} \frac{F_x^*}{3\sqrt{3}} - \frac{(1+q)}{q} \frac{F_y^*}{9},$$

where + and – denote L4 and L5, and * denotes evaluation at the classical Lagrange point ($x = x_0, y = y_0, \dot{x} = 0, \dot{y} = 0$).

Next assume a solution for the motion of a test particle perturbed from L4/L5 to be of the form,

$$x(t) = x^\dagger + X_0 e^{\lambda t} \quad y(t) = y^\dagger + Y_0 e^{\lambda t}. \quad (3.13)$$

Substituting this ansatz into the equations of motion (3.10) and keeping terms to linear order in the displacement, $X_0 e^{\lambda t}$ and $Y_0 e^{\lambda t}$, Murray (1994) finds a set of simultaneous linear equations for the displacement from equilibrium with characteristic equation for the eigenvalues λ ,

$$\lambda^4 + a_3 \lambda^3 + (1 + a_2) \lambda^2 + a_1 \lambda + \frac{27}{4} \frac{q}{(1+q)^2} + a_0 = 0, \quad (3.14)$$

which assumes that the viscous force is small by neglecting terms of $\mathcal{O}(\nu^2)$. The a_i coefficients can be written in terms of derivatives of the external force evaluated at the new equilibrium points.

In the case of our viscous force

$$\begin{aligned} a_0 &= \frac{27}{4} \nu \frac{\pm\sqrt{3}(x^\dagger)^2 + 2x^\dagger y^\dagger \mp\sqrt{3}(y^\dagger)^2}{2[(x^\dagger)^2 + (y^\dagger)^2]^2} \\ a_1 &= \frac{27}{4} \frac{\nu}{(x^\dagger)^2 + (y^\dagger)^2} \\ a_2 &= 0 \\ a_3 &= -\frac{9}{2} \frac{\nu}{(x^\dagger)^2 + (y^\dagger)^2}, \end{aligned}$$

where \pm and \mp refer to the L4 and L5 points respectively, x^\dagger, y^\dagger are given by (3.13), and we have dropped terms of order $q^2/(1+q)^2$, except in the constant term (second to last term on the LHS of Eq. (3.14)) which gives the stability criterion of $q < 0.04$ in the classical R3Bp. Solving the depressed quartic for λ we find four complex solutions which we plot in Figure 3.14 for multiple values of the viscosity. We make three observations,

- For finite viscosity and $q > 0$, there are no linearly stable orbits, only quasi-stable orbits, which are formally unstable but have long instability times (small $\text{Re}[\lambda]$). This can be seen from the solid lines in the top panel of Figure 3.14. The non-viscous case (black lines) has $\text{Re}[\lambda] = 0$ for $q < 0.04$, the classical result. For the higher-viscosity cases (blue and red lines), $|\text{Re}[\lambda]|$ is large and Eqs. (3.13) show that orbits will decay or blow up.
- If we define the critical mass ratio for linear quasi-stability to be the mass ratio where $\text{Re}[\lambda]$ has the largest derivative (hence changing quickly with q from a long to short instability timescale), then the top panel of Figure 3.14 shows that this critical mass ratio becomes smaller for larger viscous forces. The instability transition is explored further in the bottom panel of Figure 3.14 where we plot the instability timescale in units of the orbital time. The critical mass ratio is not very sensitive to the viscosity. The instability timescale is of order the viscous time below $q = 0.04$ and of order an orbital time above $q = 0.04$.
- The symmetry between the L4 and L5 points is broken. Figure 3.15 plots the difference in

maximum $\text{Re}[\lambda]$ between the L4 and the L5 points in units of the inverse binary orbital time. As expected, there is no difference in growth timescale for L4 vs. L5 for the non-viscous case (black line). The viscous (red, blue, and orange lines) cases, however, show a large difference in instability growth rate above $q = 0.04$ and a weaker difference over a viscosity dependent mass ratio range below the $q = 0.04$ transition. For smaller mass ratios, the L5 point is more unstable, and for larger mass ratios, the L4 point is more unstable.

For completeness we include the oscillatory part of the linear solutions ($\text{Im}[\lambda]$) in the top panel of Figure 3.14 (dashed lines). The different oscillation timescales are discussed for the non-viscous R3Bp in (Murray 1994).

We also integrate the full viscous R3Bp in Figure 3.16. Streams feeding the binary components exist even after 50 orbits in the case where viscous forces are allowed to destroy conservation of the Jacobi constant.

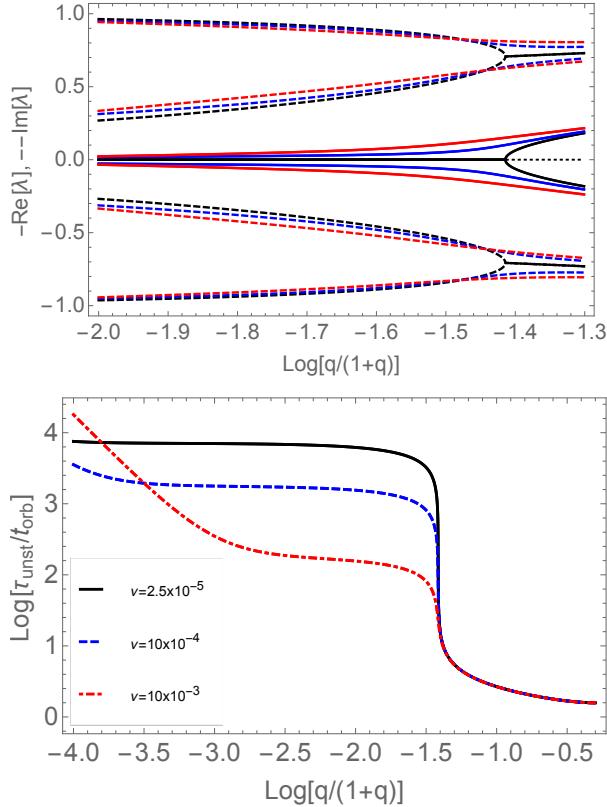


Figure 3.14: Eigenvalues of a linear-stability analysis of the L4 and L5 points with an added viscous force. The top panel shows the real (solid lines) and imaginary (dashed lines) parts for different values of the constant coefficient of kinematic viscosity, $\nu = 0.0$ (black), $\nu = 0.005$ (blue), $\nu = 0.01$ (red). The bottom panel displays instability timescales normalized to the binary orbital time.

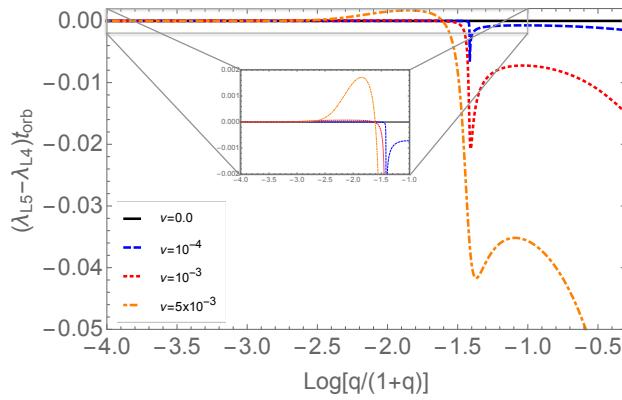


Figure 3.15: The difference in orbital out-spiral of a test-particle perturbed from the L4 and L5 points. λ_{L4} and λ_{L5} refer to the maximum real part of the complex eigenvalues found from Eq. 3.14. The inset zooms in on the region below the $q = 0.04$ transition where weak asymmetry between the L4 and L5 points exists. A value of -0.01 in this plot means that, after 100 binary orbital times, the difference in the final position of a particle perturbed from L4 is an e-fold farther from its starting position than a particle identically perturbed from L5. It is likely that a higher order perturbation in the mass ratio is required to capture the return of $\lambda_{L5} - \lambda_{L4} = 0$ for $q = 1$.

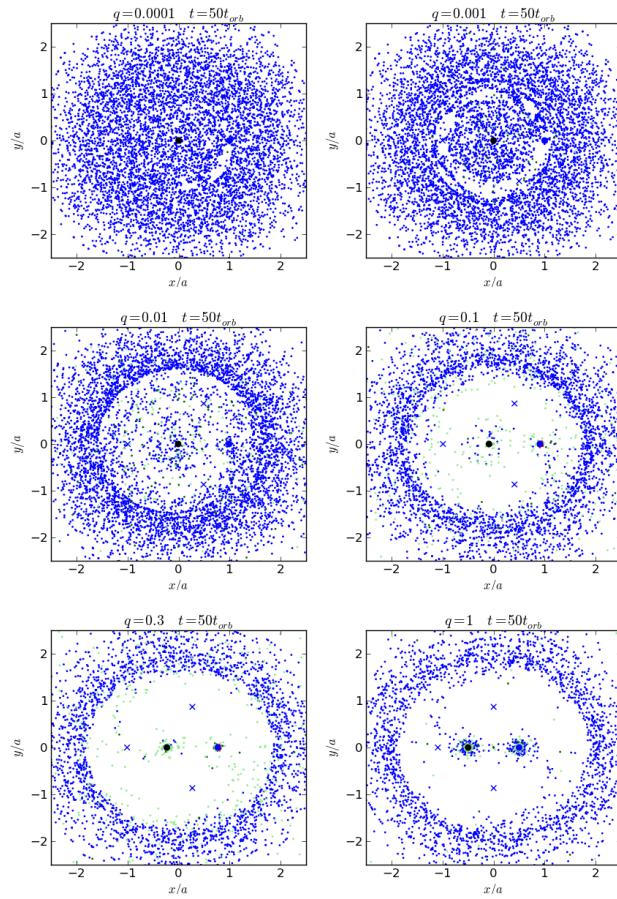


Figure 3.16: Same as Figure 3.5 except for 10^4 particles obeying the viscous R3Bp Eqs. (3.10) and (3.10) with constant $\nu = 0.001a_0^2\Omega_0$.

Chapter 4

A reduced orbital period for the supermassive black hole binary candidate in the quasar PG 1302-102?

4.1 Introduction

?, hereafter G15 recently reported strong optical variability of the quasar PG 1302-102, with an observed period of $t_{\text{obs}} = 5.2 \pm 0.2$ yr. G15 attribute the variability to the orbital motion of a super-massive black hole binary (SMBHB). Broad emission lines in the spectrum of PG 1302 imply a binary mass in the range $M = 10^{8.3-9.4} M_{\odot}$. Assuming that the binary's orbital period t_{bin} equals the rest-frame optical variability period t_{opt} , G15 derive a fiducial binary separation $a \approx (0.0084 \pm 0.0003)\text{pc} \approx (276 \pm 9)R_{\text{S}}$ for $M = 10^{8.5} M_{\odot}$, where $R_{\text{S}} = 2GM/c^2$ is the Schwarzschild radius.

Hydrodynamical simulations of a binary BH embedded in a gaseous accretion disc predict that, depending on the binary mass ratio and the physical parameters of the disc, the strongest periodicity

in the accretion rate on to the BHs may correspond to the motion of gas farther out in the disc, at a few times the binary separation, producing optical variability at several ($\sim 3\text{--}8$) times the binary orbital period. In this article, we discuss this expectation, and show that a reduced binary period, in the case of PG 1302, would have several important implications. Follow up spectroscopy and photometric monitoring can determine the true binary period.

The rich variability structure of the mass accretion rates seen in simulations can be roughly divided into four distinct categories, based on the binary mass ratio $q \equiv M_2/M_1$. For $q \lesssim 0.05$, the disc is steady and the BH accretion rate displays no strong variability (D’Orazio et al. 2013; Farris et al. 2014, D’Orazio et al. in preparation). For $0.05 \lesssim q \lesssim 0.3$, the accretion rate varies periodically on the time-scale t_{bin} , with additional periodicity at $\approx 0.5t_{\text{bin}}$. Binaries with $0.3 \lesssim q \lesssim 0.8$ clear a lopsided central cavity in the disc, causing variability on three time-scales. The dominant period, $(3 - 8)t_{\text{bin}}$ is that of an over dense lump, orbiting at the ridge of the cavity, with additional periodicities at t_{bin} and $\approx 0.5t_{\text{bin}}$ (MacFadyen & Milosavljević 2008; Shi et al. 2012; Noble et al. 2012; Roedig et al. 2012b; D’Orazio et al. 2013; Farris et al. 2014). The dominant period depends on the size of the cavity, and thus on disc parameters, such as temperature and viscosity. Finally, equal-mass ($q = 1$) binaries display variability at the longer lump period and at $\approx 0.5t_{\text{bin}}$.

Here, we consider the identification of the observed variability of PG 1302 with the long, cavity-wall period, and introduce the parameter $\chi \equiv t_{\text{opt}}/t_{\text{bin}}$, denoting the ratio, $3 \lesssim \chi \lesssim 8$, of the observed rest-frame period and the true binary period. The binary separation is then

$$a \approx (94 \pm 3)R_{\text{S}} \left(\frac{\chi}{5}\right)^{-2/3} \left(\frac{M}{10^{8.5}M_{\odot}}\right)^{-2/3}, \quad (4.1)$$

or (0.0029 ± 0.0001) pc for the fiducial choices of χ and M .

In the rest of this article, we first (§4.2) explore several implications of a reduction in the

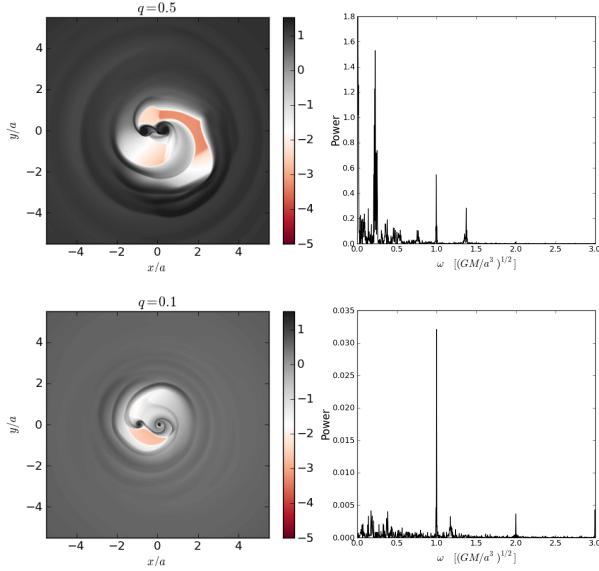


Figure 4.1: Results of 2D hydrodynamical simulations of a binary BH surrounded by a circumbinary accretion disc. The BHs clear out a central cavity and form their own minidisks. *Left-hand panels:* snapshots of the (logarithmic) surface-density of the gas discs, after reaching quasi-steady state, with mass ratios of $q = 0.5$ (top) and $q = 0.1$ (bottom). *Right-hand panels:* corresponding LSPs of the total accretion rate on to the secondary + primary BHs. The discs are locally isothermal with a Mach number of 10 and an alpha viscosity prescription ($\alpha = 0.1$).

binary's orbital period, including the nature of PG 1302's orbital decay and its ability to produce electromagnetic (EM) radiation (§4.2.1), the expected binary fraction of quasars (§4.2.2), and the detectability of gravitational waves (GWs) from PG 1302 by pulsar timing arrays (PTAs). We then (§4.3) propose possible observational tests of the underlying binary BH + circumbinary disc (CBD) model, including variations of broad line widths and centroids correlating with the optical variability (§4.3.1), additional periodic variability at the true t_{bin} caused by relativistic beaming (§4.3.2), signatures in the broad Fe K α lines (§4.3.2), and the existence of distinct secondary peaks in the periodogram (§4.3.3). We briefly summarize our main conclusions in §4.4.

4.2 Implications of a shorter orbital period

In order to demonstrate the possibility of a short orbital period for the PG 1302 binary, we have performed hydrodynamical simulations, following the set-up in our earlier work ([Farris et al. 2014](#)). The hydrodynamical equations are evolved for $\gtrsim 600$ binary orbits, using the two-dimensional code DISCO ([Duffell & MacFadyen 2011](#)), with the BHs moving on fixed circular orbits, surrounded by an isothermal (Mach number = 10) disc, obeying an α -viscosity prescription ($\alpha = 0.1$). The fluid motion around individual BHs is well resolved (with a log grid of 384 radial cells extending to $8a$ and a maximum of 512 azimuthal cells). The two runs discussed below differ only in their BH mass ratio ($q = 0.1$ and $q = 0.5$). A wider range of simulations is needed in the future, to address possibilities such as eccentric ([Roedig et al. 2012b](#)), tilted (?), or retrograde ([Nixon et al. 2011](#)) binary orbits.

The results are illustrated in Fig 4.1. The left-hand panels show snapshots of the surface density, and the right-hand panels show Lomb Scargle periodograms (LSPs) of the total accretion rate measured in the two BH minidisks over 200 binary orbits. The top panels, for $q = 0.5$, show an over dense lump orbiting at the rim of the central cavity, resulting in strong periodicity at the orbital time $\approx 6t_{\text{bin}}$ of the lump. The periodogram shows weaker peaks at t_{bin} and at $\sim 0.6t_{\text{bin}}$. This three-time-scale behaviour, with the longest time-scale dominating, is observed for $0.3 \lesssim q \lesssim 0.8$. The location of the highest frequency peak is closer to $0.5t_{\text{bin}}$ near the low end of this range ($q \sim 0.3$), and also has a weak dependence on disc temperature and viscosity which must be quantified in future work. The bottom panels, for $q = 0.1$, show no orbiting lump and exhibit accretion rate periodicity only at t_{bin} and $0.5t_{\text{bin}}$. This behaviour is found in the range $0.05 \lesssim q \lesssim 0.3$.

[Farris et al. \(2014\)](#) have shown that for unequal-mass binaries, accretion occurs preferentially on to the secondary BH, with the ratio of accretion rates as skewed as $\dot{M}_2/\dot{M}_1 \approx 10 - 20$ in the range $0.02 \lesssim q \lesssim 0.1$. Over long time-scales, this would drive the binary to more equal masses,

suggesting that mass ratios of $0.3 \lesssim q \lesssim 0.8$ may be common. Near-equal mass binaries are also preferred in cosmological models of the population of merging SMBHs ([Volonteri et al. 2003b](#)).

Although there are large uncertainties in how accretion rate fluctuations turn into luminosity variations, we do expect the simulated accretion rate variations to lead to optical luminosity variations for PG 1302. The luminosity will follow local accretion rate fluctuations when the longer of the thermal or photon diffusion time-scale is much shorter than the accretion rate fluctuation time-scale ($\sim t_{\text{bin}}$). This is indeed the case where accretion modulations occur in our simulations, at the minidisc edges. Furthermore, optical emission is generated at the minidisces edges. We compute thin disc spectra for the CBD and minidisces. For the preferred mass range of PG 1302, near unity mass ratios, and the expected range of χ , the dominant optical component of the spectrum is generated by the low-energy tail of the blackbody emission from the outer edges of the minidisces, as well as (steady) emission from the inner regions of the CBD. Disc+binary simulations by [Farris et al. \(2015b\)](#), which self-consistently compute the local effective disc temperature not assuming a steady state, find results in agreement with our analytic reasoning: luminosity variations track the accretion-rate fluctuations, except at low frequencies where the quasi-steady CBD dominates.

The above lines of evidence motivate us to examine the possibility that the apparent 5.2 yr period in PG 1302 is the (redshifted) lump period, and assess the implications.

4.2.1 Binary-Disc Decoupling

A shorter orbital period would place the binary at a later stage of its orbital decay. A critical point during the orbital decay is the decoupling of the binary from the CBD, and it is important to know whether the binary is past this point. Here, we consider the decoupling radius for which the GW decay time-scale becomes shorter than the decay time-scale due to gaseous torques (so-called secondary-dominated Type II migration; [Syer & Clarke 1995](#)), outpacing the CBD. We use simple 1D models of the binary + disc system ([Haiman et al. 2009](#), hereafter HKM09) to calculate

the separation r_{GW} at which decoupling occurs for a circular binary with mass ratio $q = 0.3$. We assume an α -viscosity $\nu = \alpha P_{\text{gas}}(\rho\Omega)^{-1}$, with gas pressure P_{gas} , density ρ , and disc angular velocity Ω . All other disc parameters are assumed to have the fiducial values given in HKM09. In Fig. 4.2, we plot the ratio a/r_{GW} as a function of the total binary mass M , with the binary separation from equation (4.1), for the range of masses in G15, and for three values of χ covering the range suggested by the hydrodynamical simulations.

Interpreting the observed variability in PG 1302 with t_{bin} , as may be justified for $q \lesssim 0.3$, it is unclear whether or not the binary has entered the GW dominated regime and decoupled from the disc. The binary would still be coupled to the disc for $M < 10^{8.7} M_{\odot}$ (the majority of the range inferred from the broad lines by G15), but GW-driven and decoupled if $M > 10^{8.7} M_{\odot}$. However, for the shorter binary periods $3 \lesssim \chi \lesssim 8$, justified for $0.3 \lesssim q \lesssim 0.8$, we find that the inferred smaller binary separation would place the binary well past decoupling. For $q > 0.3$ and $\alpha < 0.3$, the binary in PG 1302 is plunged even deeper into the GW-dominated regime.

Because the binary outpaces the disc, it has been argued that the post-decoupling BHs may be “starved” and thus dim (Milosavljević & Phinney 2005; Shapiro 2010; ?). Recent simulations (Noble et al. 2012; Farris et al. 2015a) show that high levels of accretion can persist well past the decoupling phase, delivering gas to the binary efficiently until much closer to coalescence. These simulations also exhibit the lopsided cavity which generates the χt_{bin} variability considered here. Identification of the variability in PG 1302 with the cavity wall lump period would constitute the (to our knowledge, first-ever) detection of an SMBHB which is undergoing GW dominated inspiral, yet producing bright emission, near the Eddington limit.^a

^aNote that in the precessing binary model for OJ287 (??), the orbital period is 12.2 yr, the primary is very massive ($\sim 1.8 \times 10^{10} M_{\odot}$), but the secondary is light ($\sim 1.4 \times 10^8 M_{\odot}$). The latter reduces the efficiency of GWs, but increases the impact of a gas disc; as a result, the OJ287 binary is gas-driven, well before decoupling.

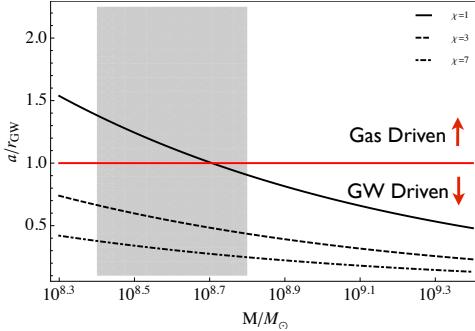


Figure 4.2: The ratio of binary separation a to the decoupling radius r_{GW} , for three different values of the ratio between the rest frame optical period and the true binary period, $\chi = t_{\text{opt}}/t_{\text{bin}}$. The shaded region marks the binary mass range inferred from the widths of broad lines measured by G15. For $\chi > 3$, the PG 1302 binary is past decoupling, for any choice of mass M .

4.2.2 Binary Fraction among Quasars

A shorter binary orbit also reduces the expected number of detectable SMBHBs in quasars, because gas driven binaries are expected to spend less time at smaller separations. A simple estimate of the fraction of quasars that would harbour binaries with an orbital period t_{bin} can be obtained from the residence time $t_{\text{res}} = a/\dot{a}$ at each separation a , and the lifetime of bright ($L_Q/L_{\text{edd}} \gtrsim 0.3$) quasars, $t_Q \sim 10^8$ yr (?). Assuming that a fraction f_{bin} of all quasars are triggered by coalescing SMBHBs (e.g ? and references therein), it follows that among bright quasars, the fraction with orbital period t_{bin} is $f_{\text{var}} = f_{\text{bin}} f_{\text{duty}}$, where $f_{\text{duty}} = t_{\text{res}}(t_{\text{bin}})/t_Q$ is the fraction of the bright quasar phase that a typical binary quasar spends at the orbital period t_{bin} .

We use the binary+disc models of HKM09 to predict the residence times of binaries. Prior to decoupling, t_{res} is determined by the binary's interaction with the gas disc. For the masses and separations relevant for PG 1302, the disc would be radiation pressure dominated, yielding a relatively shallow power-law dependence $t_{\text{res}} \propto t_{\text{bin}}^\beta$ with $0.5 \lesssim \beta \lesssim 1.5$. These scalings depend on the poorly understood physical model of the disc and its coupling to the binary. Past decoupling, the residence time is precisely known, since it is determined by the strength of GWs. The dependence is much steeper, $t_{\text{res}} \propto t_{\text{bin}}^{8/3} \propto \chi^{-8/3}$. For reference, a binary with $M = 10^{8.5} M_\odot$

and $t_{\text{bin}} = 4\text{yr}$ would be in the disc-driven stage and would have $t_{\text{res}} \approx 10^6$ yr, yielding a large $f_{\text{duty}} \approx 10^{-2}$.

The expected f_{var} can be compared with the number of periodic candidates uncovered in CRTS (???).^b There are $\approx 114,000$ quasars in the CRTS sample with luminosity higher than PG 1302, ≈ 6 of which are SMBHB candidates with period $t_{\text{obs}} \lesssim 5\text{yr}$ (Graham, private communication), amounting to an observed fraction of $f_{\text{var}}^{\text{obs}} = 5 \times 10^{-5}$. Fig. 4.3 illustrates combinations of M , χ , and f_{bin} , for which we expect 1 (light grey regions) or 10 (dark grey) candidates in the CRTS quasar sample with periods ≤ 5.2 yr. Each shaded region is bounded by the assumed fraction of quasars related to SMBHBs at all, $f_{\text{bin}} = 0.01$ (left) to $f_{\text{bin}} = 1$ (right).

If the observed period of PG 1302 is assumed to be the binary orbital period ($\chi = 1$), then the rarity of the binary candidates in CRTS require $f_{\text{bin}} < 0.14$ (< 0.19) at $q = 0.3$ ($q = 1.0$), even at the most extreme mass of $M = 10^{9.4} M_{\odot}$. Taking the G15 fiducial mass of $M = 10^{8.5} M_{\odot}$, these fractions must be as low as $f_{\text{bin}} < 0.006$. These low values would be surprising, as a large fraction of quasars are commonly believed to be triggered by mergers. This association is based on various pieces of observational evidence, as well as on the success of merger-based quasar population models to reproduce many properties of the observed quasar population (e.g. ?). If, instead, the observed period of PG 1302 is due to the 3 – 8 times longer lump-periodicity, then the SMBHB fraction and the inferred binary mass of PG 1302-12 come into wider agreement with the expectation that $f_{\text{bin}} = O(1)$, e.g. allowing $f_{\text{bin}} \sim 0.3$ with $q = 0.3$ and $M = 10^{8.5} M_{\odot}$.

4.2.3 Detectability of GWs

A reduced orbital period increases the frequency and amplitude of GWs emitted by a binary, and it is interesting to ask whether PG 1302 may be detectable by present or future PTAs. The

^bSince the expected $f_{\text{var}}(M, t_{\text{bin}})$ declines steeply with increasing M and decreasing t_{bin} , it is a good proxy for the fraction of quasars with period t_{bin} or less, and BH mass M or higher (or equivalently luminosity L or higher, further assuming a monotonic relation between L and M).

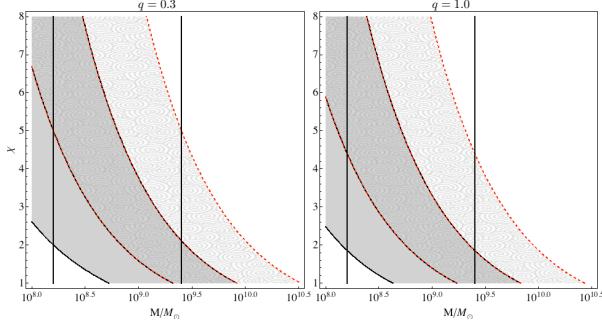


Figure 4.3: Combinations of total binary mass M and $\chi = t_{\text{opt}}/t_{\text{bin}}$ for which the predicted binary fraction of quasars matches 10 CRTS candidates with luminosity above that of PG 1302 (dark grey, bounded by solid curves), and for which it consists only of PG 1302 (light grey, bounded by red dashed curves). Vertical lines delineate the range of masses preferred by broad line widths. Each shaded region is bounded by the fraction f_{bin} of quasars which are triggered by a binary. In each case, the lines correspond to $f_{\text{bin}} = 0.01, 0.1, 1$ (left to right).

GW frequency, $f_{\text{GW}} = 2t_{\text{bin}}^{-1} \approx 61(\chi/5)$ nHz, places the binary in the range of PTA sensitivity (e.g. [?](#)). We calculate an SNR for the PG 1302 binary choosing optimistic binary parameters $M = 10^{9.4}M_{\odot}$, $q = 0.5$, and $t_{\text{bin}} = 1447\chi^{-1}$ d. The GW induced rms timing residual (for simplicity, adopting the sky and polarization averaged values) is $\delta t_{\text{GW}} = \sqrt{8/15}h/(2\pi f_{\text{GW}})\sqrt{f_{\text{GW}}T} \approx 2.6(\chi/5)^{1/6}\sqrt{T_{\text{yr}}}$ ns for a one year observation time T_{yr} ; for $f_{\text{GW}} \gtrsim 10$ nHz, the timing residual noise is nearly constant in frequency. Using noise curves of currently operating PTAs, the SNR for GW detection of the PG 1302 binary is $\sim 0.005(\chi/5)^{1/6}\sqrt{T_{\text{yr}}}$ for NANOGrav (fig. 12 in [?](#)) or $\sim 0.011(\chi/5)^{1/6}\sqrt{T_{\text{yr}}}$ for the PPTA (black curve in fig. 9 of [?](#)). The reduced binary period increases the SNR, but this increase is unfortunately modest. Future detectors, such as the international pulsar timing array (iPTA; [?](#)) as well as inclusion of the square kilometre array (SKA; [?](#)) in the PTA telescope networks will improve the SNR by about an order of magnitude, but PG 1302 will remain a factor of ~ 10 below detection.

4.3 Testing the Binary BH Scenario for PG 1302

4.3.1 Broad Line Variability and Asymmetry

? report a $\sim (150 \pm 50) \text{ km s}^{-1}$ offset between the broad and narrow line components of PG1302's H β emission line. This is much smaller than the secondary BH's orbital speed,

$v_2 = 14,500 (1.5/[1+q])(M/10^{8.5} M_\odot)^{1/3} (\chi/5)^{1/3} \text{ km s}^{-1}$, or the width of the broad lines. Such larger offsets have been predicted for binary SMBHs, assuming that the broad line region (BLR) originates from gas bound to one component of a binary SMBH and thus shares its overall orbital motion (e.g [Tsalmantza et al. 2011](#)). Here we argue that the smaller offset for PG1302 can also be attributed to a binary SMBH, assuming that the BLR is located farther out, in the circumbinary gas. Using simple toy models, we show that the lopsided geometry of the CBD gas (Fig. 4.1) could generate the small observed offset but large width of PG 1302's broad H β line. The models below are meant to be mere illustrations; a self-consistent description of the BLR is left to future work.

The idea is that the large width of a line can reflect the orbital speed of gas in the CBD (over a range of annuli), whereas the offset of the line centroid is caused only by departures from axisymmetry and can be much smaller. (In a strictly axisymmetric BLR, the blue- and redshifts from gas on opposing sides of the binary would be the same and leave no net offset). To illustrate this, we compute the line offset V_0 as the emission-weighted line-of-sight (l.o.s.) velocity,

$$V_0 = \frac{\int_0^{2\pi} \int_{\mathcal{R}} \rho^n (v_\phi/r)^m v_{\text{los}} r dr d\phi}{\int_0^{2\pi} \int_{\mathcal{R}} \rho^n (v_\phi/r)^m r dr d\phi} \quad (4.2)$$

and the width Γ as the weighted rms l.o.s. velocity

$$\left(\frac{\Gamma}{2\sqrt{2 \ln 2}} \right)^2 = \frac{\int_0^{2\pi} \int_{\mathcal{R}} \rho^n (v_\phi/r)^m (v_{\text{los}} - V_0)^2 r dr d\phi}{\int_0^{2\pi} \int_{\mathcal{R}} \rho^n (v_\phi/r)^m r dr d\phi} \quad (4.3)$$

over a surface patch \mathcal{R} in a $q = 0.5$ simulation (see Fig. 4.1).

In order to select the scaling indices n and m , and to identify a patch \mathcal{R} corresponding to the BLR, we first consider a steady thin disc model for the CBD (as in §4.2), and assume that the CBD is illuminated by a central ionizing source (i.e. the minidisks). For PG1302's parameters, the inner region of the CBD would have density $n \sim 10^{12-13} \text{ cm}^{-3}$ and would be highly opaque to ionizing radiation. In this case, recombinations in a volume corresponding to a very thin ($\Delta R \ll R$) inner annulus of the CBD would balance the central ionizing photon rate. In particular, the disc would absorb the covering fraction $2\pi RH/(4\pi R^2) = 0.05H/(0.1R)$ of the central luminosity. PG 1302's bolometric luminosity is $\sim 6 \times 10^{46} \text{ erg s}^{-1}$; assuming that $\gtrsim 1/3$ rd of this is emitted in the UV, $\gtrsim 5\%$ of which is absorbed by the CBD (i.e. $H/R \gtrsim 0.1$), this would be sufficient to provide the total power $\sim 10^{45} \text{ erg s}^{-1}$ measured in the broad lines (?). The line emission from each patch of the CBD depends only upon the number of ionizing photons incident on the CBD in that direction, i.e. proportional to the scale height H of the inner wall of the CBD. For an adiabatic scale height, assuming vertical disc hydrostatic equilibrium, $n = 1/3$ and $m = -1$. To be specific, \mathcal{R} is chosen by excising the binary plus minidisks, and imposing a surface density range $\Sigma/\Sigma_0 = (0.01, 0.5)$. While somewhat ad hoc, we find that this Σ range accurately picks out the streams edges inside the cavity, and the thin inner edges of the CBD.

Although a standard Shakura-Sunyaev CBD is optically thick outside of the inner cavity, we also consider, for generality, an alternative scenario, where the BLR emission is produced by an optically thin medium, but still resembling the lopsided geometry in our simulations. In this case, line emission would scale with the recombination rate, with $n = 2$ and $m = 0$ in equations. 4.3. We adopt the region \mathcal{R} to be an annulus with inner and outer radii $(2a, 6a)$, chosen to encompass the inner CBD.

We use the simulated surface density Σ , azimuthal velocity v_ϕ and the inferred isothermal scale height of the disc $H = rc_s/v_\phi$ (assuming vertical hydrostatic equilibrium), to compute the

volume density $\rho = \Sigma/H$ and l.o.s. velocity $v_{\text{los}} = v_\phi \cos \phi$. All l.o.s. velocities are multiplied by an additional factor of $\sin i$, where i is the CBD inclination angle, measured from face-on. We calculate V_0 and Γ 10 times per orbit for 20 orbits. Fig. 4.4 displays the variations of V_0 and Γ with time for the optically thin case. The line centroid varies with mean and range

$$V_0^{\text{thn}} = (2 \pm 167) \left[\frac{\sin i}{\sin(14.1^\circ)} \right] \frac{\text{km}}{\text{s}} \left(\frac{M}{10^{8.5} M_\odot} \right)^{\frac{1}{2}} \left(\frac{a}{94 R_S} \right)^{-\frac{1}{2}},$$

$$V_0^{\text{thk}} = (-5 \pm 285) \left[\frac{\sin i}{\sin(11.4^\circ)} \right] \frac{\text{km}}{\text{s}} \left(\frac{M}{10^{8.5} M_\odot} \right)^{\frac{1}{2}} \left(\frac{a}{94 R_S} \right)^{-\frac{1}{2}},$$

while the line full width at half-maximum (FWHM) fluctuates periodically on the lump's orbital time,

$$\Gamma^{\text{thn}} = (4450 \pm 615) \left[\frac{\sin i}{\sin(14.1^\circ)} \right] \frac{\text{km}}{\text{s}} \left(\frac{M}{10^{8.5} M_\odot} \right)^{\frac{1}{2}} \left(\frac{a}{94 R_S} \right)^{-\frac{1}{2}},$$

$$\Gamma^{\text{thk}} = (4450 \pm 636) \left[\frac{\sin i}{\sin(11.4^\circ)} \right] \frac{\text{km}}{\text{s}} \left(\frac{M}{10^{8.5} M_\odot} \right)^{\frac{1}{2}} \left(\frac{a}{94 R_S} \right)^{-\frac{1}{2}}.$$

The fiducial inclination angles above are chosen to match the observed H β FWHM.

We find that both models of a CBD BLR predict a line offset consistent with that observed in ? and which require a CBD inclination angle that also predicts consistent line widths. Additionally, we find that Γ varies by up to ~ 14 per cent of the mean in each case. It is important to emphasize that in the optically thick case, these results arise from the non-axisymmetric velocities in the gas that trace the lopsided inner wall of the CBD, whereas in the optically thin case, they are driven by the lopsided density distribution.

Since fluctuations in the latter case arise from the lump's varying position along the cavity wall, they correlate with long-term variations in the BH accretion rate. In the right-hand panel of Fig. 4.4, we plot the accretion rate on to the BHs, together with the Γ variations of the optically thin case. The phase lag between line-width maximum and accretion maximum derives from the time between lump passage near the BHs, and the lump-enhanced accretion in the minidisks. The phase difference is therefore independent of the observer's viewing angle. This is not true for the amplitude and shape of the FWHM variations, and the relative phase of $V_0(t)$, which depend on

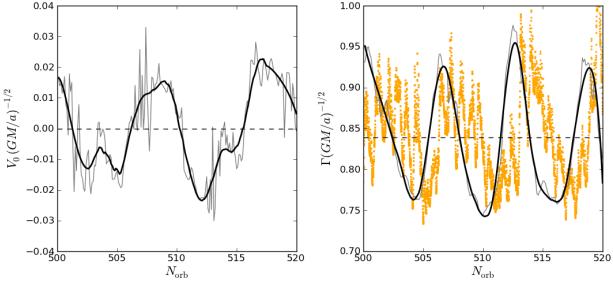


Figure 4.4: Predicted variations of the centroid V_0 (left) and FWHM Γ (right) of an emission line emanating from the inner CBD. The total accretion rate on to both black holes is over plotted in the right-hand panel in arbitrary units (orange). Dark black lines are smoothed versions of the light grey simulation data.

viewing angle. For the optically thick case, the FWHM variations are similar, except that they have a $\sim 20\%$ higher mean, and are \sim half a cycle out of phase with the accretion rate modulation. The phase difference arises because the optically thick model tracks the low density gas at the cavity and stream edges instead of the high density lump. The centroid variations are also similar in the optically thick case but have a few times higher mean and deviation because they track the eccentric cavity shape rather than a symmetric annulus.

The line characteristics which we calculate here are dependent on the existence and magnitude of the orbiting cavity wall over-density (requiring $3 \lesssim \chi \lesssim 8$). As long as the binary has mass ratio such that a cavity wall lump is generated, our BLR calculation is largely unchanged. In addition to mass ratio, disc viscosity and temperature affect the lump size and thus the magnitude of broad line variations. The line shape also depends on the broad line emission model. Hence, a full study of CBD broad lines, which examines more sophisticated recombination models and a range of disc parameters, binary mass ratios and viewing angles, is warranted in a future study. Parameter dependences aside, observation of line variability, matched to luminosity variability, would provide evidence for the CBD model and the origin of the BLR as well as identification of the CBD cavity wall period.

4.3.2 Relativistic Effects

Beaming. ? have shown that if PG1302 consists of a massive ($M \gtrsim 10^9 M_\odot$) but unequal-mass ($0.03 \lesssim M_2/M_1 \lesssim 0.1$) SMBH binary, seen within $\lesssim 30^\circ$ of edge-on, then the entire 0.14 mag variability of PG 1302 can be explained by relativistic Doppler boost. In the hydrodynamical explanation proposed here, where the 5.2-yr modulation arises from variations in the accretion rate, relativistic boost would also inevitably imprint additional sinusoidal modulations at the true (shorter) binary period. The effect would be enhanced, because the secondary’s velocity is higher by a factor of $\chi^{1/3}$, potentially causing a detectable second peak in the periodogram at $5.2\chi^{-1}$ yr. Requiring consistency with § 4.3.1, we use the maximum binary mass and minimum mass ratio ($q = 0.03$) to put an upper limit on the secondary’s l.o.s. velocity v_{los} . The relativistic beaming factor is $[\Gamma(1 - v_{\text{los}}/c)]^{\alpha-3}$, where Γ is the Lorentz factor. ? have estimated the spectral index $\alpha = 1.1$ from an average over the continuum in the V band. We find that the corresponding maximum velocity imprints a 0.07 mag amplitude modulation on PG 1302’s light curve. PG 1302’s periodogram does not show a significant secondary peak with sub-5.2 yr periods, but noise modelling suggests that such second peaks would be detectable only at amplitudes of $\gtrsim 0.07$ mag, \sim half of the 5.2-yr modulation (?; see the next section).

Iron K α lines. Because the binary separation can be reduced below $\lesssim 100 R_S$, FeK α lines generated at such small separations can have characteristic binary-related features, such as ‘missing wings’ (due to the central cavity), or ‘see-saw oscillations’ of the red and blue wings (due to Doppler-shifting of the emission from minidisks; McKernan et al. 2013). These may be detectable with the upcoming Astro-H mission (?).

4.3.3 Orbital time-scale Variability

The binary+CBD model discussed above generically predicts multiple periodic variations. If the observed period of PG 1302 is the true binary period, then its periodogram could contain lower frequency, higher-amplitude, and also higher frequency, lower amplitude peaks. These could be revealed in future data, combined with more sophisticated search algorithms for periodicity (e.g. ? and references therein). It will be helpful in such a search that two of the periodicities occur at t_{bin} and $\approx 0.5t_{\text{bin}}$, i.e. with a characteristic 1:2 ratio in frequency. These are expected to be the only two peaks present for $0.05 \lesssim q \lesssim 0.3$. The variability at t_{bin} can disappear entirely, but this happens only in the limit of $q \rightarrow 1$ (presumably rarely realized in nature) . Thus the detection of a secondary peaks, and the characterization of the full variability structure, can help confirm the binary nature of PG 1302, and constrain its parameters. ? searched PG 1302’s available photometric data for the existence of additional peaks at frequencies above or below the strongest and unambiguous 5.2 yr period. No significant peaks were detected, and an upper limit of $\delta m \gtrsim 0.07 - 0.14$ mag (depending on frequency) was derived for the amplitude of additional modulations.

4.4 Conclusions

For binaries with mass ratio in the range $0.3 \lesssim q \lesssim 0.8$, hydrodynamical simulations of CBDs predict dominant luminosity variations at 3 – 8 times the binary orbital period, due to a dense lump in the CBD (Fig. 4.1). If the periodic variability observed in quasar PG 1302 is identified with this lump period, rather than the orbital period of a putative SMBHB, a two to four times smaller binary separation is inferred. This would place the PG 1302 binary securely in the GW-driven regime, making it the first EM detection of such a system, and proving that gas can follow the binary past decoupling. This is encouraging for the possibility of locating EM counterparts of GW sources. Because binaries spend less time at smaller separations, a shorter t_{bin} is in better agreement with the small number of SMBHB candidates reported by G15. The higher orbital velocity of the binary

increases the effects of relativistic beaming, causing optical variability at the orbital period, and also on inferred broad line widths.

The binary+CBD model can be tested as it predicts variability at multiple, well-defined frequencies which depend on binary mass ratio and disc parameters. Since a recent search (?) did not reveal secondary variability in the optical light curve of PG 1302, follow up observations are required. Finally, associating the BLR with the inner annuli of a lumpy CBD, we find that the FWHM of the lines can vary at the period of the continuum variability by ± 14 per cent; we also predict a much smaller shift of the broad line centroids. These predictions are consistent with existing observations of the width and offset of the H β broad line. Follow-up spectra, sampling PG 1302's apparent 5.2 yr period, could test this interpretation of the BLR and aid in identifying the nature of PG 1302's variability.

Acknowledgements

Resources supporting this work were provided by the NASA High-End Computing (HEC) Programme through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center and by the High Performance Computing resources at Columbia University. The authors thank Maria Charisi, Imre Bartos, Adrian Price-Whelan, Jules Halpern, and Roman Rafikov for useful discussions. We also thank Matthew Graham and George Djorgovski for useful information on PG 1302, as well as for providing the light curve in electronic form. We also thank the anonymous referee for comments that helped to improve this Letter. We acknowledge support from a National Science Foundation Graduate Research Fellowship under Grant no. DGE1144155 (DJD) and NASA grant NNX11AE05G (ZH and AMF).

Chapter 5

Relativistic boost as the cause of the periodicity in a massive black hole binary candidate

Assuming PG 1302-102 is a binary, it is natural to attribute its optical emission to gas that is bound to each black hole, forming circumprimary and circumsecondary accretion flows. Such flows, forming “minidisks”, are generically found in high-resolution 2D and 3D hydrodynamical simulations that include the black holes in their simulated domain?????????. Assuming a circular orbit, the velocity of the lower-mass secondary black hole is

$$v_2 = \left(\frac{2\pi}{1+q} \right) \left(\frac{GM}{4\pi^2 P} \right)^{1/3} = 8,500 \left(\frac{1.5}{1+q} \right) \left(\frac{M}{10^{8.5} M_\odot} \right)^{1/3} \left(\frac{P}{4.04 \text{ yr}} \right)^{-1/3} \text{ km s}^{-1},$$

or $\sim 0.03c$ for the fiducial parameters above, where $M = M_1 + M_2$ is the total binary mass, $M_{1,2}$ are the individual masses, $q = M_2/M_1$ is the mass ratio, P is the orbital period, and c is the speed of light. The primary’s orbital velocity is $v_1 = qv_2$. Even if a minidisk has a steady intrinsic

rest-frame luminosity, its apparent flux on Earth is modulated by relativistic Doppler beaming. The photon frequencies suffer relativistic Doppler shift by the factor $D = [\Gamma(1 - \beta_{||})]^{-1}$, where $\Gamma = (1 - \beta^2)^{-1/2}$ is the Lorentz factor, $\beta = v/c$ is the three-dimensional velocity v in units of the speed of light, and $\beta_{||} = \beta \cos \phi \sin i$ is the component of the velocity along the line of sight, with i and ϕ the orbital inclination and phase. Because the photon phase-space density $\propto F_\nu/\nu^3$ is invariant in special relativity, the apparent flux F_ν at a fixed observed frequency ν is modified from the flux of a stationary source F_ν^0 to $F_\nu = D^3 F_{D^{-1}\nu}^0 = D^{3-\alpha} F_\nu^0$. The last step assumes an intrinsic power-law spectrum $F_\nu^0 \propto \nu^\alpha$. To first order in v/c , this causes a sinusoidal modulation of the apparent flux along the orbit, by a fractional amplitude $\Delta F_\nu/F_\nu = \pm(3 - \alpha)(v \cos \phi/c) \sin i$. This modulation is analogous to periodic modulations from relativistic Doppler boost predicted? and observed for extrasolar planets?? and for a double white dwarf binary?, but here it has a much higher amplitude.

The light-curve of PG 1302-102 is well measured over approximately two periods (≈ 10 years). The amplitude of the variability is ± 0.14 mag (measured in the optical V band?), corresponding to $\Delta F_\nu/F_\nu = \pm 0.14$. The spectrum of PG 1302-102 in and around the V band is well approximated by a double power-law, with $\alpha \approx 0.7$ (between $0.50 - 0.55\mu\text{m}$) and $\alpha \approx 1.4$ (between $0.55 - 0.6\mu\text{m}$), apart from small deviations caused by broad lines. We obtain an effective single slope $\alpha_{\text{opt}} = 1.1$ over the entire V band. We conclude that the 14% variability can be attributed to relativistic beaming for a line-of-sight velocity amplitude of $v \sin i = 0.074 c = 22,000 \text{ km s}^{-1}$.

While large, this velocity can be realised for a massive (high M) but unequal-mass (low q) binary, whose orbit is viewed not too far from edge-on (high $\sin i$). In Fig. 1, we show the required combination of these three parameters that would produce a 0.14 mag variability in the sum-total of Doppler-shifted emission from the primary and the secondary black hole. As the figure shows, the required mass is $\gtrsim 10^{9.1} M_\odot$, consistent with the high end of the range inferred for PG 1302-102. The orbital inclination can be in the range of $i = 60 - 90^\circ$. The mass ratio q has to be low $q \lesssim$

0.3, which is consistent with expectations based on cosmological galaxy merger models Volonteri et al. (2003a), and also with the identification of the optical and binary periods (for $q \gtrsim 0.3$, hydrodynamical simulations predict that the mass accretion rates fluctuate with a period several times longer than the orbital period?).

As Fig. 1 shows, fully accounting for the observed optical variability also requires that the bulk ($f_2 \gtrsim 80\%$) of the optical emission arises from gas bound to the faster-moving secondary black hole. We find that this condition is naturally satisfied for unequal-mass black holes. Hydrodynamical simulations have shown that in the mass ratio range $0.03 \lesssim q \lesssim 0.1$, the accretion rate onto the secondary is a factor of 10–20 higher than onto the primary?. Because the secondary captures most of the accreting gas from the circumbinary disk, the primary is “starved”, and radiates with a much lower efficiency. In the (M, q) ranges favoured by the beaming scenario, we find that the primary contributes less than 1%, and the circumbinary disk contributes less than 20% to the total luminosity, leaving the secondary as the dominant source of emission in the three-component system (see details in *Methods*).

The optical light curve of PG 1302-102 appears remarkably sinusoidal compared to the best-studied previous quasi-periodic quasar binary black hole candidate, which shows periodic bursts?. Nevertheless, the light curve shape deviates from a pure sinusoid. In order to see if such deviations naturally arise within our model, we maximised the Bayesian likelihood over five parameters (period P , velocity amplitude K , eccentricity e , argument of pericentre ω , and an arbitrary reference time t_0) of a Kepler orbit? and fit the observed optical light-curve. In this procedure, we accounted for additional stochastic physical variability with a broken power-law power spectrum (i.e. a “damped random walk” ?) described by two additional parameters. This analysis returns a best-fit with an eccentricity of 0.09, but which does not provide a better model than a pure sinusoid (see *Methods*). We have considered an alternative model to explain PG 1302-102’s optical variability, in which the luminosity variations track the fluctuations in the mass accretion rate predicted

in hydrodynamical simulations?????. However, the amplitude of these hydrodynamic fluctuations are large (order unity), and their shape is bursty, rather than sinusoid-like??; as a result, we find that they provide a poorer fit to the observations (see Fig. 2 and *Methods*). For mass ratios $q \gtrsim 0.05$, hydrodynamical simulations predict a characteristic pattern of periodicities at multiple frequencies, but an analysis of the periodogram of PG 1302-102 has not uncovered evidence for multiple peaks?.

A simple test of relativistic beaming is possible due to the strong frequency-dependence of PG 1302-102’s spectral slope $\alpha = d \ln F_\nu / d \ln \nu$. PG 1302-102’s continuum spectrum is nearly flat with a slope $\beta_{\text{FUV}} \equiv d \ln F_\lambda / d \ln \lambda = 0$ in the far UV (0.145-0.1525 μm) band, and shows a tilt with $\beta_{\text{NUV}} = -0.95$ in the 0.20-0.26 μm near UV range (see Fig. 3 and *Methods*). These translate to $\alpha_{\text{FUV}} = -2$ and $\alpha_{\text{NUV}} = -1.05$ in these bands, compared to the value $\alpha = 1.1$ in the optical. The UV emission can be attributed to the same minidisks responsible for the optical light, and would therefore share the same Doppler shifts in frequency. These Doppler shifts would translate into UV variability that is larger by a factor of $(3 - \alpha)_{\text{FUV}} / (3 - \alpha)_{\text{opt}} = 5/1.9 = 2.63$ and $(3 - \alpha)_{\text{NUV}} / (3 - \alpha)_{\text{opt}} = 4.05/1.9 = 2.13$ compared to the optical, and reach the maximum amplitudes of $\pm 37\%$ (FUV) and 30% (NUV).

PG 1302-102 has five separate UV spectra dated between 1992 and 2011, taken with instruments on the Hubble Space Telescope (HST) and on the GALEX satellite (see Fig. 3), as well as additional photometric observations with GALEX at 4 different times between 2006 and 2009 (shown in Fig. 2). The brightness variations in both the far- and near-UV bands show variability resembling the optical variability, but with a larger amplitude. Utilizing our best fit sinusoid model, and allowing only the amplitude to vary, we find that the UV data yields the best fit variability amplitudes of $\Delta F_\nu / F_\nu |_{\text{FUV}} = \pm(35.0 \pm 3.9)\%$ and $\Delta F_\nu / F_\nu |_{\text{NUV}} = \pm(29.5 \pm 2.4)\%$ (shown in Fig. 2). These amplitudes are factors of (2.57 ± 0.28) and (2.17 ± 0.17) higher than in the optical, in excellent agreement with the values 2.63 and 2.13 expected from the corresponding spectral

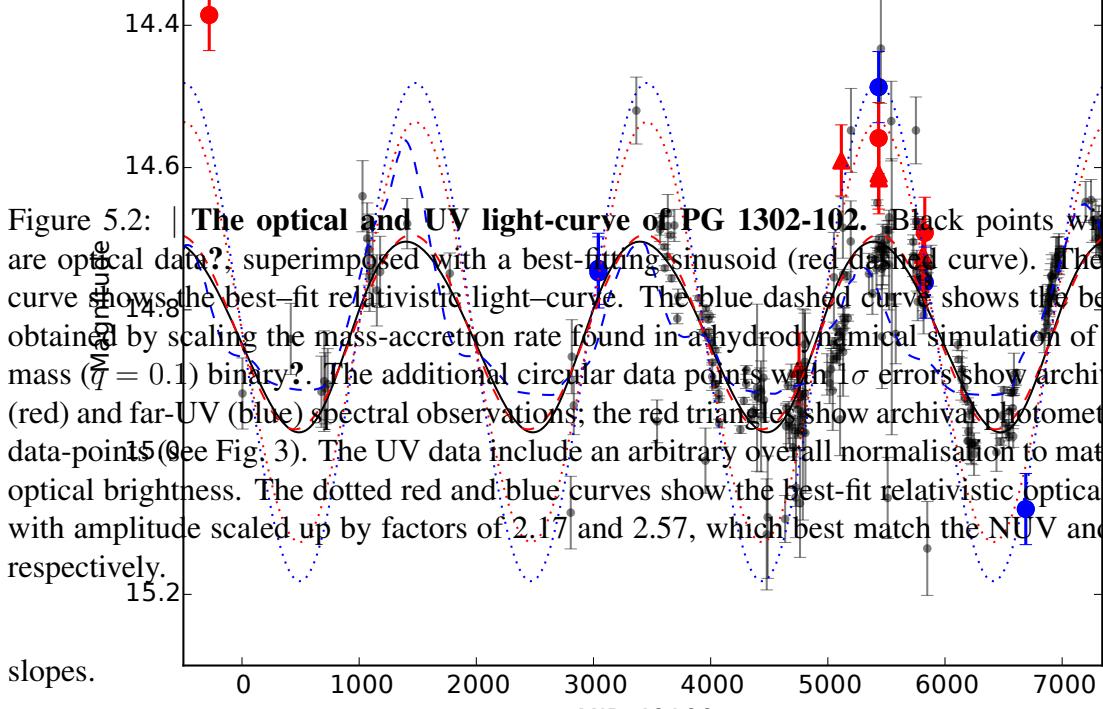
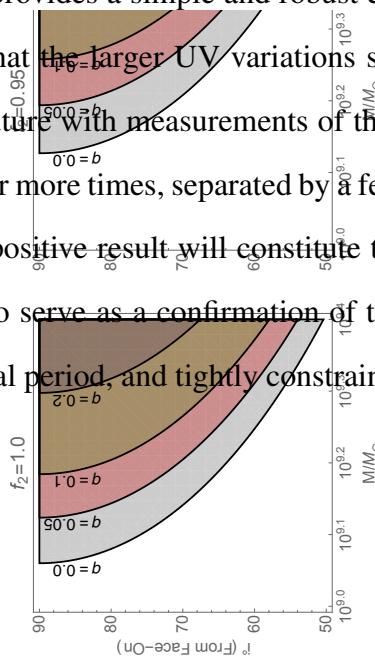


Figure 5.2: **The optical and UV light-curve of PG 1302-102.** Black points with 1σ errors are optical data? superimposed with a best-fitting sinusoid (red dashed curve). The solid black curve shows the best-fit relativistic light-curve. The blue dashed curve shows the best-fit model obtained by scaling the mass-accretion rate found in a hydrodynamical simulation of an unequal-mass ($\tilde{\eta} = 0.1$) binary? The additional circular data points with 1σ errors show archival near-UV (red) and far-UV (blue) spectral observations; the red triangles show archival photometric near-UV data-points (see Fig. 3). The UV data include an arbitrary overall normalisation to match the mean optical brightness. The dotted red and blue curves show the best-fit relativistic optical light curve with amplitude scaled up by factors of 2.17 and 2.57, which best match the NUV and FUV data, respectively.

slopes.

Relativistic beaming provides a simple and robust explanation of PG 1302-102’s optical periodicity. The prediction that the larger UV variations should track the optical light-curve can be tested rigorously in the future with measurements of the optical and UV brightness at or near the same time, repeated two or more times, separated by a few months to ~ 2 years, covering up to half of the optical period. A positive result will constitute the first detection of relativistic black hole binary motion; it will also serve as a confirmation of the binary nature of PG 1302-102, remove the ambiguity in the orbital period, and tightly constrain the binary parameters to be close to those shown in Fig. 1.



METHODS

5.0.1 V-band emission from a three-component system in PG 1302-102.

Here we assume that the PG 1302-102 supermassive black hole (SMBH) binary system includes three distinct luminous components: a circumbinary disk (CBD), as well as an actively accreting primary and secondary SMBH. The optical brightness of each of the three components can be estimated once their accretion rates and the BH masses M_1 and M_2 are specified. Using the absolute V-band magnitude of PG 1302, $M_V = -25.81$ and applying a bolometric correction $BC \approx 10$, we infer the total bolometric luminosity of $L_{\text{bol}} = 6.5(BC/10) \times 10^{46} \text{ erg s}^{-1}$. Bright quasars

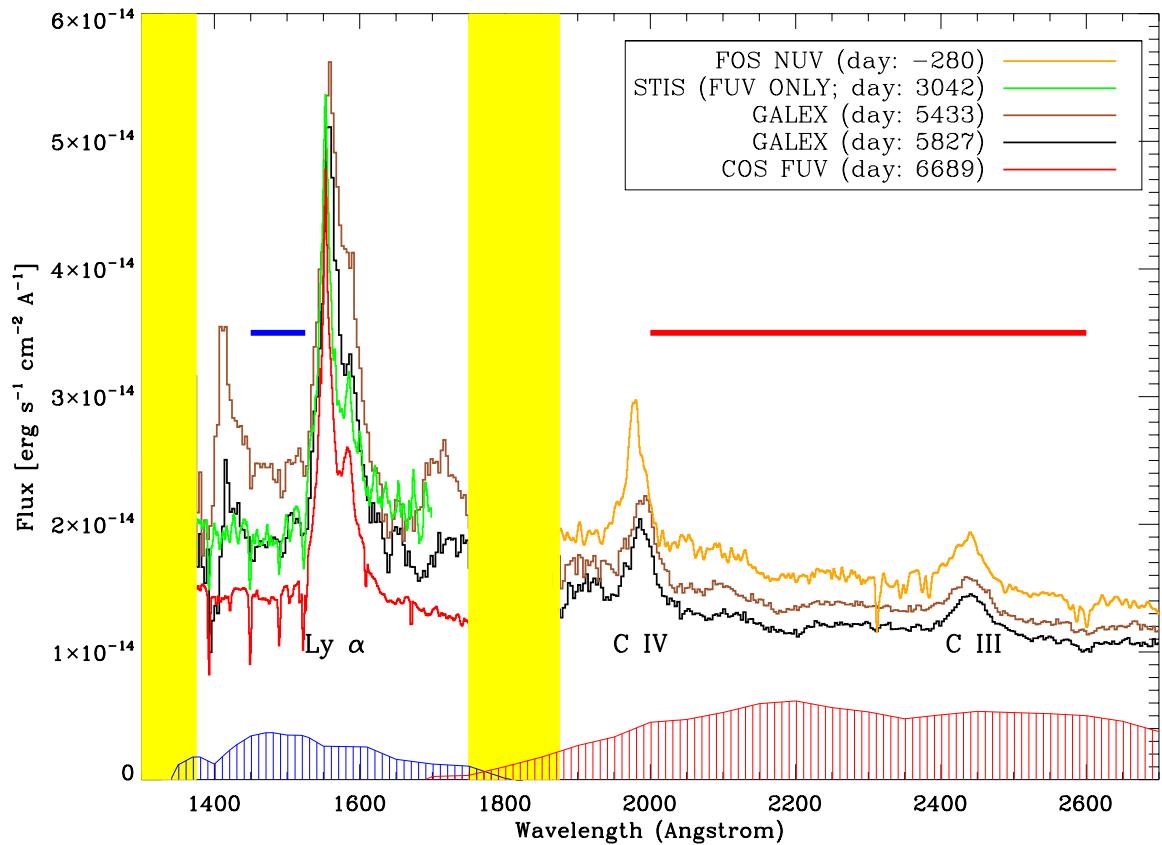


Figure 5.3: | Archival ultraviolet spectra of PG 1302-102 from 1992-2011. Far- and near-UV spectra obtained by the FOS and STIS instruments on the Hubble Space Telescope (HST) and by GALEX are shown. Dates are in MJD (modified Julian day)-49100. Vertical yellow bands mark regions outside the spectroscopic range of both GALEX and HST and contain no useful spectral data. From each spectrum, average flux measurements were computed in one or both of the two UV bands (shown in Fig. 2). GALEX photometric band shapes for FUV and NUV photometry are shown for reference as shaded blue and red curves, respectively. Additional GALEX NUV photometric data were also used in Fig. 2. The UV spectra show an offset by as much as $\pm 30\%$, close to the value expected from relativistic boost (see *Methods*).

with the most massive SMBHs ($M \gtrsim 10^9 M_\odot$), have a typical radiative efficiency of $\epsilon = 0.3$?. Adopting this value, the implied accretion rate is $\dot{M}_{\text{tot}} = L_{\text{bol}}/(\epsilon c^2) = 3.7 M_\odot \text{ yr}^{-1}$.

We identify this as the total accretion rate through the CBD, and require that at small radii, the rate is split between the two black holes $\dot{M}_{\text{tot}} = \dot{M}_2 + \dot{M}_1$ with the ratio $\eta \equiv \dot{M}_2/\dot{M}_1$. Hydrodynamical simulations? have found that the secondary captures the large majority of the gas, with $10 \lesssim \eta \lesssim 20$ for $0.03 \lesssim q \lesssim 0.1$ (where $q \equiv M_2/M_1$). Defining the Eddington ratio of the i^{th} disk as its accretion rate scaled by its Eddington-limited rate $f_{i,\text{Edd}} \equiv \dot{M}_i/\dot{M}_{i,\text{Edd}}$ with $\dot{M}_{\text{Edd}} \equiv L_{\text{Edd}}/0.1c^2$ (here L_{Edd} is the Eddington luminosity for the i^{th} BH, and we have adopted the fiducial radiative efficiency of 0.1 to be consistent with the standard definition in the literature), we have

$$\begin{aligned} f_{\text{CBD,Edd}} &\approx 0.068 \left(\frac{M_{\text{tot}}}{10^{9.4} M_\odot} \right)^{-1}, \\ f_{1,\text{Edd}} &= f_{\text{CBD,Edd}} \frac{(1+q)}{1+\eta} \sim 0.0034 \left(\frac{f_{\text{CBD,Edd}}}{0.068} \right) \left(\frac{1+q}{1.05} \right) \left(\frac{21}{1+\eta} \right), \\ f_{2,\text{Edd}} &= \eta \frac{f_{1,\text{Edd}}}{q} \simeq 1.37 \left(\frac{f_{1,\text{Edd}}}{0.0034} \right) \left(\frac{\eta}{20} \right) \left(\frac{0.05}{q} \right), \end{aligned} \quad (5.1)$$

where the subscripts 1 and 2 refer to the primary and the secondary, and $M_{\text{tot}} \equiv M_1 + M_2$. We adopt a standard, radiatively efficient, geometrically thin, optically thick Shakura-Sunyaev (SS) disk modelShakura & Sunyaev (1973) to compute the luminosities produced in the CBD and the circum-secondary disk (CSD). Although the secondary is accreting at a modestly super-Eddington rate, recent 3D radiation magneto-hydrodynamic simulations of super-Eddington accretion find radiative efficiencies comparable to the values in standard thin disk models?. On the other hand, the primary is accreting below the critical rate $\dot{M}_1 \lesssim \dot{M}_{\text{ADAF}} \approx 0.027(\alpha/0.3)^2 \dot{M}_{\text{Edd}}$ (with α the viscosity parameter) for which advection dominates the energy balance?. We therefore estimate its luminosity from a radiatively inefficient advection-dominated accretion flow (ADAF)??, rather than a SS disk. This interpretation is supported by the fact that PG1302 is known to be an extended

radio source, with evidence for a jet and bends in the extended radio structure?, features that are commonly associated with sub-Eddington sources?.

For the radiatively efficient CBD and CSD, the frequency-dependent luminosity is given by integrating the local modified blackbody flux over the area of the disk

$$\begin{aligned} L_\nu &= 2\pi \int_{R_{\text{in}}}^{R_{\text{out}}} F_\nu[T_p(r)] r dr \\ F_\nu &= \pi \frac{2\epsilon_\nu^{1/2}}{1 + \epsilon_\nu^{1/2}} B_\nu \quad \epsilon_\nu = \frac{\kappa_\nu^{\text{abs}}}{\kappa_\nu^{\text{abs}} + \kappa^{\text{es}}} \end{aligned} \quad (5.2)$$

where B_ν is the Planck function, κ_ν^{abs} is the frequency-dependent absorption opacity, and κ^{es} is the electron scattering opacity. We compute the radial disk photosphere temperature profile T_p by equating the viscous heating rate to the modified blackbody flux

$$\begin{aligned} \left[\frac{3GM\dot{M}}{8\pi r^3} \left[1 - \left(\frac{r_{\text{ISCO}}}{r} \right)^{1/2} \right] \right] &= \zeta(\nu, T_p) \sigma T_p^4 \quad (5.3) \\ \zeta(\nu, T_p) &= \frac{15}{\pi^4} \int \frac{2\epsilon_\nu^{1/2}(x)}{1 + \epsilon_\nu^{1/2}(x)} \frac{x^3 e^{-x}}{1 - e^{-x}} dx \quad x \equiv \frac{h\nu}{k_B T_p}. \end{aligned}$$

where σ is the Stefan-Boltzmann constant and k_B is Boltzmann's constant. In solving for the photosphere temperature we work in the limit that $\kappa_\nu^{\text{abs}} \ll \kappa^{\text{es}}$ following Appendix A of ref. ?, and we adopt $r_{\text{ISCO}}^i = 6GM^i/c^2$ and $(R_{\text{in}}, R_{\text{out}}) = (2a, 200a), (r_{\text{ISCO}}^s, a(q/3)^{1/3})$ for the inner and outer radii of the CBD and CSD, respectively. Here the superscript i refers to the i^{th} disk, a is the binary separation, $6GM$ is the location of the innermost stable circular orbit for a Schwarzschild black hole (our results are insensitive to this choice) and $a(q/3)^{1/3}$ is the secondary's Hill radius (which provides an upper limit on the size of the CSD?).

The optical luminosity of an ADAF is sensitive to the assumed microphysical parameters and its computation is more complicated than for a thin disk. Here we first compute a reference thin-disk luminosity L_{SS} for the primary, and multiply it by the ratio of the bolometric luminosity of an

ADAF to an equivalent thin-disk luminosity from ref. ?,

$$\frac{L_{ADAF}}{L_{SS}} \sim 0.008 \left(\frac{\dot{M}/\dot{M}_{Edd}}{0.0034} \right) \left(\frac{\alpha}{0.3} \right)^{-2}. \quad (5.4)$$

For calculating the reference L_{SS} , we adopted parameters consistent with ref. ?, in particular $\epsilon = 0.1$. Although the above ratio is for the bolometric luminosities, we find that it agrees well with the factor of 100 difference in the V band shown in Figure 6 of ref. ? between ADAF and thin disk spectra with parameters similar to PG 1302 ($10^9 M_\odot$, $\dot{M} = \dot{M}_{ADAF} = 10^{-1.5} \dot{M}_{Edd}$, $\alpha \approx 0.3$).

Extended Data Fig. 1 shows the thin-disk CBD and CSD spectra for a total Eddington ratio of $f_{CBD,Edd} = 0.07$, consistent with the high-mass estimates for PG 1302 needed for the beaming scenario ($M = 10^{9.4} M_\odot$ and $q = 0.05$). The red dot shows the reduced V-band luminosity of an ADAF onto the primary. The secondary clearly dominates the total V band luminosity, with the primary contributing less than 1%, and the CBD contributing $\sim 14\%$. In practice, the contribution from the CBD becomes non-negligible only for the smallest binary masses and lowest mass ratios (reaching 20% for $M < 10^9 M_\odot$ and $q < 0.025$).

We compute the contributions of each of the three components to the total luminosity, $L_{tot}^V = L_1^V + L_2^V + L_{CBD}^V$, and the corresponding total fractional modulation amplitude $\Delta L_{tot}^V/L_{tot}^V = (\Delta L_1^V + \Delta L_2^V)/L_{tot}^V$, for each value of the total mass M and mass ratio q . The primary is assumed to be Doppler-modulated with a line-of sight velocity $v_1 = -qv_2$ while the emission from the CBD is assumed constant over time ($\Delta L_{CBD}^V = 0$). Extended Data Fig. 2 shows regions in the M, q, i parameter space where the total luminosity variation due to relativistic beaming exceeds 14%. This recreates Fig. 1 of the main text, but using the luminosity contributions computed self-consistently in the above model, rather than assuming a constant value of f_2 . Because the secondary is found to be dominant, the relativistic beaming scenario is consistent with a wide range of binary parameters.

5.0.2 Model fitting to the PG 1302 optical light curve.

We fit models to the observed light curve of PG 1302 by maximising the Bayesian likelihood $\mathcal{L} \propto \det|\text{Cov}^D \text{Cov}^{ph}|^{-1/2} \exp(-\chi^2/2)$, where

$$\chi^2 \equiv \mathbf{Y}^T (\text{Cov})^{-1} \mathbf{Y}, \quad (5.5)$$

and $\mathbf{Y} \equiv \mathbf{O} - \mathbf{M}$ is the difference vector between the mean flux predicted in a model and the observed flux at each observation time t_i . Here Cov is the covariance matrix of flux uncertainties, allowing for correlations between fluxes measured at different t_i . We include two types of uncertainties: (1) random (uncorrelated) measurement errors,

$$\text{Cov}_{ij}^{ph} = \begin{cases} \sigma_i^2 & i = j \\ 0 & i \neq j \end{cases} \quad (5.6)$$

where σ_i^2 is the variance in the photometric measurement for the i^{th} data point (as reported in ref. ?), and (2) correlated noise due to intrinsic quasar variability, with covariance between the i^{th} and j^{th} data points,

$$\text{Cov}_{ij}^D = \sigma_D^2 \exp \left[\frac{-|t_i - t_j|}{(1+z)\tau_D} \right]. \quad (5.7)$$

where a factor of $(1+z)$ converts the rest frame coherence time to the observer's frame where the t_i are measured. The parameters σ_D and τ_D determine the amplitude and coherence time of correlated noise described by the damped random walk (DRW) model?. These parameters enter the normalisation of the Bayesian likelihood, and this normalisation must therefore be included when maximising the likelihood over these parameters?. The covariance matrix for the total noise

is given by $\text{Cov} = \text{Cov}^D + \text{Cov}^{ph}$. We assume both types of noise are Gaussian, which provides a good description of observed quasar variability?.

We then fit the following four different types of models to the data:

- *Relativistic beaming model* with 5+2=7 model parameters: eccentricity, argument of peri-center, amplitude, phase and orbital period, plus the two noise parameters σ_D and τ_D .
- *Accretion rate model* with 3+2=5 model parameters: amplitude, phase and period, plus the two noise parameters. This model assumes that PG1302's light-curve tracks the mass accretion rates predicted in hydrodynamical simulations. For near-equal mass binaries, several studies have found that the mass accretion rates fluctuate periodically, but they resemble a series of sharp bursts, unlike the smoother, sinusoid-like shape of PG1302's light-curve. To our knowledge, only three studies to date have simulated unequal-mass ($q \leq 0.1$) SMBH binaries???. The accretion rates for these binaries are less bursty; among all of the cases in these three studies, the $q = 0.075$ and $q = 0.1$ binaries in ref ? resemble PG1302's light-curve most closely (shown in Extended Data Fig. 3). Here we adopt the published accretion curve for $q = 0.1$, and perform a fit to PG1302 by allowing an arbitrary linear scaling in time and amplitude, as well as a shift in phase; this gives us the three free parameters for this model. (We find that the $q = 0.075$ case provides a worse fit.)
- *Sinusoid model* with 3+2 parameters: amplitude, phase and period, plus the two noise parameters. This model is equivalent, to first order in v/c , to the beaming model restricted to a circular binary orbit.
- *Constant luminosity model* with 2 parameters: for reference only – to quantify how poor the fit is with only an amplitude plus the two noise parameters.

In each of these models above, we have fixed the mean flux to equal its value inferred from the optical data. We have found that allowing the mean to be an additional free parameter did

not change our results. The highest maximum likelihood is found for the beaming model, with best-fit values of ($P = [1996]_{-35}^{+29}$ days, $K = [0.065]_{-0.006}^{+0.007}$, $c, e = 0.09_{-0.06}^{+0.07}$, $\cos \omega = [-0.65]_{-0.06}^{+1.2}$, $t_0 = [718]_{-34}^{+422}$ days), where the reference point t_0 is measured from MJD–49100. Uncertainties are computed using a Markov Chain Monte Carlo algorithm ^a in which we sample the 7D posterior probability of the model given the G15 data. In practice, we employ 28 individual chains to sample the posterior for 1024 steps each. Throwing away the first 600 steps ('burning in'), we run for 424 steps and quote parameters which sample the maximum posterior probability with errors given by the 85th and 15th percentile values. The best-fit noise parameters are found to be $(\sigma_D, \tau_D) = (0.049 \text{ mag}, 37.6 \text{ days})$. The best-fit model has a reduced $\chi^2/(N - 1 - 7) \approx 2.1$, where $N = 245$ is the number of data points.

To assess which of the above model is favoured by the data, we use the Bayesian Information Criterion (BIC), a standard method for comparing different models, penalising models with more free parameters?. Specifically, $BIC = -2 \ln \mathcal{L} + k \ln N$, where the first term is evaluated using the best-fit parameters in each of the models and where k is the number of model parameters. We find the following differences ΔBIC between pairs of models:

$$BIC_{\text{Acc}} - BIC_{\text{Beam}} = 4.0 \text{ (Beaming model preferred over accretion model)}$$

$$BIC_{\text{Acc}} - BIC_{\text{Sin}} = 14.9 \text{ (Sin model strongly preferred over accretion model)}$$

$$BIC_{\text{Sin}} - BIC_{\text{Beam}} = -10.9 \text{ (Sin model strongly preferred over Beaming model)}$$

$$BIC_{\text{Const}} - BIC_{\text{Beam}} = 11.5 \text{ (Beaming model strongly preferred over pure noise)}$$

$$BIC_{\text{Const}} - BIC_{\text{Sin}} = 22.4 \text{ (Sin model strongly preferred over pure noise).}$$

We conclude that a sinusoid, or equivalently the beaming model for a circular binary, is the preferred model. In particular, this model is very strongly favoured over the best fit accretion models (see Extended Data Fig. 3), $\Delta BIC > 14.9$. For the assumed Gaussian distributions, this corresponds to an approximate likelihood ratio of $\exp(-14.9/2) \approx 5.7 \times 10^{-4}$. Although our best fit beaming model has a small non-zero eccentricity, we find no evidence for nonzero eccentricity;

^aWe use the *emcee* code ?.

the 7-parameter eccentric model is disfavoured (by $\Delta BIC = 11.5$) over the 5-parameter circular case.

We have conservatively allowed the amplitude of accretion rate fluctuations to be a free parameter in the accretion models, but we note that the accretion rate variability measured in hydrodynamic simulations exhibits large (order unity) deviations from the mean, even for $0.05 < q < 0.1$ binaries ??? . In the accretion rate models, an additional physical mechanism needs to be invoked to damp the fluctuations to the smaller $\sim 14\%$ amplitude seen in PG 1302 (such as a more significant contribution from the CBD and/or the primary).

5.0.3 Disk Precession.

The lowest BIC model, with a steady accretion rate and a relativistic boost from a circular orbit, has a reduced $\chi^2 = 2.1$, indicating that the relativistic boost model with intrinsic noise does not fully describe the observed light-curve. The residuals could be explained by a lower-amplitude periodic modulation in the mass accretion rate, which is expected to have a non-sinusoidal shape (i.e. with sharper peaks and broader troughs, as mentioned above?). Alternatively, the minidisks, which we have implicitly assumed to be co-planar with the binary orbit, could instead have a significant tilt?.

A circum-secondary minidisk that is tilted with respect to the binary's orbital plane will precess around the binary angular momentum vector, causing additional photometric variations due to the changing projected area of the disk on the sky. The precession timescale can be estimated from the total angular momentum of the secondary disk and the torque exerted on it by the primary black hole. The ratio of the precession period to the binary's orbital period is ?,

$$\frac{P_{\text{prec}}}{P_{\text{orb}}} = -\frac{8}{\sqrt{3}} \frac{\sqrt{1+q}}{\cos \delta}, \quad (5.8)$$

where we have chosen the outer edge of the minidisk to coincide with the secondary's Hill sphere

$R_H = (q/3)^{1/3}a$, for binary semi-major axis a . This choice gives the largest secondary disk and the shortest precession rates. The angle δ between the disk angular momentum vector and the binary angular momentum vector can range from $-\pi/2$ to $\pi/2$. For small binary mass ratios, consistent with the relativistic beaming scenario, the precession can be as short as $4.8P_{\text{orb}}$, causing variations on a timescale spanning the current observations of PG 1302-102. The precession timescale would be longer ($> 20P_{\text{orb}}$) for a smaller secondary disk tidally truncated at $0.27q^{0.3}a$?, and with a more inclined (45°) disk.

5.0.4 Archival UV data.

FUV (0.14-0.175 μm) and NUV (0.19-0.27 μm) spectra of PG1302 were obtained by the Hubble Space Telescope (HST) and the Galaxy Evolution Explorer (GALEX) since 1992. HST/Faint Object Spectrograph (FOS) NUV spectra were obtained on July 17, 1992 (pre-COSTAR)?. HST/Space Telescope Imaging Spectrograph (STIS) FUV spectra were obtained on August 21, 2001?. GALEX FUV and NUV spectra were obtained on March 8, 2008 and April 6, 2009 and HST/Cosmic Origins Spectrograph (COS) FUV spectra were obtained on January 28, 2011. All data are publicly available through the Mikulski Archive for Space Telescopes (MAST) at *archive.stsci.edu*. All measurements have been spectrophotometrically calibrated, and binned or smoothed to 1-3 \AA resolution. The spectra (shown in Fig. 2 in the main text) have errors per bin typically less than 2% and published absolute photometric accuracies are better than 5%.

From each spectrum, average flux measurements (shown in Fig. 2 in the main text) were obtained in one or both of two discrete bands: FUV continuum from 0.145-0.1525 μm for FUV (a range chosen to avoid the Ly α line) and NUV continuum from 0.20-0.26 μm . For the GALEX NUV photometric data (also used in Fig. 2) we adopted a small correction (0.005 mag) for the transformation from the GALEX NUV to our NUV continuum band. GALEX FUV photometric data were not used because of the significant contribution from redshifted Ly α . Note that the broad

lines in the UV spectra (in Fig. 3) do not show a large $\Delta\lambda = (v/c)\lambda \approx 140$ Doppler shift. This is unsurprising since the broad lines widths (2,500-4,500 km s⁻¹) much smaller than the inferred relativistic line-of-sight velocities, and are expected to be produced by gas at larger radii, unrelated to the rapidly orbiting minidisks producing the featureless thermal continuum emission?.

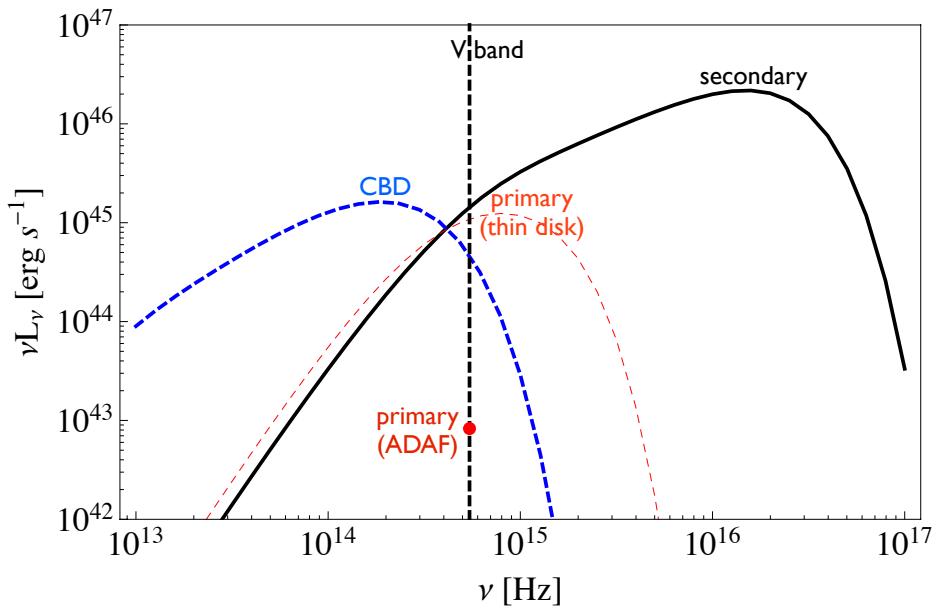


Figure 5.4: Extended Data Figure 1 | Model spectrum of PG1302. Circumbinary (dashed blue) and circumsecondary (solid black) disk spectra for a total binary mass of $10^{9.4}$, binary mass ratio $q = 0.05$, and ratio of accretion rates $\dot{M}_2/\dot{M}_1 = 20$. A vertical dashed line marks the center of the V-band and the approximate flux from an advection-dominated accretion flow (ADAF) is shown as a red dot for the V-band contribution of the primary. The spectrum for a radiatively efficient, thin disk around the primary is shown by the thin red dashed curve for reference.

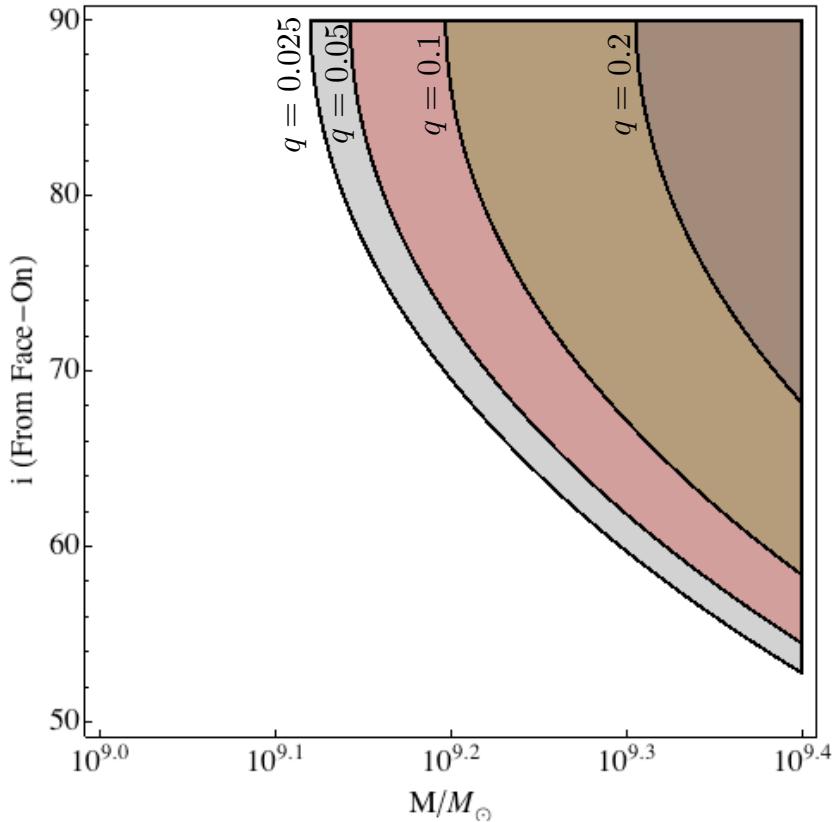


Figure 5.5: Extended Data Figure 2 | Parameter combinations for which the combined V-band luminosity of the three-component system varies by the required 0.14 mag. M is the binary mass, q is the mass ratio, and i is the orbital inclination angle. This figure is analogous to Fig. 1, except instead of adopting a fractional luminosity contribution f_2 by the secondary, the luminosities of each of the three components are computed from a model: the primary's luminosity is assumed to arise from an ADAF, while the secondary's luminosity is generated by a modestly super-Eddington thin disk. Emission from the circumbinary disk is also from a thin disk, and is negligible except for binaries with the lowest mass ratio $q \lesssim 0.01$ (see text).

Figure 5.6: Extended Data Figure 3 | Model fits to PG1302's optical light curve. Best-fitting curves assuming relativistic boost from a circular binary (solid black curves), a pure sinusoid (red dotted curves) and accretion rate variability adopted from hydrodynamical simulations? (blue dashed curves) for a $q = 0.075$ (a) and a $q = 0.1$ (b) mass-ratio binary. The grey points with 1σ errors bars show the data for PG1302?

Part II

Neutron Star - Black Hole Binaries

“Magnets, how do they work?”

– The Insane Clown Posse

Chapter 6

Big Black Hole, Little Neutron Star: Magnetic Dipole Fields in the Rindler Spacetime

6.1 Introduction

Although intrinsically dark, a black hole (BH) can potentially act as a battery in an electromagnetic circuit – a battery that can power great luminosities when connected to other elements in the circuit ([MacDonald & Thorne 1982](#)).

Blandford-Znajek famously proposed a BH battery as the power source for quasar jets [Blandford & Znajek \(1977\)](#). In their well-known model, a spinning BH twists a strong magnetic field anchored in an accretion disk to create an emf that powers an energetic jet. The BH spins down as energy is lost to the luminosity of the jet. In a related yet novel scenario, it was recently proposed [McWilliams & Levin \(2011\)](#) that a magnetized neutron star (NS) in orbit with a BH could light up. When the BH orbits within the magnetosphere of the NS, the relative motion of the BH

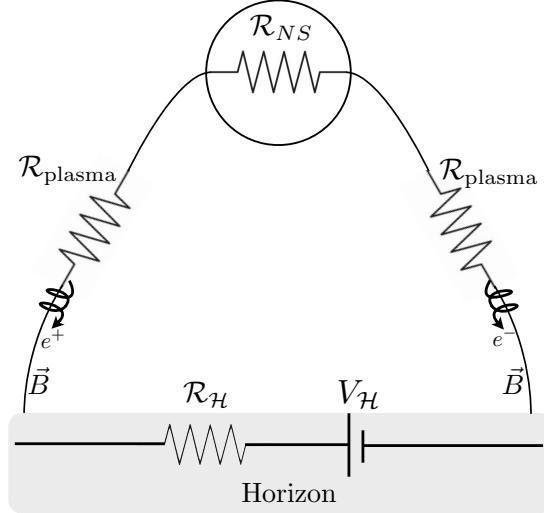


Figure 6.1: Neutron star - Rindler horizon effective circuit diagram. Magnetic field lines act as wires connecting the neutron star to the horizon. Current flows in (out) of the horizon via positively (negatively) charged particles spiraling in tight Larmor radii around magnetic field lines into the horizon.

through the NS dipole field could generate an emf. The BH acts as a battery, the field lines as wires, the charged particles of the NS magnetosphere as current carriers, and the NS itself behaves as a resistor. In principle, the orbit would wind down as angular momentum is lost to the circuit, although in practice gravitational radiation drains angular momentum by far the faster. The circuit is illustrated schematically in Figure 6.1. (See also Refs. [Piro \(2012\)](#); [Lai \(2012\)](#); [?](#); [Palenzuela et al. \(2010, 2013\)](#) for related systems.)

BH-NS pairs may generate gamma-ray bursts during merger via the Blandford-Znajek mechanism; when the NS is tidally disrupted, accreting material tows a magnetic field into the ringing BH ([Narayan et al. 1992a](#); [Lee et al. 2005](#); [Faber et al. 2006](#); [Shibata & Uryu 2007](#); [Shibata & Taniguchi 2008](#); [Etienne et al. 2009](#); [Rezzolla et al. 2011](#); [Etienne et al. 2012](#); [East et al. 2012](#); [Giacomazzo et al. 2013](#)). If the BH is big enough, however, the NS will not be tidally disrupted prior to merger but instead will be swallowed whole, prohibiting the post-merger gamma-ray burst. Since AdLIGO will be most sensitive to binaries with larger BH's (?), it is important to note that the electromag-

netic circuit of (McWilliams & Levin 2011) may be the only electromagnetic counterpart to the gravitational-wave signal.

In this paper, we describe the BH-NS circuit in an analytic calculation valid for large BHs. Very near a large BH, the event horizon looks like a flat wall and in this limit the Schwarzschild metric can be approximated by a Rindler metric – the metric of a flat spacetime as measured by observers with uniform proper-accelerations. In the Rindler limit, calculations are simplified while some of the key physics is retained. Due to acceleration, the Rindler observer also sees a flat-wall event horizon and so the relevant interaction of the EM field with an horizon is present. Also, since the Rindler observer is just on a special worldline in flat spacetime, calculations can be carried out in the Minkowski spacetime and transformed to Rindler, a significant calculational advantage. (The Rindler limit is also used by (MacDonald & Suen 1985) to investigate the fields of a point charge interacting with an horizon.)

We consider a magnetic dipole on an arbitrary worldline near the flat-wall event horizon and derive analytic expressions for the electromagnetic fields. We find that a battery is established when the worldline of the source incorporates motion parallel to the horizon and the pair have approached within the light cylinder of the NS.^a As the pair draws closer under the effects of gravitational radiation, the power of the battery and luminosity of the circuit hits a maximum, just prior to merger. We evaluate the maximum power the black-hole battery would provide to a completed circuit, and thus the maximum luminosity generated. In addition, we estimate the maximum energy to which plasma particles could be accelerated, and thus the type of emission the circuit is capable of producing. As a preview of the conclusions, we quote here the rough scaling

^aBecause we do not capture effects from spatial curvature, an actual BH-NS pair may establish a battery even with head-on motion.

of the voltage and luminosity:

$$\begin{aligned} V_{\mathcal{H}}^{\max} &\sim 3.3 \times 10^{16} \left(\frac{B_p}{10^{12} \text{ G}} \right) \left(\frac{M}{10 M_{\odot}} \right)^{-2} \text{ statvolts} \\ \mathcal{L}^{\max} &\sim 1.3 \times 10^{42} \left(\frac{B_p}{10^{12} \text{ G}} \right)^2 \left(\frac{M}{10 M_{\odot}} \right)^{-4} \frac{\text{erg}}{\text{s}} \end{aligned} \quad (6.1)$$

where M is the mass of the BH and B_p is the magnetic field strength at the poles of the NS. (Readers who prefer to skip the derivations in favor of the conclusions can fast-forward to the results of §6.8.) These scalings only apply at a fixed height $3M$ above the horizon and are dependent on the unknown resistivities of the plasma and of the NS. Eqs. (6.1) should therefore be taken as a guide only. Still, even with these caveats, the conclusion is that a BH-NS circuit could power high-energy bursts of radiation visible to current missions, especially for the special case of magnetar-strength NS fields. This intriguing possibility calls for more detailed predictions of the timescales and spectra of emission, a topic for future explorations. We hope that, in addition to the above estimates, the electro-vacuum example this paper provides will be a resource for further analytic studies and numerical experiments.

6.2 Set-up and Limits

6.2.1 Rindler Spacetime

Consider the line element in Minkowski spacetime

$$ds^2 = -dT^2 + dX^2 + dY^2 + dZ^2 \quad (6.2)$$

The following coordinate transformation,

$$\begin{aligned} T &= z \sinh(g_H t) & X &= x \\ Z &= z \cosh(g_H t) & Y &= y, \end{aligned} \tag{6.3}$$

leads to the Rindler line element,

$$\begin{aligned} ds^2 &= -\alpha^2 dt^2 + dx^2 + dy^2 + dz^2 \\ \alpha &= g_H z \end{aligned} \tag{6.4}$$

where the lapse function α measures the difference in Rindler observer proper time τ_R and Rindler coordinate time t . For reference, the inverse transformation is given by

$$\begin{aligned} t &= \frac{1}{g_H} \tanh^{-1} \left[\frac{T}{Z} \right] \\ z &= \sqrt{Z^2 - T^2}, \end{aligned} \tag{6.5}$$

and we may write the non-inertial, uniformly accelerated trajectory in Minkowski coordinates as,

$$\begin{aligned} X_R^\mu &= (T, 0, 0, Z) = (z_R \sinh(g_H t), 0, 0, z_R \cosh(g_H t)) \\ u_R^\mu &= \frac{dX_R}{d\tau_R} = (\gamma_R, 0, 0, \gamma_R \beta_R) \\ a_R^\mu &= \frac{du_R}{d\tau_R} = z_R^{-1} (\gamma_R \beta_R, 0, 0, \gamma_R) \end{aligned} \tag{6.6}$$

where

$$\begin{aligned}\beta_R &= \frac{dZ}{dT} = \frac{T}{Z} \\ \gamma_R &= (1 - \beta_R^2)^{-1/2}.\end{aligned}\tag{6.7}$$

It is also useful to express these in Rindler coordinates

$$\begin{aligned}\gamma_R &= \cosh(g_H t) \\ \gamma_R \beta_R &= \sinh(g_H t).\end{aligned}\tag{6.8}$$

The 4-acceleration has constant magnitude:

$$a_R^\mu a_\mu^R = z_R^{-2}\tag{6.9}$$

and so observers of constant Rindler coordinate z_R have a 4-acceleration of constant magnitude according to a Minkowski observer.

Figure 6.2 is a Minkowski spacetime diagram demonstrating the wedge occupied by the Rindler spacetime (shaded region). The worldline of a stationary Minkowski observer is denoted by the vertical dotted line while the worldline of a Rindler observer is denoted by the dashed hyperbolic trajectory (Eqs. (6.6)). Due to their accelerations, Rindler observers are causally disconnected from the non-shaded region of Minkowski space in Figure 6.2 and thus experience an event horizon at $z = 0$ ($T = \pm Z$, $Z > 0$).

With the choice of $g_H = 1/(4M)$ and the transformations,

$$x = 2M\phi \quad y = 2M(\theta - \pi/2) \quad z = 4M(1 - 2M/r)^{1/2}\tag{6.10}$$

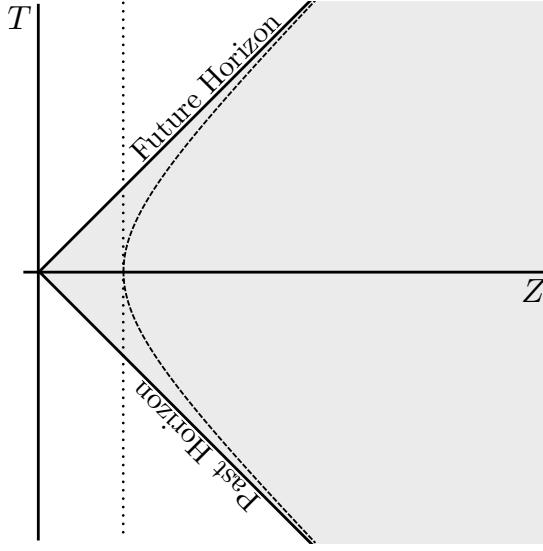


Figure 6.2: Rindler space is the shaded wedge given by $T > \pm Z$, $Z > 0$ on the Minkowski spacetime diagram. The dotted vertical line is the trajectory of a Minkowski observer and the dashed hyperbolic line is that of a Rindler observer as viewed by a Minkowski observer. In the frame of the Rindler observer an event horizon exists at $T = \pm Z$.

the Rindler line element approximates the Schwarzschild line element around the point $(r, \phi, \theta) = (2M, 0, \pi/2)$. Errors of order unity in the approximation to the Schwarzschild spacetime occur when $z \rightarrow 4M$ ($\alpha \rightarrow 1$) and $y \rightarrow 2M$ (see e.g. [\(MacDonald & Suen 1985\)](#)). The Rindler limit retains some key features of the spacetime, including gravitational red-shifting and time dilation as well as the event horizon, although it necessarily misses elements of spatial curvature.

6.2.2 Electrodynamical Properties of an Event Horizon and the Horizon Battery

To understand and interpret power generation by the BH-NS circuit, we first review some key features of horizon electrodynamics [Thorne et al. \(1986\)](#). We consult observers who are at a fixed location *relative to the event horizon*. These fiducial observers can tell us if the event horizon has established charge separation and therefore a battery. Around a BH, these observers must

accelerate to maintain a fixed location and avoid plunging into the BH. Similarly, in Rindler space, our fiducial observers accelerate to maintain a fixed location z from the event horizon. So while stationary relative to the horizon, our fiducial observers are not stationary in an absolute sense – they are non-inertial and so must burn fuel to stay at their Rindler-coordinate location.

We are therefore after the electric \mathbf{E}_R and magnetic \mathbf{B}_R fields measured by a Rindler observer – fields due to a magnetic dipole source on an arbitrary worldline – and we want to determine these fields everywhere outside the event horizon. Problematically, the fields of a Rindler observer will necessarily experience divergences at the event horizon due to infinite time dilation. Following the membrane paradigm [Thorne et al. \(1986\)](#), we construct a timelike hyper-surface stretched over the true, null-horizon. On the stretched horizon, fields will be finite. We then apply electromagnetic boundary conditions on this fictitious surface. Since electric field lines can only terminate or originate on sources, the stretched horizon is assigned hypothetical surface charge to satisfy the boundary conditions of any normal \mathbf{E}_R component. Similarly, the stretched horizon is assigned hypothetical surface current to satisfy the boundary conditions of any tangential \mathbf{B}_R component. As can be derived from local versions of Gauss's law and Ampere's law, the fictitious charge density and surface current are given by

$$\mathbf{E}_{\mathbf{R}} \cdot \mathbf{n} \Big|_{\mathcal{H}} = 4\pi\sigma_{\mathcal{H}} \quad \alpha\mathbf{B}_{\mathbf{R}} \Big|_{\mathcal{H}} = 4\pi\mathcal{J}_{\mathcal{H}} \times \mathbf{n} \quad (6.11)$$

where \mathbf{n} is the unit normal to the horizon and \mathcal{H} denotes evaluation at the stretched horizon. The interpretation then is that electric fields terminate on charges in the stretched horizon, and magnetic fields parallel to the horizon are sourced by surface currents.

Combining (6.11) along with the horizon normal component of the differential form of Am-

pere's law gives charge conservation on the horizon

$$\frac{\partial \sigma_{\mathcal{H}}}{\partial t} + \nabla \cdot \mathcal{J}_{\mathcal{H}} = -(\alpha j_n)_{\mathcal{H}} \quad (6.12)$$

where $(\alpha j_n)_{\mathcal{H}}$ is the normal component of currents entering (positive charges flowing in) and leaving (negative charges flowing in) the horizon in units of universal time. The divergence is the two-dimensional divergence computed on the horizon.

The lesson of the membrane paradigm: when electromagnetic field boundary conditions are applied, the horizon behaves as if it were a conductor with the resistivity of free space^b. On this (hypothetical) conductor may exist (hypothetical) surface charges and currents. Eq. (6.12) tells us that charge is conserved as current flows. In the vacuum calculations presented here, the right hand side of (6.12) will always be zero (to fractional errors of order $\alpha_{\mathcal{H}}$ arising from stretching the horizon) as there is no plasma to carry current off the horizon.

We will be particularly interested in the case where the motion of a magnetic field relative to the Rindler observers induces an electric field that has normal components to the horizon. These normal components source a surface charge density on the horizon that must, when integrated over the black hole area, amount to zero net charge for an initially uncharged black hole. Therefore, charge separation is induced on the horizon and that gradient can be interpreted as creating a battery. If an external circuit is connected to the horizon then the horizon emf associated with the charge separation will drive a current in the circuit. The instantaneous emf of such a horizon battery is given by

$$V_{\mathcal{H}} = \int [\alpha \mathbf{E}_{\mathbf{R}}]_{\mathcal{H}} \cdot ds , \quad (6.13)$$

^bThis follows from (6.11) as well as using stationary observers to measure the fields. See Ch. 2 of [Thorne et al. \(1986\)](#).

remembering that an electric potential is only well-defined for electric fields that originate and terminate on source charges, albeit hypothetical source charges in this case. The situation is analogous to a conventional chemical battery. In the horizon battery, energy of motion of the magnetic field source replaces the chemical energy. In the specific case of a NS orbiting a Schwarzschild BH, the energy source is the spin and orbital energy of the binary.

Figure 6.1 shows the equivalent electrical circuit of such a system. The horizon battery drives current in the form of charged magnetosphere particles spiraling along the NS magnetic field lines. Current enters the horizon via positive-charge carriers (positrons) riding magnetic field lines into the horizon and leaves the horizon via negative-charge carriers (electrons) flowing into the horizon. The current flows through three resistors comprised of the NS, the plasma, and the BH. If we know the electric field induced from the orbital motion of the magnetic dipole we can compute an horizon battery voltage. We may calculate the power, as observed at infinity, dissipated by the i^{th} resistive component of the system,

$$\mathcal{P} = \frac{V_{\mathcal{H}}^2}{(\mathcal{R}_{\mathcal{H}} + \mathcal{R}_{NS} + 2\mathcal{R}_{\text{plasma}})^2} \mathcal{R}_i, \quad (6.14)$$

to approximate the luminosity generated by that component. While the resistance of the BH horizon is set by the resistivity of free space Thorne et al. (1986), the resistances of the NS and plasma are interesting unknowns. Although the primary calculations done here are all in vacuum, in §6.8 we use Eq. (7.5) to estimate the power and find that there is potential for significant bursts of energy from black hole batteries.

First, we find exact closed form solutions for the electromagnetic fields of a magnetic dipole on an arbitrary worldline. We then implement those solutions for specific dipole trajectories.

6.3 A magnetic Dipole in arbitrary motion

6.3.1 The Electromagnetic Four-Potential

In Minkowski spacetime, Maxwell's equations for the 4-potential A^α are,

$$\square A^\alpha(x) - \partial^\alpha (\partial_\beta A^\beta) = \frac{4\pi}{c} J^\alpha(x) \quad (6.15)$$

where $J^\alpha(x)$ is the 4-current as a function of the coordinates.^c We choose to work in the Lorentz Gauge $\partial_\beta A^\beta = 0$. Then Maxwell's equations for the 4-potential become sourced wave-equations,

$$\square A^\alpha(x) = \frac{4\pi}{c} J^\alpha(x). \quad (6.16)$$

We choose the 4-current for a point dipole source

$$J^\alpha(x) = \nabla_\mu \int Q^{\alpha\mu}(\tau) \delta^{(4)}[x - x_S(\tau)] d\tau \quad (6.17)$$

where τ is the proper time of the dipole source, not to be confused with the proper time of the Rindler observers τ_R , and the antisymmetric dipole tensor,

$$Q^{\alpha\mu}(\tau) = V^\alpha p^\mu - p^\alpha V^\mu + \epsilon^{\alpha\mu}_{\rho\sigma} V^\rho m^\sigma, \quad (6.18)$$

is the decomposition of electric p and magnetic m parts (Ribaric & Sustersic 1995; Peter Rowe & Rowe 1987). (See appendix §6.11 for more detail.) Notice that V is the instantaneous 4-velocity of the *source*. Also, hereafter X will denote observer coordinates and X_S will denote the coordinates

^cWe use Gaussian units to write Maxwell's equations. In writing Maxwell's equations we have included the proper factors of c . However, everywhere else, in writing the Rindler metric and the 4-velocities etc. we have set $G = c = 1$.

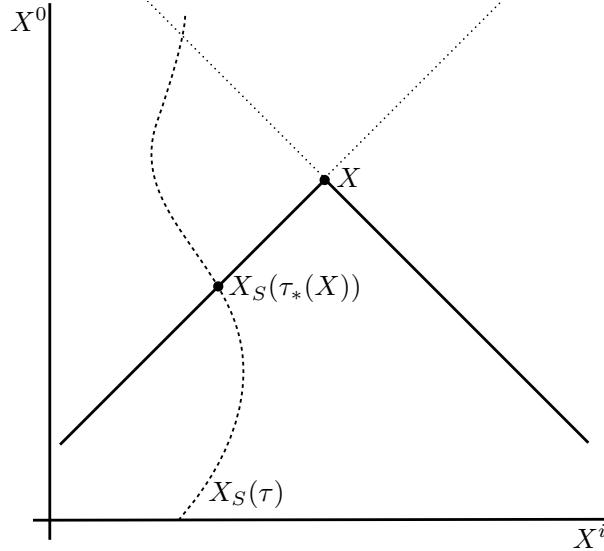


Figure 6.3: Diagram demonstrating the relationship between source trajectory coordinates $X_S(\tau)$, observer coordinates X , and the retarded proper time $\tau_*(X)$ at the intersection of the past light cone of an observer at X and the source trajectory.

along the trajectory of the dipole source. The antisymmetric tensor is fixed by $\epsilon^{0123} = 1$.

The solution for A^α is derived in Appendix 7.7 and can be written in the Minkowski frame, off of the worldline of the source, as

$$A^\alpha = \nabla_\mu \left[\frac{Q^{\alpha\mu}}{r \cdot V} \right]_* . \quad (6.19)$$

By $r \cdot V$ between 4-vectors we mean the inner product $g_{\mu\nu} r^\mu V^\nu$. Since the source may be moving, we must account for the fact that an observer at X will observe fields due to the source in the past, it taking the speed of light for the source information to get to the observer. Therefore, the 4-potential is always evaluated at the retarded time T_* as represented graphically in Figure 6.3. The retarded time is found as a function of observer coordinates by imposing the null condition. We define the relative distance between an observer and a point on the source trajectory in Minkowski

coordinates as

$$r^\mu = \begin{pmatrix} r^0 \\ \mathbf{r} \end{pmatrix} \equiv \begin{pmatrix} T - T_S \\ \mathbf{X} - \mathbf{X}_S(T_S) \end{pmatrix} . \quad (6.20)$$

The null condition is $(r_\mu r^\mu)_* = 0$, with subscript $*$ denoting evaluation at T_* . Then

$$r^0 = \sqrt{\mathbf{r} \cdot \mathbf{r}} = r \quad (6.21)$$

and

$$(\mathbf{X} - \mathbf{X}_S(T_*))^2 = (T - T_*)^2 . \quad (6.22)$$

It should be noted that the covariant derivative in (6.19) is taken with respect to the coordinates X^μ . Since the solution to the null condition for T_* is dependent on the position of the observer, $T_* = T_*(X^\mu)$, T_* is acted on by the covariant derivative.

For completeness, we expand the 4-potential further. Since $(r_\mu r^\mu)_* = 0$ we may write,

$$r_\mu \nabla_\nu r^\mu |_* = 0 = r_\mu (\delta_\nu^\mu - \nabla_\nu \tau V^\mu) |_* = r_\nu - (r \cdot V) \nabla_\nu \tau |_* .$$

Given this extremely useful relation, we can compile a list of gradients that we will need in order to evaluate Eq. (6.19) and construct the field tensor:

$$\begin{aligned} \nabla_\mu \tau &= \frac{r_\mu}{(r \cdot V)} \\ \nabla_\mu r^\nu &= \delta_\mu^\nu - \frac{r_\mu V^\nu}{(r \cdot V)} \\ \nabla_\mu V^\nu &= \frac{r_\mu a^\nu}{(r \cdot V)} \\ \nabla_\mu a^\nu &= \frac{r_\mu}{(r \cdot V)} \dot{a}^\nu \\ \nabla_\mu m^\nu &= \frac{r_\mu}{(r \cdot V)} \dot{m}^\nu \end{aligned} \quad (6.23)$$

where an overdot denotes a τ derivative. Evaluation at $*$ is implied in the relations (6.23).

When there is only a magnetic dipole moment \mathbf{m} , then $Q^{\alpha\mu} = \epsilon^{\alpha\mu}_{\rho\sigma} V^\rho m^\sigma$ and we can expand the 4-potential as:

$$A^\alpha(x) = \epsilon^{\alpha}_{\mu\rho\sigma} r^\mu \left[\frac{a^\rho m^\sigma + V^\rho \dot{m}^\sigma}{(r \cdot V)^2} - \frac{V^\rho m^\sigma (1 + r \cdot a)}{(r \cdot V)^3} \right] \Big|_* . \quad (6.24)$$

Note that the RHS of Eq. (6.24) reduces to the usual stationary dipole solution for a constant dipole at rest ($V^\alpha = (1, 0, 0, 0)$ and $\dot{m} = 0$), as it must.

Given the 4-potential, the electromagnetic field tensor is

$$F_{\alpha\beta} = \nabla_{[\alpha} A_{\beta]} = \nabla_\alpha A_\beta - \nabla_\beta A_\alpha \quad (6.25)$$

Using (6.23) and (6.24) to evaluate (6.25) we find

$$\begin{aligned} F_{\alpha\beta}(x) = & -\epsilon_{[\alpha\beta]\rho\sigma} \frac{a^\rho m^\sigma + V^\rho \dot{m}^\sigma}{(r \cdot V)^2} \\ & + \frac{r_{[\alpha\epsilon\beta]\mu\rho\sigma}}{(r \cdot V)^3} \left\{ r^\mu (2a^\rho \dot{m}^\sigma + V^\rho \ddot{m}^\sigma + \dot{a}^\rho m^\sigma) - V^\mu (a^\rho m^\sigma + V^\rho \dot{m}^\sigma) \right\} \\ & - 2 \frac{V_{[\alpha\epsilon\beta]\mu\rho\sigma}}{(r \cdot V)^3} \left\{ r^\mu (a^\rho m^\sigma + V^\rho \dot{m}^\sigma) \right\} - \frac{a_{[\alpha\epsilon\beta]\mu\rho\sigma}}{(r \cdot V)^3} r^\mu V^\rho m^\sigma + \epsilon_{[\alpha\beta]\rho\sigma} \frac{V^\rho m^\sigma}{(r \cdot V)^3} (1 + r \cdot a) \\ & - \frac{r_{[\alpha\epsilon\beta]\mu\rho\sigma}}{(r \cdot V)^4} \left\{ 3r^\mu (a^\rho m^\sigma + V^\rho \dot{m}^\sigma) (1 + r \cdot a) + r^\mu V^\rho m^\sigma (r \cdot \dot{a}) - V^\mu V^\rho m^\sigma (1 + r \cdot a) \right\} \\ & + 3 \frac{V_{[\alpha\epsilon\beta]\mu\rho\sigma}}{(r \cdot V)^4} r^\mu V^\rho m^\sigma (1 + r \cdot a) \\ & + 3 \frac{r_{[\alpha\epsilon\beta]\mu\rho\sigma}}{(r \cdot V)^5} r^\mu V^\rho m^\sigma (1 + r \cdot a)^2. \end{aligned} \quad (6.26)$$

Again, evaluation at $*$ is implied in the expression (6.26)^d.

^dNote that the $(1 + r \cdot a)$ term, in (6.24) and (6.26), becomes $(c^2 + r \cdot a)$ upon restoring units.

The electromagnetic fields for an observer with 4-velocity u^β are

$$E^\alpha = F^{\alpha\mu} u_\mu \quad B^\alpha = \frac{1}{2} \epsilon^{\alpha\mu\gamma\delta} F_{\gamma\delta} u_\mu \quad (6.27)$$

A stationary Minkowski observer has 4-velocity $u^\mu = (1, 0, 0, 0)$ and the Minkowski fields drop out:

$$\begin{aligned} \mathbf{E}_M &= -\nabla A^0 - \nabla_0 \mathbf{A} \\ \mathbf{B}_M &= \nabla \times \mathbf{A} \quad . \end{aligned} \quad (6.28)$$

In vector notation,

$$\begin{aligned} A^0 &= \frac{(\mathbf{a} \times \mathbf{m}) \cdot \mathbf{r} + (\mathbf{V} \times \dot{\mathbf{m}}) \cdot \mathbf{r}}{(r \cdot V)^2} - \frac{(\mathbf{V} \times \mathbf{m}) \cdot \mathbf{r}}{(r \cdot V)^3} (1 + r \cdot a) \Big|_* \\ \mathbf{A} &= \frac{r^0 (\mathbf{a} \times \mathbf{m}) + r^0 (\mathbf{V} \times \dot{\mathbf{m}}) - a^0 (\mathbf{r} \times \mathbf{m}) - V^0 (\mathbf{r} \times \dot{\mathbf{m}}) + m^0 (\mathbf{r} \times \mathbf{a}) + \dot{m}^0 (\mathbf{r} \times \mathbf{V})}{(r \cdot V)^2} \Big|_* \\ &\quad - \frac{r^0 (\mathbf{V} \times \mathbf{m}) - V^0 (\mathbf{m} \times \mathbf{r}) + m^0 (\mathbf{r} \times \mathbf{V})}{(r \cdot V)^3} (1 + r \cdot a) \Big|_* \end{aligned} \quad (6.29)$$

where the source kinematics can be expressed as

$$\begin{aligned} V^\alpha &= (V^0, \mathbf{V}) = (\gamma_S, \gamma_S \boldsymbol{\beta}_S) \\ \boldsymbol{\beta}_S &= \frac{d\mathbf{X}_S}{dT} \\ a^\alpha &= \left(\gamma_S^4 \left(\boldsymbol{\beta}_S \cdot \frac{d\boldsymbol{\beta}_S}{dT} \right), \gamma_S^2 \frac{d\boldsymbol{\beta}_S}{dT} + \gamma_S^4 \left(\boldsymbol{\beta}_S \cdot \frac{d\boldsymbol{\beta}_S}{dT} \right) \boldsymbol{\beta}_S \right) \end{aligned} \quad (6.30)$$

where γ_S , $\boldsymbol{\beta}_S$ are the instantaneous Lorentz factor and Lorentz boost of the *source*, not to be confused with γ_R , $\boldsymbol{\beta}_R$ of the Rindler *observer*. A dipole with only a magnetic rest frame moment

\mathbf{m}_S moving at β relative to our Minkowski observer has moments

$$\begin{aligned} p^\alpha &= \left(0, \vec{0}\right) \\ m^\alpha &= \left(\gamma_S \boldsymbol{\beta}_S \cdot \mathbf{m}_S, \mathbf{m}_S + (\gamma_S - 1)(\hat{\boldsymbol{\beta}}_S \cdot \mathbf{m}_S)\hat{\boldsymbol{\beta}}_S\right) \end{aligned} \quad (6.31)$$

as explained in more detail in appendix 6.11. With the values in Eqs. (6.30)-(6.31), the Minkowski fields can be computed. Rindler fields are then transformed from the Minkowski fields.

We find the Rindler fields \mathbf{E}_R and \mathbf{B}_R (primed) expressed in terms of Minkowski fields \mathbf{E}_M , \mathbf{B}_M (unprimed) and the Rindler 4-velocity (6.6) via the transformations,

$$E^{\alpha'} = \frac{\partial x^{\alpha'}}{\partial x^\alpha} F^{\alpha\mu} u_\mu^R \quad B^{\alpha'} = \frac{1}{2} \frac{\partial x^{\alpha'}}{\partial x^\alpha} \epsilon^{\alpha\mu\gamma\delta} F_{\gamma\delta} u_\mu^R. \quad (6.32)$$

Here u_R^μ is the Rindler velocity according to a Minkowski observer and the coordinate transformation expresses the components of the fields in the Rindler basis.

A compact way to expand Eqs. (6.32) exploits the fact that any vector can be decomposed as

$$\begin{aligned} \mathbf{E}_R &= \hat{\boldsymbol{\beta}}_R (\hat{\boldsymbol{\beta}}_R \cdot \mathbf{E}_R) + \hat{\boldsymbol{\beta}}_R \times (\mathbf{E}_R \times \hat{\boldsymbol{\beta}}_R) \\ &= \mathbf{E}_R^\perp + \mathbf{E}_R^\parallel \end{aligned} \quad (6.33)$$

where \perp and \parallel refer to components perpendicular to the Rindler horizon and parallel to the Rindler horizon respectively. (So \perp is parallel to β_R and \parallel is perpendicular to β_R .) The fields as measured by a Rindler observer are then expressed conveniently in terms of the fields as measured by a

Minkowski observer as

$$\begin{aligned}\mathbf{E}_R &= \mathbf{E}_M^\perp + \gamma_R \mathbf{E}_M^{\parallel} + \gamma_R (\boldsymbol{\beta}_R \times \mathbf{B}_M^{\parallel}) \\ \mathbf{B}_R &= \mathbf{B}_M^\perp + \gamma_R \mathbf{B}_M^{\parallel} - \gamma_R (\boldsymbol{\beta}_R \times \mathbf{E}_M^{\parallel})\end{aligned}. \quad (6.34)$$

Although we will focus on computing a charge gradient on the horizon to gauge the power output of the BH-circuit in the following examples, it is also instructive to consider the Poynting flux driven by the Rindler dipole. Given the electromagnetic fields as measured by the Rindler observer, we may compute the Poynting vector in Rindler space as seen by an observer at infinity,

$$\mathbf{S} = \frac{\alpha^2}{4\pi} \mathbf{E}_R \times \mathbf{B}_R, \quad (6.35)$$

where one factor of α converts from locally measured energy to energy at infinity and the second factor converts from proper time measured by the local stationary observer to the universal time of the 3+1 split (see §6.2.2).

To understand the meaning of the Poynting flux in this case, we integrate Poynting's theorem over the entire Rindler 3-volume, bounded at infinity and the horizon,

$$\frac{dU}{dt} = - \int \mathbf{S}_\infty \cdot dA - \int \mathbf{E}_H \cdot \mathcal{J}_H dA. \quad (6.36)$$

The last term on the right is evaluated over the stretched horizon since this is the only location in the volume where there are non-zero currents (we could of course add a plasma and get more currents). In the absence of radiation at infinity, we see that any change in EM energy U must be due to ohmic dissipation from horizon surface currents.

Generally, the Poynting flux perceived by a Rindler observer can be expressed in terms of

Minkowski fields as

$$\begin{aligned}
4\pi\alpha^{-2}\mathbf{S} = & \gamma_R E_M^\perp \left[(\hat{\boldsymbol{\beta}}_R \times \mathbf{B}_M^\parallel) + \beta_R \mathbf{E}_M^\parallel \right] \\
& + \gamma_R B_M^\perp \left[-(\hat{\boldsymbol{\beta}}_R \times \mathbf{E}_M^\parallel) + \beta_R \mathbf{B}_M^\parallel \right] \\
& - \gamma_R^2 \boldsymbol{\beta}_R \left[(B_M^\parallel)^2 + (E_M^\parallel)^2 \right] \\
& + \gamma_R^2 (1 + \boldsymbol{\beta}_R^2) \left[\mathbf{E}_M^\parallel \times \mathbf{B}_M^\parallel \right]
\end{aligned} \tag{6.37}$$

The first two terms represent flux parallel to the horizon. The third term is always into the horizon and is due solely to the Rindler motion. The final term can be in or out of the horizon and is proportional to the in or out ($\pm Z$) Poynting flux that would be observed by a Minkowski observer. This final term is the only term that could contribute to power coming out of the dipole-horizon system. However, it can be negated by the inward flux due to Rindler observer motion.

In our vacuum calculations, the Poynting flux can only tell us about radiation from the moving dipole fields, since we have not included a plasma. Instead, we look for the existence of a battery to ascertain if there is a power source. When a magnetosphere is added, the black-hole battery will power an outward Poynting flux at infinity delivering radiation to a distant observer.

6.4 A Freely Falling Dipole Solution

As a check of the above dipole solutions, we consider a dipole source that is stationary in Minkowski space at the location ($X_S = 0, Y_S = 0, Z_S = \text{constant}$). According to the Rindler observer, the magnetic dipole appears to fall straight into the event horizon. In Minkowski coordinates, the world

line is characterized by

$$r^\alpha = \begin{pmatrix} T - T_S \\ X \\ Y \\ Z - Z_S \end{pmatrix}$$

$$V^\alpha = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad a^\alpha = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (6.38)$$

The trajectory is plotted as the dotted worldline in Figure 6.4 as seen by Minkowski observers (top panel) and by Rindler observers (bottom panel). The retarded time can be found in closed form:

$$T_* = T - \sqrt{X^2 + Y^2 + (Z - Z_S)^2}.$$

Because our source is stationary in Minkowski spacetime, there is no dependence on T in the field solutions and

$$(r \cdot V)_* = -r_* \quad (6.39)$$

Then our 4-potential becomes simply

$$A^\alpha(x) = \frac{\epsilon^\alpha_{\mu 0\sigma} r^\mu m^\sigma}{r^3} \Big|_* \quad (6.40)$$

which, written more familiarly, is the potential of a stationary magnetic dipole

$$A^0 = 0, \quad \mathbf{A}(x) = \frac{\mathbf{m}_S \times \hat{\mathbf{r}}}{r^2}. \quad (6.41)$$

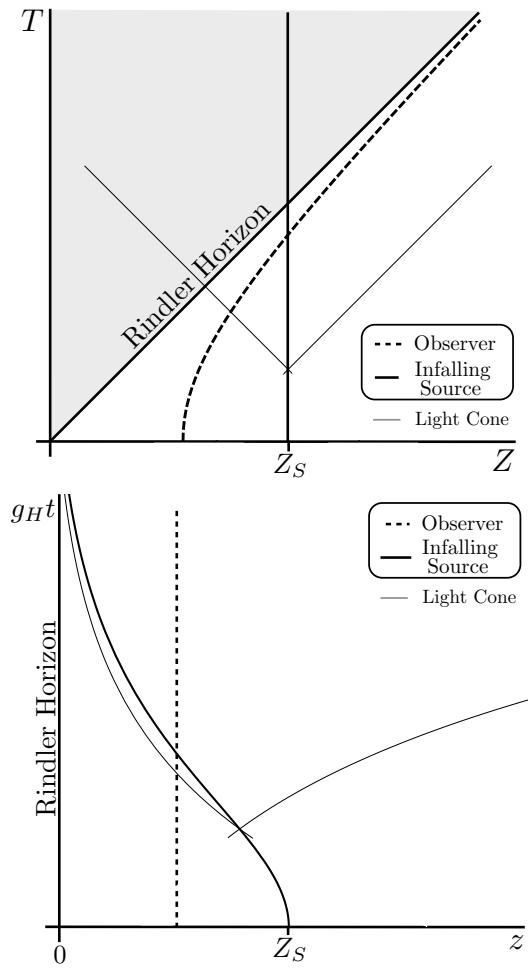


Figure 6.4: Spacetime diagrams for the infalling Rindler dipole of section §6.4. Also shown is the worldline of a Rindler observer. The top panel is drawn by Minkowski observers, the bottom panel is drawn by Rindler observers. Note that $Z_S = z_S$ at $T = t = 0$, hence the labeling of the initial source position.

where $\hat{\mathbf{r}} = \mathbf{r}/r$ and \mathbf{m}_S is the source 3-dipole moment.

The nonzero field components as viewed by the Minkowski observer in the rest frame of the dipole are

$$\begin{aligned}\mathbf{E}_M &= 0 \\ \mathbf{B}_M &= \frac{[3(\mathbf{m}_S \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}_S]}{r^3}.\end{aligned}\tag{6.42}$$

The fields as measured by our Rindler observer are related to the Minkowski observer fields according to the transformation law Eq. (6.34), which in this case simplifies to

$$\begin{aligned}\mathbf{E}_R &= \gamma_R(\boldsymbol{\beta}_R \times \mathbf{B}_M^\parallel) = \gamma_R(\boldsymbol{\beta}_R \times \mathbf{B}_M) \\ \mathbf{B}_R &= \mathbf{B}_M^\perp + \gamma_R \mathbf{B}_M^\parallel.\end{aligned}\tag{6.43}$$

We can plot the fields observed by a Rindler observer in Rindler coordinates if we express \mathbf{r} in Rindler coordinates

$$\mathbf{r} = (x, y, z\gamma_R - Z_S),\tag{6.44}$$

with Z_S just a number for this example.

A slightly different path to the same answer is to transform the 4-potential directly into Rindler coordinates and build the Rindler observer's electromagnetic field tensor. Both approaches give the same result, as they must.

Eq. (6.11) gives the horizon current and charge density,

$$\begin{aligned}\sigma_{\mathcal{H}} &\equiv \frac{\mathbf{E}_{\mathbf{R}}^{\perp}}{4\pi} \Big|_{\mathcal{H}} = 0 \\ \mathcal{J}_{\mathcal{H}} &\equiv \left[\frac{1}{4\pi} \hat{\boldsymbol{\beta}}_R \times \alpha \mathbf{B}_{\mathbf{R}}^{\parallel} \right]_{\mathcal{H}} \\ &= \frac{g_H z_{\mathcal{H}}}{4\pi} \left[\hat{\boldsymbol{\beta}}_R \times \mathbf{B}_{\mathbf{M}}^{\parallel} \gamma_R \right]_{z=z_{\mathcal{H}}} \end{aligned} \quad (6.45)$$

where $\hat{\boldsymbol{\beta}}_R$ is the unit normal to the Rindler horizon and $z_{\mathcal{H}}$ is the position of the stretched Rindler horizon. Since there is no charge on the horizon, there is no potential drop on the horizon – that is, no battery has been established. A freely falling dipole does not generate a power supply in the Rindler limit.

Since $\mathbf{E}_M = 0$, we already know from Eq. (6.37) that there is no outward directed Poynting flux anywhere. Neither the Rindler observer nor the Minkowski observer sees any radiation. For completeness, we write the Rindler Poynting vector explicitly

$$\mathbf{S} = \frac{g_H^2 z^2}{4\pi} \left[\gamma_R \beta_R B_M^{\perp} \mathbf{B}_M^{\parallel} - \gamma_R^2 (B_M^{\parallel})^2 \boldsymbol{\beta}_R \right] \quad (6.46)$$

and plot streamlines of \mathbf{S} in Figure 6.5. Notice there is a component of the Poynting flux parallel to the horizon and there is a component of the Poynting flux into the horizon, both due to the observer's motion outward.

For the sake of illustration, we write out the components of \mathbf{E}_R and \mathbf{B}_R from Eq. (6.43) for the infalling dipole explicitly for the case $\mathbf{m}_S = m \hat{\mathbf{e}}_y$. Using $\gamma_R = \cosh(g_H t)$ and $\gamma_R \beta_R = \sinh(g_H t)$

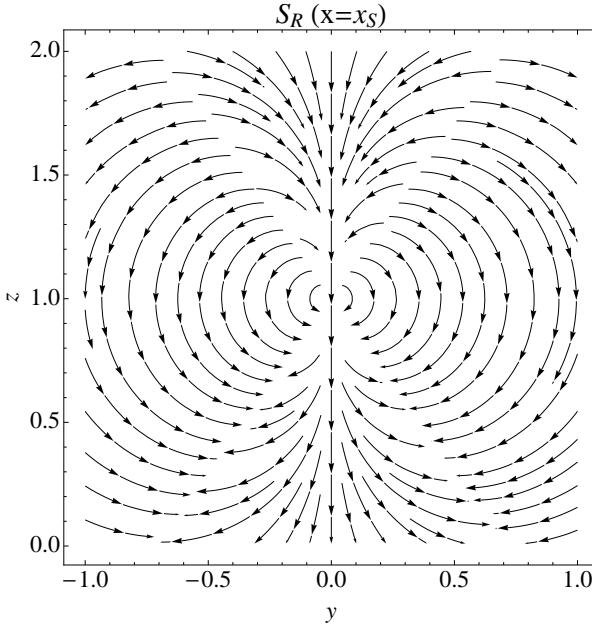


Figure 6.5: The $x = x_S = 0$ slice (plane containing the dipole) of the Poynting flux for the infalling dipole as viewed by Rindler observers. The axes are in units of Z_S .

and the magnitude of r from Eq. (6.44):

$$\begin{aligned}
 B_R^x &= \frac{3mxy}{r^5} \cosh[g_H t] \\
 B_R^y &= -\frac{m(r^2 - 3y^2)}{r^5} \cosh[g_H t] \\
 B_R^z &= \frac{3my(z \cosh[g_H t] - Z_S)}{r^5} \\
 E_R^x &= \frac{m(r^2 - 3y^2)}{r^5} \sinh[g_H t] \\
 E_R^y &= \frac{3mxy}{r^5} \sinh[g_H t] \\
 E_R^z &= 0
 \end{aligned} \tag{6.47}$$

Using (6.47), we plot the fields, and horizon charge densities and currents at three different times during the infall in Figure 6.6. As we have already seen from Eqs. (6.45), there are no charges set

up on the horizon and thus no battery. However, there are currents moving in circles along the horizon. These are the currents implied in the discussion surrounding (6.36) which are responsible for dissipating the energy in the EM fields as they pass through the horizon. Note that the divergence of \mathcal{J}_H in this case is 0, as can be seen from the purely rotational nature in Figure 6.6. Recalling Eq. (6.12), we see that this must be the case for charge conservation to hold in vacuum where currents normal to the horizon, j_n , must be zero.

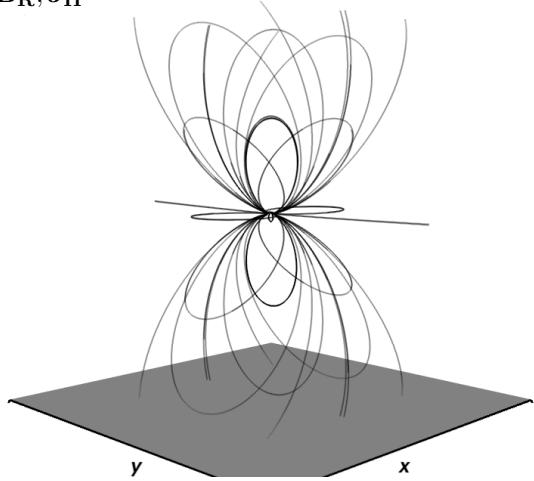
The freely-falling worldline provides a helpful test of our solutions, but no power for an electromagnetic circuit. This system would remain dark, unlike the orbit we explore in the next section. (For an actual BH-NS system at separations that probe spatial curvature, a battery may be established even with pure infall. The effect is not captured here in a flat-wall limit.)

6.5 A Boosted, Freely Falling Dipole Solution

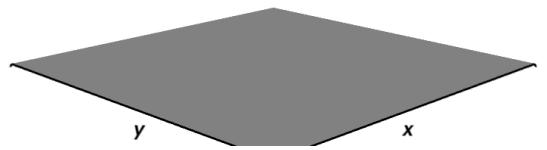
As a second test of the solutions, consider a source that stays at constant Z_S in Minkowski but is boosted in the X, Y plane. Relative to our Rindler observer, the dipole will appear to fall through the horizon but on an arc.

Taking the boost to be at constant velocity in the X -direction, as seen by the Minkowski ob-

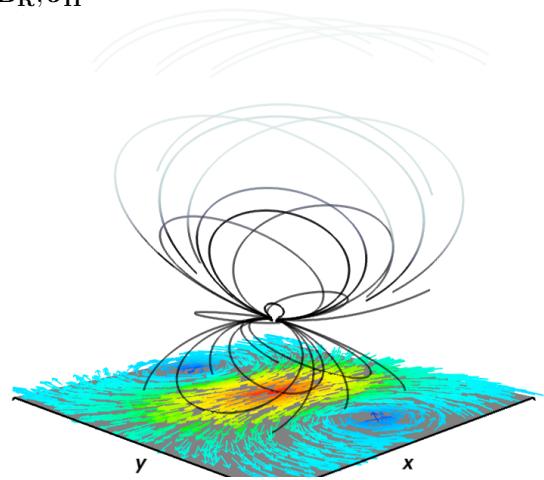
$t = 0$
 $\mathbf{B}_R, \mathbf{J}_H$



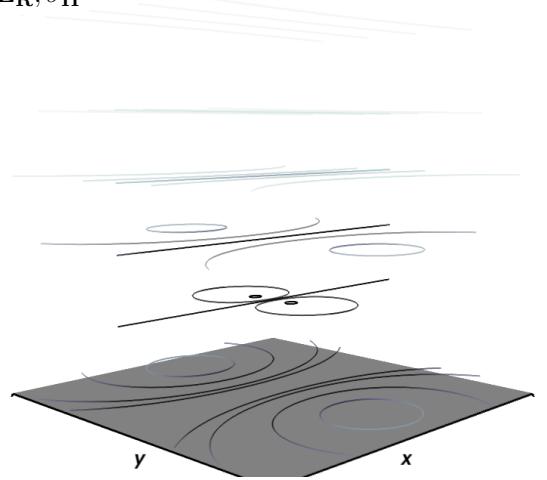
$t = 0$
 \mathbf{E}_R, σ_H



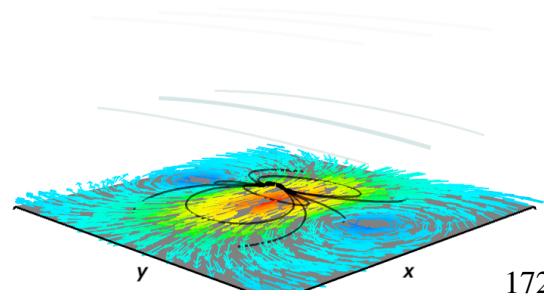
$g_H t = 3/2$
 $\mathbf{B}_R, \mathbf{J}_H$



$g_H t = 3/2$
 \mathbf{E}_R, σ_H



$g_H t = 3$
 $\mathbf{B}_R, \mathbf{J}_H$



$g_H t = 3$
 \mathbf{E}_R, σ_H

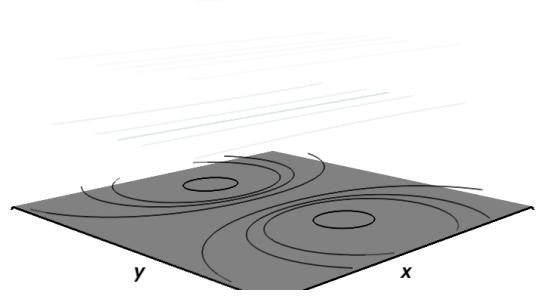


Figure 6.6: 3D visualization of the magnetic dipole field lines of a dipole falling from initial height $z_S(t=0) = Z_S$ above the Rindler horizon, denoted by the gray plane at $z = 0$. The visualization

server, we have

$$r^\alpha = \begin{pmatrix} T - T_S \\ X - \beta_S T_S \\ Y \\ Z - Z_S \end{pmatrix} \quad (6.48)$$

$$V^\alpha = \begin{pmatrix} \gamma_S \\ \gamma_S \beta_S \\ 0 \\ 0 \end{pmatrix} \quad a^\alpha = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The coordinate Z_S is again simply a number, β_S and $\gamma_S = (1 - \beta_S^2)^{-1/2}$ are also constant. It is important to note that although the Minkowski observer sees the dipole boosted in the X -direction at a constant velocity, the Rindler observer sees the dipole slow down in the x -direction as it speeds up in the z . This is illustrated in Figure 6.7.

The retarded time can be found in closed form:

$$T_* = \gamma_S^2(T - \beta_S X) - \sqrt{\gamma_S^2 (T - \beta_S X)^2 - T^2 + X^2 + Y^2 + (Z - Z_S)^2} \quad (6.49)$$

so that

$$(r \cdot V)_* = - [\gamma_S^2(X - \beta_S T)^2 + Y^2 + (Z - Z_S)^2]^{1/2} \quad (6.50)$$

Eq. (6.24) with $a = 0$ and $\dot{\mathbf{m}} = 0$ then gives the 4-potential for a boosted Minkowski dipole. From the 4-potential or the field tensor, Minkowski \mathbf{E}_M and \mathbf{B}_M can be derived from Eq. (6.28) or

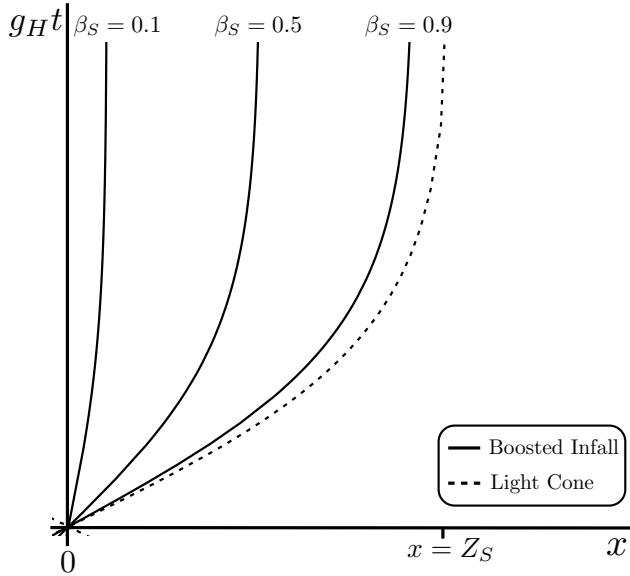


Figure 6.7: Spacetime diagram depicting the x -component of the infalling, boosted dipole worldline of §6.5 from the Rindler observer's perspective, for three different values of β_S . The z -component of the worldline is identical to that portrayed for the infalling dipole in the bottom panel of Figure 6.4 except the light cone structure is altered. Because the infalling boosted dipole approaches the speed of light in the z -direction, the motion in the x -direction must go to zero ($dx/dt \rightarrow 0$). This is evident from the worldlines in this figure which asymptote to vertical lines.

(6.27).

As a check, we may also derive the electromagnetic fields by writing the field tensor for a dipole in the rest frame of a Minkowski observer, transform to a boosted frame, and then transform to the accelerated Rindler frame.

Let the reference frame of the Minkowski observer at rest with respect to the dipole be denoted by a double prime, the frame of the Minkowski observer boosted relative to the source by a single prime, and the Rindler frame by no prime. Then the field tensor $F^{\alpha''\beta''}(X^{\mu''})$ is constructed from Eq. (6.42). The field tensor in the Minkowski boosted frame is given by

$$F^{\alpha'\beta'}(\mathbf{X}') = \Lambda_{\sigma''}^{\alpha'} \Lambda_{\rho''}^{\beta'} F^{\sigma''\rho''}(\mathbf{X}'') \quad (6.51)$$

where $\Lambda_{\sigma''}^{\alpha'}$ is the Lorentz transformation for a boost in the X direction. The boosted coordinates $X^{\mu'}$ are given in terms of the rest frame coordinates $X^{\mu''}$ via an inverse Lorentz transformation. The Rindler \mathbf{E}_R and \mathbf{B}_R fields are then found via Eq. (6.6) and (6.32),

$$\begin{aligned} E_R^\alpha(\mathbf{x}) &= \frac{\partial x^\alpha}{\partial x^{\alpha'}} F^{\alpha'\mu'}(\mathbf{X}') u_{\mu'}^R \\ B_R^\alpha(\mathbf{x}) &= \frac{1}{2} \frac{\partial x^\alpha}{\partial x^{\alpha'}} \epsilon^{\alpha'\mu'\gamma'\delta'} F_{\gamma'\delta'}(\mathbf{X}') u_{\mu'}^R. \end{aligned} \quad (6.52)$$

where $u_{\mu'}^R$ are the components of the Rindler observer's 4-velocity as viewed by the boosted Minkowski observer. The $X^{\mu'}(x^\mu)$ are given by Eqs. (6.3) and we have again kept Rindler coordinates lower case while Minkowski coordinates are upper case.

Carrying out the above procedure, we start with an observer co-moving with the dipole. This observer sees fields,

$$\begin{aligned} \mathbf{E}_{\mathbf{M}''} &= 0 \\ \mathbf{B}_{\mathbf{M}''} &= \frac{3\hat{\mathbf{r}}''(\mathbf{m}_S \cdot \hat{\mathbf{r}}'') - \mathbf{m}_S}{{\mathbf{r}''}^3} \end{aligned} \quad (6.53)$$

where \mathbf{m}_S is the constant rest-frame value of the dipole's magnetic moment and \mathbf{r}'' is a radial coordinate in the rest-frame of the dipole. There exists another Minkowski observer boosted by $-\beta_S$ relative to the source who measures the fields

$$\begin{aligned} \mathbf{E}_{\mathbf{M}'} &= -\gamma_S (\boldsymbol{\beta}_S \times \mathbf{B}_{\mathbf{M}''}) \\ \mathbf{B}_{\mathbf{M}'} &= \mathbf{B}_{\mathbf{M}''} + (\gamma_S - 1) \hat{\boldsymbol{\beta}}_S \times (\mathbf{B}_{\mathbf{M}''} \times \hat{\boldsymbol{\beta}}_S) \end{aligned} \quad (6.54)$$

We obtain the Rindler fields by an application of Eq. (6.34):

$$\begin{aligned}
\mathbf{E}_R &= \gamma_S \gamma_R (\boldsymbol{\beta}_R \times \mathbf{B}_{M''}) \\
&\quad - \left[\gamma_S \hat{\boldsymbol{\beta}}_R \cdot (\boldsymbol{\beta}_S \times \mathbf{B}_{M''}) \right] \hat{\boldsymbol{\beta}}_R \\
&\quad - \left[\gamma_R \mathbf{B}_{M''} \cdot \left((\gamma_S - 1) |\boldsymbol{\beta}_R| \hat{\boldsymbol{\beta}}_S - \gamma_S |\boldsymbol{\beta}_S| \hat{\boldsymbol{\beta}}_R \right) \right] (\hat{\boldsymbol{\beta}}_R \times \hat{\boldsymbol{\beta}}_S) \\
\mathbf{B}_R &= \gamma_S \gamma_R \mathbf{B}_{M''} \\
&\quad - \left[\gamma_S (\gamma_R - 1) (\mathbf{B}_{M''} \cdot \hat{\boldsymbol{\beta}}_R) \right] \hat{\boldsymbol{\beta}}_R \\
&\quad - \left[\gamma_R \mathbf{B}_{M''} \cdot \left((\gamma_S - 1) \hat{\boldsymbol{\beta}}_S - \gamma_S |\boldsymbol{\beta}_S| \boldsymbol{\beta}_R \right) \right] \boldsymbol{\beta}_S
\end{aligned} \tag{6.55}$$

The horizon charge and current densities are

$$\begin{aligned}
\sigma_{\mathcal{H}} &\equiv \frac{\mathbf{E}_{\mathbf{R}}^{\perp}}{4\pi} \Big|_{\mathcal{H}} = - \frac{\hat{\boldsymbol{\beta}}_R \cdot (\gamma_S \boldsymbol{\beta}_S \times \mathbf{B}_{M''}^{\parallel})_{z=z_{\mathcal{H}}}}{4\pi} \\
\mathcal{J}_{\mathcal{H}} &\equiv \left[\frac{1}{4\pi} \hat{\boldsymbol{\beta}}_R \times \alpha \mathbf{B}_{\mathbf{R}}^{\parallel} \right]_{\mathcal{H}} \\
&= \gamma_R \frac{g_H z_{\mathcal{H}}}{4\pi} \left[\gamma_S \left(\hat{\boldsymbol{\beta}}_R \times \mathbf{B}_{M''}^{\parallel} \right) \right. \\
&\quad \left. - \mathbf{B}_{M''} \cdot \left((\gamma_S - 1) \hat{\boldsymbol{\beta}}_S - \gamma_S |\boldsymbol{\beta}_S| \boldsymbol{\beta}_R \right) (\hat{\boldsymbol{\beta}}_R \times \hat{\boldsymbol{\beta}}_S) \right]_{z=z_{\mathcal{H}}}
\end{aligned} \tag{6.56}$$

This example manifests charge separation and therefore a voltage drop across the event horizon.

We have established a BH battery.

To express these Rindler fields in Rindler coordinates, we perform a Lorentz transformation on the Minkowski 4-vector r'' for a boost in the x-direction and use Eqs. (6.3) to write,

$$\mathbf{r}'' = (\gamma_S (x - \beta_S z \sinh[g_H t]), y, z \cosh[g_H t] - Z_S). \tag{6.57}$$

Eqs. (6.57), (6.53), and (6.55) then give the Rindler fields in Rindler coordinates.

The Rindler fields derived in this manner agree with the fields derived from inserting (6.48) into the 4-potential as they must.

Choosing $\mathbf{m}_S = m\hat{\mathbf{e}}_y$, given that we boost in the x -direction, leads to the simplest form for the observed 4-dipole moment,

$$m_R^\mu = m_{M''}^\mu = m_{M'}^\mu = (0, \mathbf{m}_S). \quad (6.58)$$

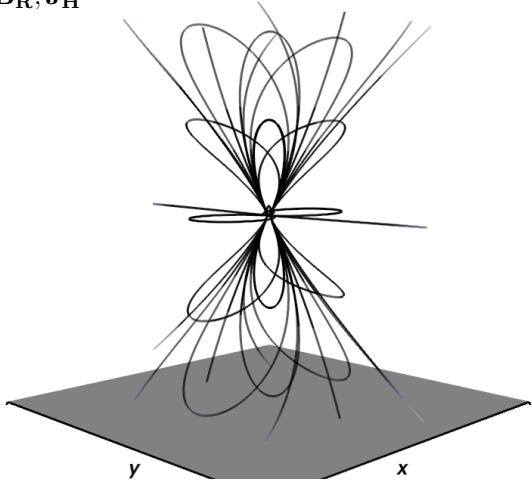
We write out the components of \mathbf{E}_R and \mathbf{B}_R for the boosted, infalling dipole explicitly:

$$\begin{aligned} B_R^x &= \gamma_S \frac{3my(x\cosh[g_H t] - \beta_S Z_S \sinh[g_H t])}{r^5} \\ B_R^y &= -\gamma_S \frac{m(r^2 - 3y^2)}{r^5} \cosh[g_H t] \\ B_R^z &= \gamma_S \frac{3my(z\cosh[g_H t] - Z_S)}{r^5} \\ E_R^x &= \gamma_S \frac{m(r^2 - 3y^2)}{r^5} \sinh[g_H t] \\ E_R^y &= \gamma_S \frac{3my\{x\sinh[g_H t] + \beta_S(z - Z_S \cosh[g_H t])\}}{r^5} \\ E_R^z &= \gamma_S \frac{m(r^2 - 3y^2)}{r^5} \beta_S \end{aligned} \quad (6.59)$$

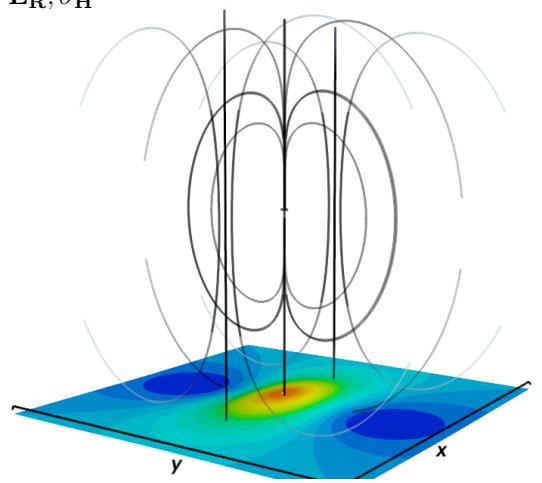
where r is the RHS of (6.50) in Rindler coordinates. Using the above, we plot the fields, and horizon charge and current densities, given by Eqs. (6.56), at three different times during the inspiral in Figure 6.8.

Figure 6.9 shows the Poynting flux generated by the above fields for $\beta_S = 0.1, 0.5, 0.9$. The Poynting flux is directed into the horizon below the dipole signifying the dissipation of the field energy into the horizon (via ohmic dissipation from horizon currents). The increasingly uniform z component of the Poynting flux for increasing β_S is due to the increasing disparity between t and t_* (observers see further into the relative past of the dipole) for larger β_S and smaller z . There is no observed Poynting flux at infinity in this case and hence no radiation from the moving dipole in vacuum. We elaborate on the above points further in §6.7.

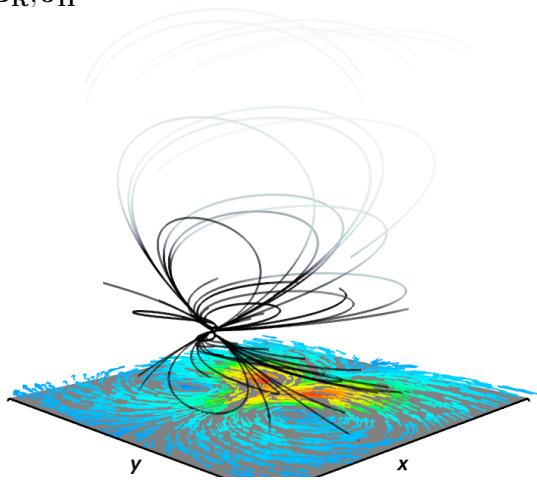
$t = 0$
 $\mathbf{B}_R, \mathbf{J}_H$



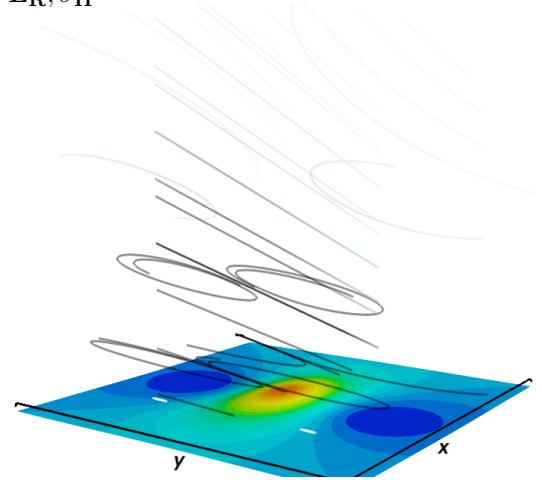
$t = 0$
 \mathbf{E}_R, σ_H



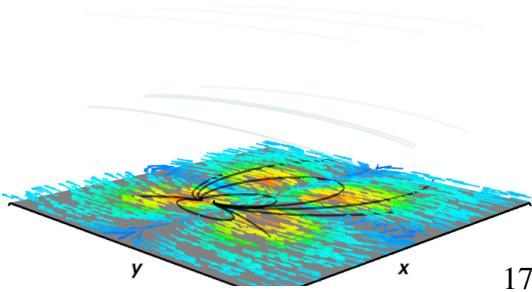
$g_H t = 3/2$
 $\mathbf{B}_R, \mathbf{J}_H$



$g_H t = 3/2$
 \mathbf{E}_R, σ_H



$g_H t = 3$
 $\mathbf{B}_R, \mathbf{J}_H$



$g_H t = 3$
 \mathbf{E}_R, σ_H

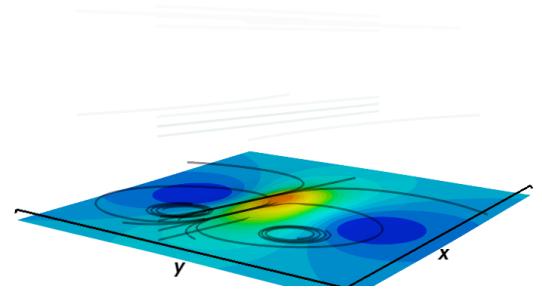


Figure 6.8: 3D visualization of the field lines of a dipole spiraling into the Rindler horizon from initial height $z_S(t = 0) = Z_S$ with an initial boost of $\beta_S = 0.9$ in the x -direction. The visualization

Figure 6.8, as well as the expression for E_R^z and Eq. (6.11) for the horizon charge density, indeed confirm that charge separation occurs on the stretched horizon. Figure 6.10 explores the horizon charge density further. As we will elaborate in §6.7, this charge separation can be considered a result of tangential components of the dipole magnetic field sourcing horizon currents via (6.11). Because no currents are entering or leaving the horizon in the vacuum case, these horizon currents pile up charge on the horizon, that is the divergence of the horizon current in (6.12) is not 0, but cancels a time changing charge density. This is why the currents seem to flow towards regions of positive charge in Figure 6.10. We see also that, as might have been expected, the magnitude of the charge separation grows with the speed of the boost, i.e. the more energy given to boost the dipole along the horizon, the higher the voltage of the horizon battery.

The line of zero charge density in the plane of the stretched horizon is given by,

$$\begin{aligned} x|_{\sigma_H=0} &= \beta_S z_H \sinh[g_H t] \\ &\pm \gamma_S^{-1} \sqrt{2y^2 - (z_H \cosh[g_H t] - Z_S)^2}. \end{aligned} \quad (6.60)$$

On the true horizon,

$$x|_{\sigma_H=0} = \pm \gamma_S^{-1} \sqrt{2y^2 - Z_S^2}. \quad (6.61)$$

On the stretched horizon the shape of the charge separation boosts along the horizon at late times, when $z_H \sinh[g_H t]$ becomes large. On the true horizon however the charge separation is stationary reflecting the freezing in of fields on the horizon.

As can be seen in Figure 6.10, the γ_S^{-1} pre-factor in equation (6.61) morphs the geometry of the charge separation from that of roughly equal parts positive and negative charge at low β_S , to that of smaller regions of larger negative charge density squeezed to the sides of the dipole in the direction of its motion for larger β_S .

The charge separation and corresponding battery emf is a direct consequence of the boosted motion parallel to the horizon. We will see this feature again in the final example (§6.7). In the penultimate section §6.8, we estimate the power produced by a black-hole battery, the luminosities attained in the circuit, and the energy scale of the emission.

6.6 Rindler Dipole

Now suppose there is a magnetic dipole that is uniformly accelerated so that it lives at constant Rindler coordinate z_S . While the Minkowski observers see this dipole accelerate and asymptote to a null trajectory, the Rindler observers see a source dipole at fixed coordinate distance above the horizon.

This is the first case for which we no longer have a check of our solutions. Nor do we have an obvious alternative method of calculation. We must compute fields from our exact solution for the 4-potential from §6.3. The kinematics of the accelerated source are characterized by

$$r^\mu = \begin{pmatrix} T - T_S \\ X \\ Y \\ Z - Z_S(T_S) \end{pmatrix} \quad V^\mu = \gamma_S \begin{pmatrix} 1 \\ 0 \\ 0 \\ \beta_S \end{pmatrix}$$

$$a^\mu = \frac{\gamma_S^2}{Z_S} \begin{pmatrix} \beta_S \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \dot{a}^\mu = \frac{\gamma_S^3}{Z_S^2} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \beta_S \end{pmatrix}$$

where, in this case, $\beta_S = \tanh(g_H t_S) = T_S/Z_S$, $\gamma_S = \cosh(g_H t_S) = (1 - (T_S/Z_S)^2)^{-1/2}$, and $Z_S = \sqrt{z_S^2 + T_S^2} = z_S \cosh[g_H t_S]$, where z_S is the constant height of the Rindler dipole above the

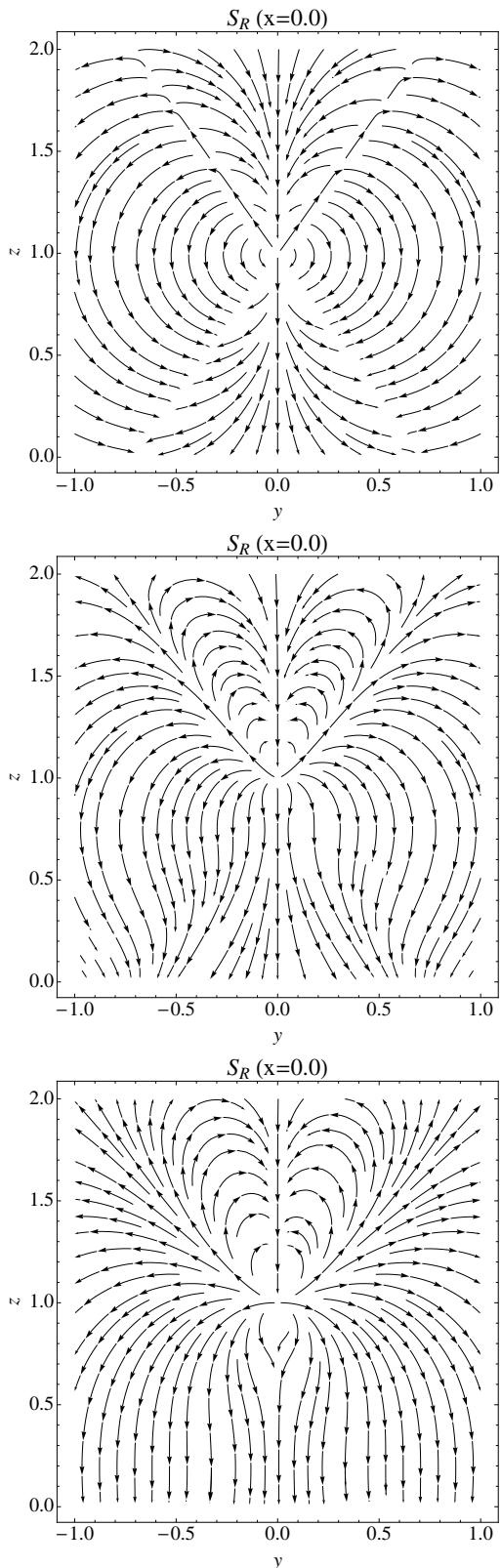


Figure 6.9: An $x = x_S = 0$ slice of the Poynting¹⁸¹ flux for the infalling boosted dipole of §6.5 as viewed by Rindler observers for three different boost magnitudes in the x -direction, $\beta_S = 0.1, 0.5, 0.9$. The Poynting flux is 0 at infinity despite outward components of the field in the region plotted here. The axes are in units of Z_S .

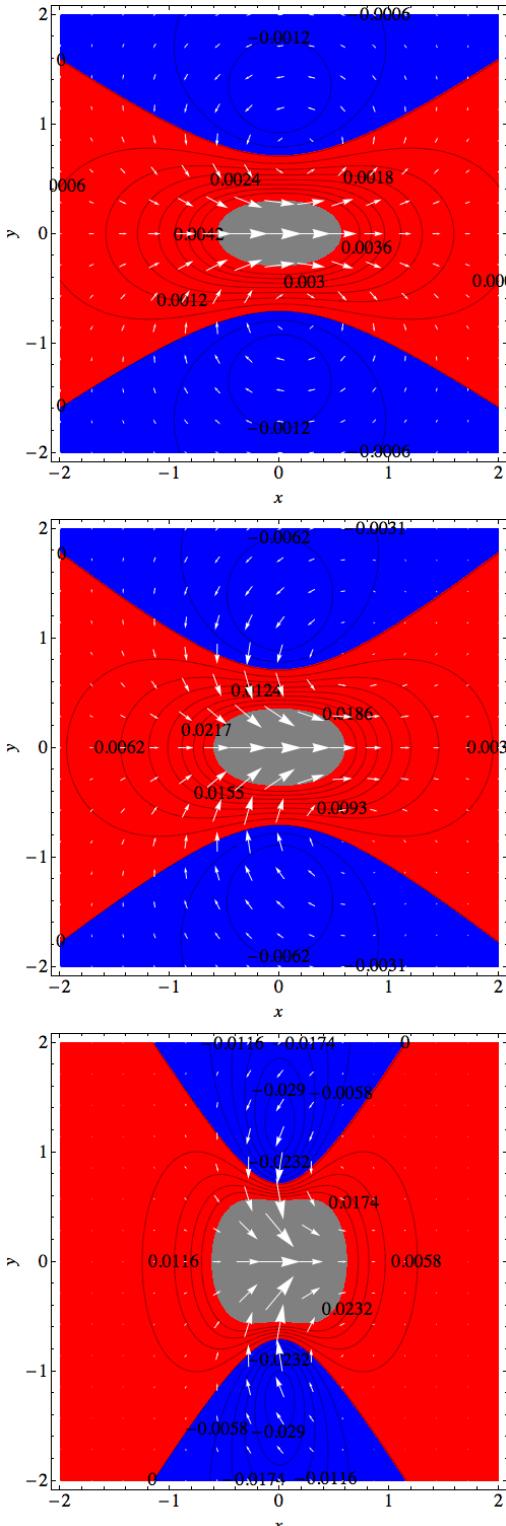


Figure 6.10: Current density vectors (white) overlaid on contours of charge density on the stretched horizon ($\alpha_{\mathcal{H}} = 10^{-4}$) of the infalling boosted dipole with rest frame magnetic moment in the y -direction. From top to bottom, the magnitude of the boost in the x -direction increases from $\beta_S = 0.1, 0.5, 0.9$. As inferred from the last of Eqs. (6.59), the magnitude of the charge density increases with β_S . Also the shape of the charge separation is squeezed in the direction of source boost as indicated by Eq. (6.60). All of the snapshots are taken at $q_{\mathcal{H}} t = 1$ and the contour labels

horizon.

Here, the light cone condition is easier to solve in Rindler coordinates. Evaluating the source at the retarded time, $t_S = t_*$, the light cone condition is

$$x^2 + y^2 + z^2 + z_S^2 - 2zz_S \cosh[g_H(t - t_*)] = 0. \quad (6.62)$$

For $z_S = \text{constant}$, we find

$$t_*(x) = t - g_H^{-1} \cosh^{-1} \left(\frac{x^2 + y^2 + z^2 + z_S^2}{2zz_S} \right) \quad (6.63)$$

and $T_* = z_S \sinh(g_H t_*)$.

Now that we have an expression for the retarded time, we can find the fields for our Rindler observer following the prescription of §6.3. For the sake of illustration, we write the Rindler fields for the specific case where $\mathbf{m}_S = m\hat{\mathbf{e}}_y$:

$$\begin{aligned} B_R^x &= -\frac{48mxyzS^2}{(r_-r_+)^5} [r_+^2 - 2zz_S] \\ B_R^y &= \frac{8mzz_S^2}{(r_-r_+)^3} + B_R^x \frac{y}{x} \\ B_R^z &= \frac{16myzS^2}{(r_-r_+)^5} [(r_-r_+)^2 + 6z^2z_S^2] + B_R^x \frac{z}{x} \\ \mathbf{E}_R &= 0 \\ r_\pm &= \sqrt{x^2 + y^2 + (z \pm z_S)^2} \end{aligned} \quad (6.64)$$

We plot the above fields in Figure 6.11 from 4 different points of view; looking down each coordinate axis and looking from a position half way between the x and y axes. From the above expressions we see that even though the Minkowski dipole is accelerated, Rindler observers see no radiation field, nor do they see any electric field at all. This is surprising since the Minkowski observers

see a Poynting flux as well as radiation^e. However, from the expression for the Rindler Poynting flux in terms of Minkowski fields (6.37), we see that the purely Minkowski term $\left[\mathbf{E}_M^{\parallel} \times \mathbf{B}_M^{\parallel} \right]$ is exactly balanced by terms due to accelerations of the Rindler observers, which account for field energy moving past them as they accelerate. Note that this is also consistent with our choice of the rest frame moments \mathbf{m}_S and \mathbf{p}_S .

As can be seen in Figure 6.11, at the horizon, the fields align themselves perpendicular to the horizon, i.e $\mathbf{B}_{\mathcal{H}} = 0$. This is a consequence of the conductor-like properties of the horizon, Eqs. (6.11), along with the ingoing wave boundary conditions, $\mathbf{B}_{\mathcal{H}} = -\hat{e}_z \times \mathbf{E}_{\mathcal{H}}$ and $\mathbf{E}_{\mathcal{H}} = \hat{e}_z \times \mathbf{B}_{\mathcal{H}}$ which are a result of choosing stationary observers to measure the fields.

Since $\mathbf{E}_R = 0$, no battery is established for the Rindler dipole. But then, no battery would be expected from this configuration given that the dipole is fixed relative to the horizon. When we introduce relative motion, as we do in the next section, we will once again see a power source generated in the form of an event-horizon battery.

6.7 Rindler Dipole Boosted Parallel to the Horizon

We would like to imagine a worldline for the source dipole that mimics a magnetized NS in orbit around a BH. The physical motion we want to represent is best imitated by a source dipole at some fixed Rindler height above the horizon z_S , but moving parallel to the horizon with some fixed Rindler velocity, $v_{S,x} = \text{constant}$, so that

$$x_S = v_{S,x} \alpha_S t_S . \quad (6.65)$$

^eNote that the field tensor (6.26) has terms which fall off as $1/|r|$ and hence generate a radiation field, as long as there is a non-zero dipole acceleration (See also (Heras 1998))

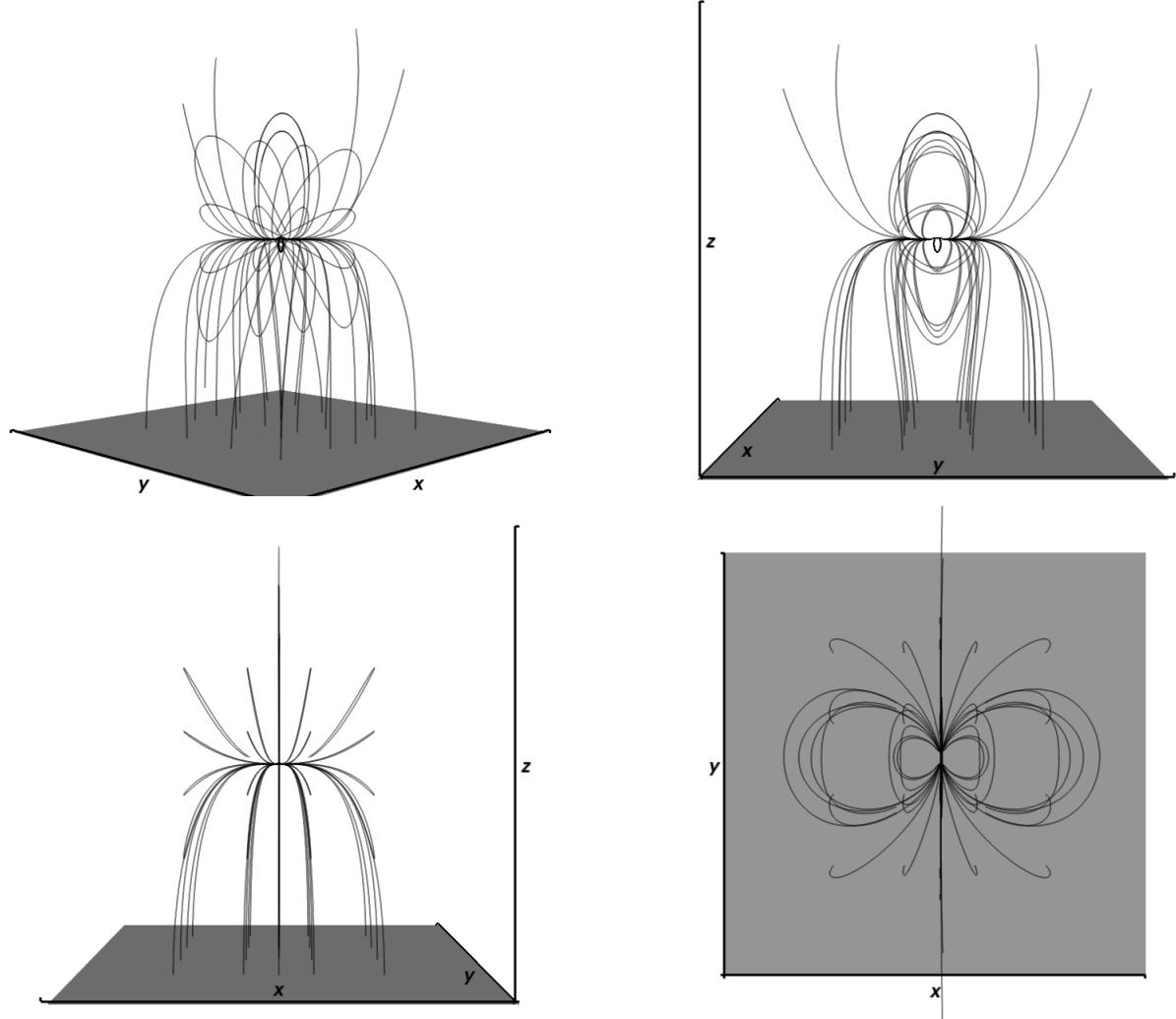


Figure 6.11: 3D visualization of the magnetic dipole field lines hovering at constant height z_S above the Rindler horizon, denoted by the gray plane at $z = 0$. The visualization region is a cube with side length $2z_S$.

The kinematic ingredients are then expressed in Minkowski coordinates as

$$r^\mu = \begin{pmatrix} T - T_S \\ X - X_S(T_S) \\ Y \\ Z - Z_S(T_S) \end{pmatrix} \quad V^\mu = \gamma_S \begin{pmatrix} 1 \\ \beta_{S,X} \\ 0 \\ \beta_{S,Z} \end{pmatrix}$$

$$a^\mu = \frac{\gamma_S^2}{Z_S} \begin{pmatrix} \beta_{S,Z} \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \dot{a}^\mu = \frac{\gamma_S^3}{Z_S^2} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \beta_{S,Z} \end{pmatrix}$$

Here γ_S is the total Lorentz factor computed with $\beta_S = \beta_{S,X}\mathbf{e}_X + \beta_{S,Z}\mathbf{e}_Z$ and

$$\beta_{S,X} = v_{S,x} \frac{z_S}{Z_S}$$

$$\beta_{S,Z} = \frac{T_S}{Z_S} \tag{6.66}$$

and again $Z_S = \sqrt{z_S^2 + T_S^2} = z_S \cosh[g_H t_S]$, with z_S the constant height of the Rindler dipole above the horizon. Notice that as long as $|v_{S,x}| \leq 1$, the source will travel slower than the speed of light at all times, $\beta_S \leq 1$.

The light-cone condition in Rindler coordinates can no longer be found in closed form for $t_S = t_*$. We can however write A^α or $F^{\alpha\beta}$ in terms of t_* (or T_*) and solve numerically for the retarded time. It is extremely helpful that we never have to take explicit derivatives of t_* since the first relation in Eq. (6.23) allows us to re-express derivatives in terms of more transparent variables.

Figure 6.13 plots the fields of a parallel-boosted dipole for the choice of a magnetic dipole moment in the y -direction and a boost in the x -direction. Each panel plots streamlines of the magnetic (blue) or electric (red) fields in the $y - z$ plane containing the source. Also plotted are

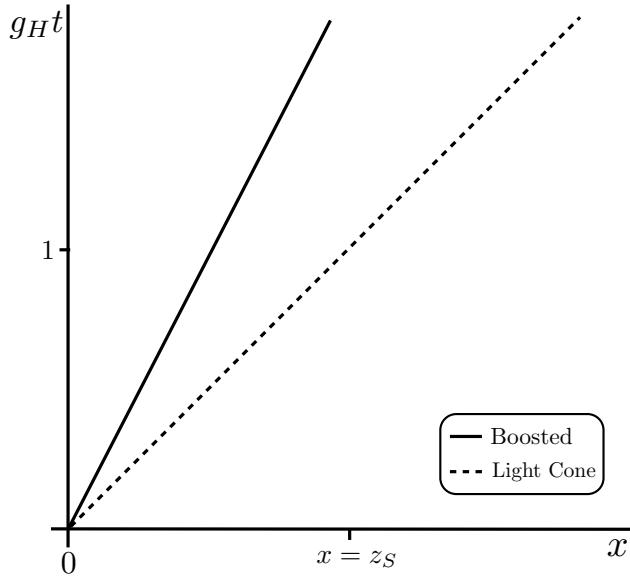


Figure 6.12: Spacetime diagram in the Rindler frame depicting the worldline of a source boosted parallel to the Rindler horizon (§6.7). The light ray (dotted line) has slope $dt/dz = (g_H z_S)^{-1}$ and the worldline has slope $(v_{S,x} g_H z_S)^{-1}$, where z_S is the constant position of the source above the Rindler horizon.

contours of $B_R^x / \sqrt{(B_R^y)^2 + (B_R^z)^2}$ or $E_R^x / \sqrt{(E_R^y)^2 + (E_R^z)^2}$ to give a sense of the 3D nature of the fields. Successive rows correspond to increases in $v_{S,x}$. The fields do not evolve in time except for their constant (universal-time) velocity motion in the x -direction.

The dipolar magnetic field structure flattens near the horizon due to time dilation. Observers below the dipole source see the dipole as it was further in the past, when the dipole was further away in the negative x -direction, than do observers the same distance above the source. This leads to an overall dragging of field lines along the horizon as explained in more detail in the figure captions.

We also see this effect in Figure 6.16 which is a slice of the Poynting-flux vector field in the $y - z$ plane containing the source. For large $v_{S,x}$, observers at small z see fields from when the dipole is relatively far away and thus do not see the dipole structure of the field energy flowing past them, only nearly uniform z and y -components.

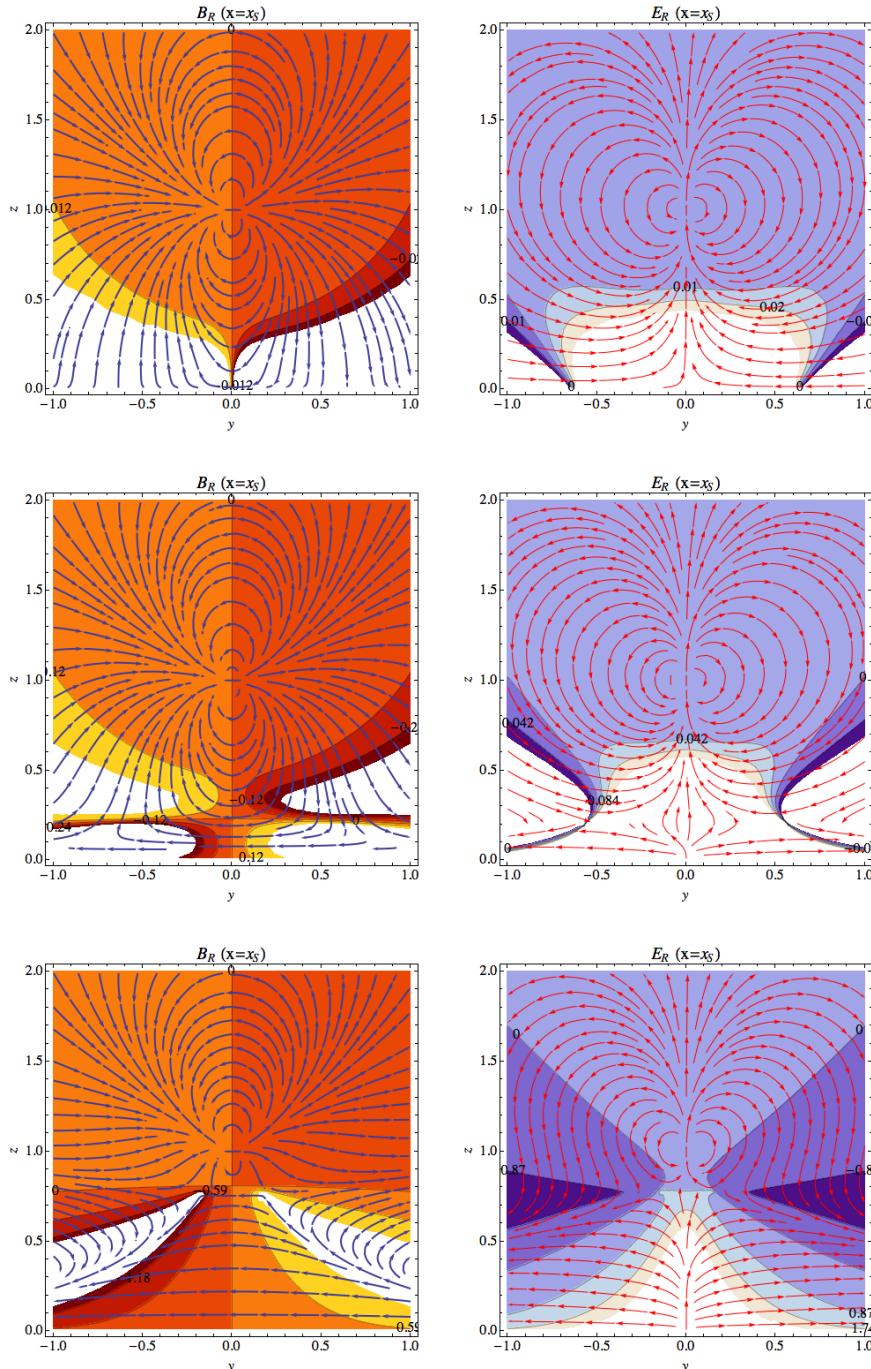


Figure 6.13: Streamlines of the magnetic (left) and electric (right) fields in the plane $x = 0$ for a magnetic dipole with dipole moment $\mathbf{m} \propto \hat{\mathbf{e}}_y$ and with three different boost velocities increasing from top to bottom $v_{S,x} = 0.1, 0.5, 0.9$ in the x -direction. Plotted over the streamlines are contours of the x -components of the fields relative to the $y-z$ magnitude. Darker regions represent negative values and lighter regions represent positive values. The white regions are clipped to better view the contour structure. The snapshots here are taken at $g_H t = 1/4$, however the fields retain the same structure for all time except for their motion in the x -direction (out of the page). The axes are in units of z_S . Since the source is boosted, observers near the horizon see the fields as they would have been when the dipole was further away in the negative x -direction. The result is an

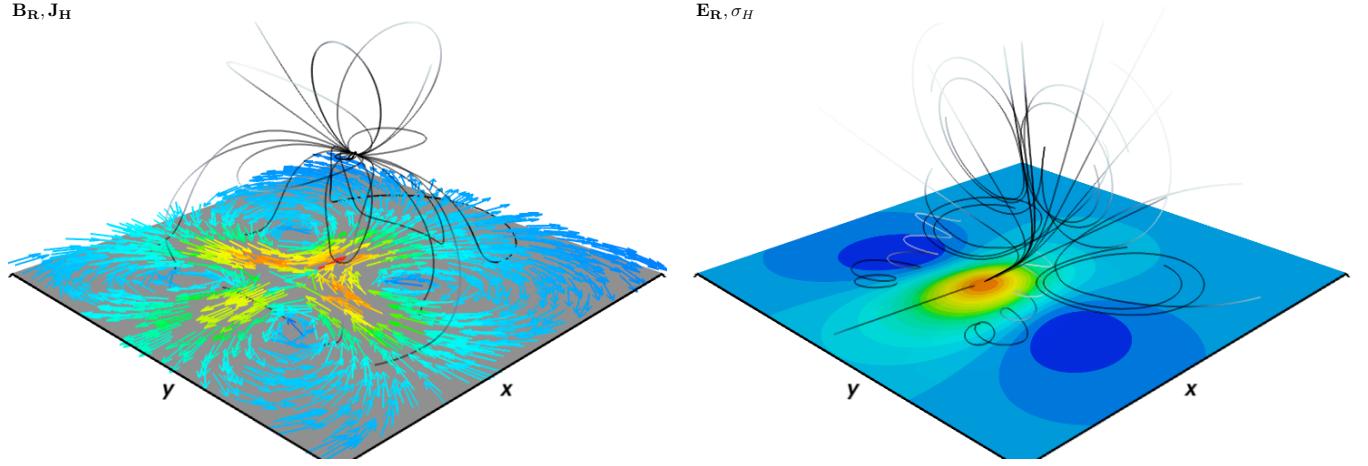


Figure 6.14: A 3D visualization of the magnetic fields lines and corresponding horizon currents $\mathbf{J}_\mathcal{H}$ (left) and electric field lines with the corresponding horizon charges $\sigma_\mathcal{H}$ (right) for the boosted Rindler dipole. The case shown is for $v_{S,x} = 0.2$.

For small $v_{S,x}$, the fields resemble those in the stationary case, threading the horizon nearly perpendicularly. As $v_{S,x}$ is increased, the dragging effect causes the fields near the horizon to lay down tangentially to the horizon as the source moves along. Figure 6.14 shows a 3D representation of the dragging effect for the $v_{S,x} = 0.2$ case. In the left panel of Figure 6.14 we see that the tangential magnetic fields source horizon currents. Via horizon charge conservation, these currents build up horizon charge density which we observe in the right panel of Figure 6.14 and interpret as the normal components of the induced electric fields. Figure 6.15 shows the horizon charge and current densities for three different $v_{S,x}$, all at $g_H t = 10$. As time progresses, these same charge and current distributions are dragged behind the dipole on the stretched horizon at a lag distance which increases as $v_{S,x}$ increases, and also as the distance between the stretched and true horizons decreases.

An interpretation of this behavior follows similarly to that of (MacDonald & Suen 1985) for the case of an electric point-charge boosted parallel to the Rindler horizon. In the electric point-charge case, the charge distribution induced on the horizon is also dragged behind the boosted

source. Via charge conservation, this necessitates horizon currents to redistribute charges. Such horizon currents can be thought of as due to tangential components of magnetic fields induced by the moving point-charge.

The key result from this example is the explicit charge separation and therefore voltage drop across the event horizon. We have established an event-horizon battery, a power source for a BH-NS electromagnetic circuit. The boosted Rindler dipole provides a proxy for a NS in orbit around a big BH. We will use this case to estimate some astrophysically relevant scales in the following section.

6.8 Consequences for the BH-NS Binary

6.8.1 Voltage, Luminosity, and Energy

We utilize the electromagnetic field solutions of the boosted, Rindler dipole of §6.7 to estimate the power output and maximum energy of radiation a BH battery can supply. We treat the BH-NS system as a series circuit containing resistors and a battery with voltage given via (6.13) from the electromagnetic field solutions. For simplicity, we imagine sticking one wire of the circuit into the point of maximum horizon potential, and the other wire, a distance of $2M$ away in the y -direction. From (6.10), this separation can be compared to to a circuit connecting the pole and equator of a Schwarzschild BH. (Although this seems arbitrary, there is little dependence on the distance. We could have stuck the other wire at infinity with little difference in results.) For the boosted Rindler dipole solutions with $\mathbf{m}_S = m\mathbf{e}_y$, the y -component of the vector potential vanishes. Then from the potential of Eq. (6.13) and the electric field Eq. (6.28) in a Rindler coordinate frame, we have $\nabla_y V_H = \alpha E_y^R = \alpha^2 \nabla_y A_R^0$, and therefore $V_H = \alpha^2 A_R^0$ across regions of charge separation estimates the voltage drop on the horizon. In terms of the Rindler retarded time, and with physical

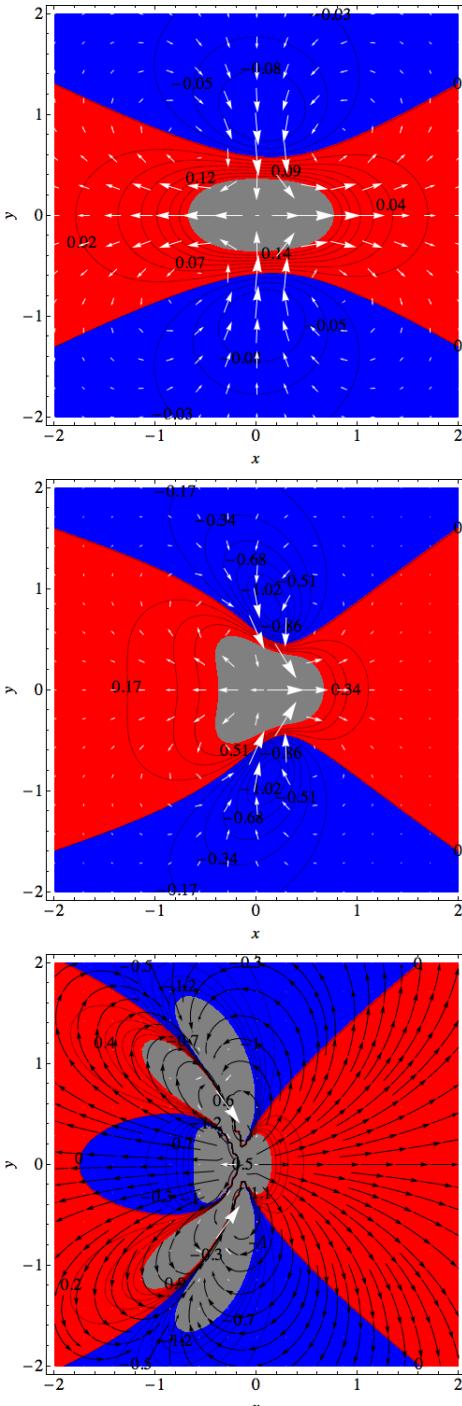


Figure 6.15: An identical plot to Figure 6.10 but for the boosted case. Current density vectors (white) are overlaid on contours of charge density on the stretched horizon ($\alpha_H = 10^{-4}$) of the boosted dipole with magnetic moment in the y -direction. The bottom panel also plots streamlines of the currents. From top to bottom, the magnitude of the boost in the x -direction increases from $v_{S,x} = 0.1, 0.5, 0.9$. Each snapshot is taken at $q_H t = 10$. The configuration drags along the stretched horizon keeping a constant lag distance behind the moving source. The induced currents can be thought of as redistributing charge in order to slide the charge distribution along behind the boosted dipole. The gray regions are regions of steeply increasing σ_H which have been removed to more clearly view the contour structure. The axes are in units of z_S .

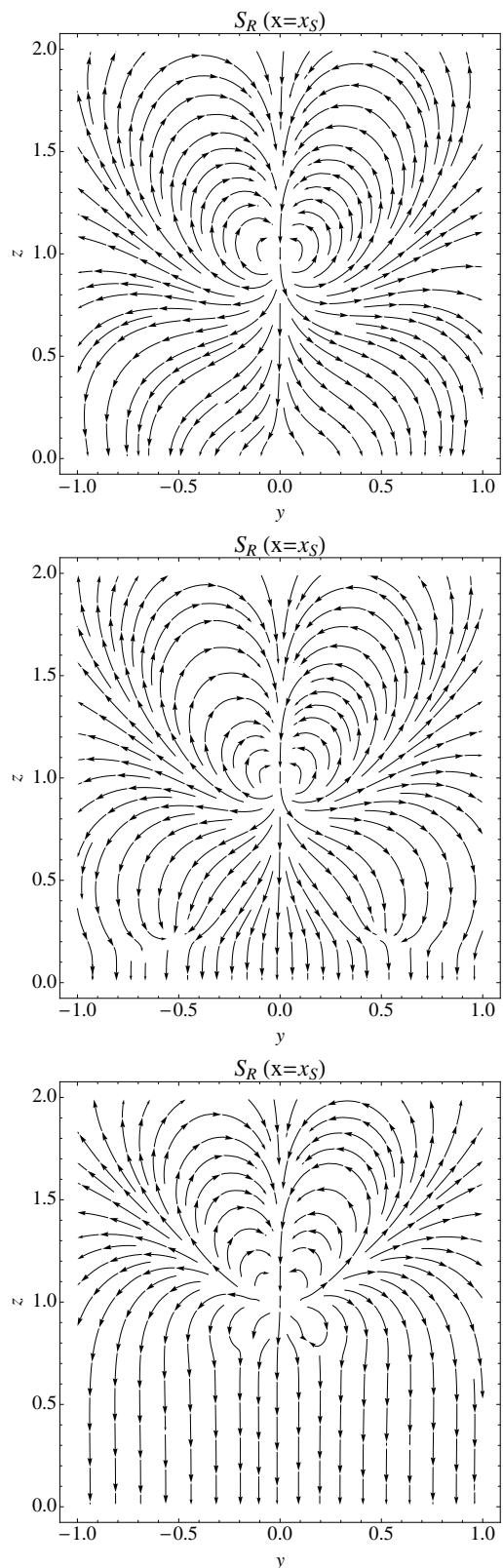


Figure 6.16: An $x = x_S = 0$ slice through the Poynting flux. The plane contains the source as viewed by Rindler observers for three different boost magnitudes in the x -direction, $v_{S,x} = 0.1, 0.5, 0.9$. The Poynting flux is 0 at infinity despite outward components of the field in the region plotted here. The axes are in units of z_S .

constants restored, the horizon voltage is,

$$V_{\mathcal{H}} = \frac{c^2 v_{S,x} z_H m}{8GM} \left(v_{S,x}^2 z_S \frac{g_H}{c} t_* - v_{S,x} x + z_H \tilde{\mathcal{S}} \right)^{-3} \times \\ \left[z_H \left(1 + 2v_{S,x}^2 + \tilde{\mathcal{C}}2 \right) + 2v_{S,x} \tilde{\mathcal{S}} \left(x - v_{S,x} z_S \frac{g_H}{c} t_* \right) - 2\tilde{\mathcal{C}} \left(\frac{x^2 + v_{S,x}^2 z_S^2 + z_H^2}{z_S} + v_{S,x} \frac{g_H}{c} t_* \left(v_{S,x} z_S \frac{g_H}{c} t_* - 2 \right) \right) \right] \quad (6.67)$$

with

$$\tilde{\mathcal{C}} = \cosh \left[\frac{g_H}{c} (t - t_*) \right], \quad \tilde{\mathcal{S}} = \sinh \left[\frac{g_H}{c} (t - t_*) \right], \quad \tilde{\mathcal{C}}2 = \cosh \left[2 \frac{g_H}{c} (t - t_*) \right]$$

where m is the NS rest frame dipole moment (See (6.68)) and the retarded time t_* is a function of the (Rindler) observer coordinates.

We compute power radiated by such a circuit from (7.5). To do so, we estimate the physically relevant values of the various parameters. Very near the Schwarzschild horizon, in physical units, the gravitational acceleration is

$$g_H = \frac{c^4}{4GM} \simeq 1.5 \times 10^{14} \left(\frac{10M_\odot}{M} \right) \frac{\text{cm}}{\text{s}^2}$$

about 100 billion times that on Earth for a $10M_\odot$ black hole. The magnitude of the NS's magnetic dipole moment written in terms of the magnetic field strength at the NS's poles B_p and the radius of the NS R_{NS} is of order,

$$m = \frac{B_p R_{NS}^3}{2} \simeq 5 \times 10^{29} \left(\frac{B_p}{10^{12} G} \right) \left(\frac{R_{NS}}{10^6 \text{cm}} \right)^3 \text{G cm}^3. \quad (6.68)$$

We must also approximate the resistances in our astrophysical circuit diagramed in Figure 6.1. We have three resistors to consider: the horizon with resistance \mathcal{R}_H ^f, the NS crust with resistance \mathcal{R}_{NS} and the plasma of the NS magnetosphere denoted by $\mathcal{R}_{\text{plasma}}$. The horizon resistance is

^fA material with resistivity ρ has resistance $\mathcal{R} = \rho \frac{L}{A}$ where L and A are the length and cross sectional-area of the material as seen by the current. In the case of a black hole horizon, A and L can be taken to both be of order $\pi 2M$.

known from the membrane paradigm to be [Thorne et al. \(1986\)](#)

$$\mathcal{R}_H \simeq \frac{4\pi}{c} = 4.2 \times 10^{-10} \text{s cm}^{-1} = 377\Omega . \quad (6.69)$$

The resistivity of the NS crust is likely very small compared to \mathcal{R}_H , on the order of 10^{-24}s^{-1} (see *e.g.* [\(Piro 2012\)](#)); thus we set $\mathcal{R}_{NS} = 0$.

The value of $\mathcal{R}_{\text{plasma}}$ is an interesting unknown and requires numerical exploration beyond the scope of this article. For our present purposes, as a rough guide, we choose an effective value of $\mathcal{R}_{\text{plasma}} = \mathcal{R}_H/2$, because it gives maximum power output through $\mathcal{R}_{\text{plasma}}$. Ref. [McWilliams & Levin \(2011\)](#) choose $\mathcal{R}_{\text{plasma}}$ based on equating the power dissipated due to curvature radiation with the power dissipated due to ohmic dissipation $I^2\mathcal{R}_{\text{plasma}}$. Upon solving for $\mathcal{R}_{\text{plasma}}$, they find $\mathcal{R}_{\text{plasma}} = \mathcal{R}_H$ when the plasma velocity is $\sim 0.7c$. However, the estimate is sensitive to the plasma particle velocity. The dependence of power output on $\mathcal{R}_{\text{plasma}}$ for a similar NS-NS circuit with non-zero \mathcal{R}_{NS} is explored in [Piro \(2012\)](#), however the NS-BH case is simpler since the denominator of the power formula (7.5) is dominated by \mathcal{R}_H .

Since we have set the NS resistance to 0, we focus on the power radiated in the space between the NS and BH, *i.e.* $\mathcal{R}_i = \mathcal{R}_{\text{plasma}}$ in (7.5). Since the horizon potential is symmetric around the line $y = y_S$, which contains the maximum of the potential and thus one of the circuit wires, we multiply the above luminosity by a factor of 2. The combination of our choices for \mathcal{R}_H and $V_{\mathcal{H}}$ will correspond to maximum achievable bolometric luminosities when \mathcal{R}_{NS} is ignored.

The circuit is connected if the BH is within the light cylinder of the NS:

$$\begin{aligned} R_{lc} &= \frac{c}{\Omega_{NS}} = 5 \times 10^9 \left(\frac{P}{1s} \right) \text{cm} \\ &\sim 3 \times 10^3 \left(\frac{P}{1s} \right) \left(\frac{10M_{\odot}}{M} \right) \frac{GM}{c^2} \end{aligned} \quad (6.70)$$

where P is the period of the NS spin and in the last line we quote the radius in units of M . We choose a fiducial horizon distance of $z_S = 3GM/c^2$ where the Rindler limit is valid and the pair has approached extremely close prior to merger. For BH's with $M \gtrsim 10^4 M_\odot$ our fiducial value of z_S is larger than the light cylinder of the NS and the pair is unplugged. We address the case for these larger black holes in the next section. For lighter BHs, the circuit will connect when the NS is a distance above the horizon $\sim R_{lc}$ and the power supplied will grow until it reaches a maximum around our fiducial distance $z_s = 3GM/c^2$ just prior to merger. Realistically, the compact objects will plunge extremely rapidly at such close separations so we only use these values to get a sense of the maximum blast of luminosity. For a $10M_\odot$ BH, $R_{lc} \gg z_S$ and so the circuit is connected for many orbits before maximum is reached.

The maximum horizon voltage and corresponding luminosities are plotted for $10M_\odot$, $10^2 M_\odot$, and $10^3 M_\odot$ BH's in Figure 6.17 for $v_{S,x}$ varying from 0.01 to 0.95. For comparison, $v_{S,x} \sim 0.5$ at the last stable circular Schwarzschild orbit.

Finally then we have our answer. We estimate the voltage of our BH battery to be $\sim 10^{16}$ statvolts for a $10M_\odot$ BH, $\sim 10^{14}$ statvolts for a $10^2 M_\odot$ BH, and $\sim 10^{12}$ statvolts for a $10^3 M_\odot$ BH when $v_{S,x} \sim 0.5$. The luminosities in this limited approximation are $\sim 10^{42}$ erg/s, 10^{38} erg/s, and 10^{34} erg/s respectively.

As is suggested by the pre-factor in (6.67), the horizon voltage and the luminosity decrease with increasing BH mass. For $z_S \rightarrow 0$, the horizon voltage scales as M^{-1} and the luminosity as M^{-2} . Otherwise, terms proportional to g_H inside the brackets in Eq. (6.67) dominate and thus the voltage goes as M^{-2} causing the luminosity to scale as M^{-4} . Comparison of the three panels of Figure 6.17 confirms this scaling with BH mass.

This scaling also agrees with (McWilliams & Levin 2011) where the BHNS battery was first proposed. In (McWilliams & Levin 2011) the physical mechanism was sketched out in a non-relativistic calculation. The analytic, relativistic solutions obtained here are in agreement with the

findings of (McWilliams & Levin 2011). Specifically, (McWilliams & Levin 2011)'s Eq. (6) for \mathcal{L} exhibits the same scaling with black hole mass and NS magnetic field strength as does our Eq. (6.1). Also, (McWilliams & Levin 2011)'s Eq. (6) is calculated for similar parameters that we use in our calculation. They choose $\mathcal{R}_{\text{plasma}} = \mathcal{R}_H$, $\mathcal{R}_{NS} = 0$, $B_p = 10^{12}$, $M = 10M_\odot$, and a separation corresponding to the NS orbiting at the light ring of a Schwarzschild BH. Hence we may also compare magnitudes of the computed luminosities in each study, and we find that they agree in order of magnitude. Note that the inverse BH mass dependence arises because we are comparing luminosities for different BH masses while holding the distance of the dipole source from the horizon at a fixed number of gravitational radii (which scales with M).

We can also compute the maximum energy given to magnetosphere particles by the horizon battery. To be clear, we are not calculating the spectrum – which promises to be complicated – just the maximum energy scale. The magnitude of horizon voltages plotted in Figure 6.17 makes evident that the highest energy particles accelerated via the horizon battery will radiate their energy via curvature radiation. A Rindler observer at the instantaneous location of an accelerating plasma particle will measure a local energy given by the characteristic energy of curvature radiation,

$$\epsilon_R = \frac{3hc}{4\pi} \frac{\gamma_p^3}{\eta R_{LC}} \quad (6.71)$$

where h is Planck's constant, γ_p is the Lorentz factor of the plasma particle (electron or positron) measured by a Rindler observer,^g and we have parameterized the radius of curvature of a magnetic field line by a constant η times the NS light cylinder radius. We choose $\eta = 0.1$ throughout. The energy measured by an observer at infinity is found by multiplying by a factor of $\alpha = (g_H/c^2)z = (4GM/c^2)^{-1}z$, which accounts for the gravitational redshift, which is to be evaluated at the Rindler

^gThe Lorentz factor in units of Rindler proper time are related to the Lorentz factor in units of universal time by $\gamma_p(\tau_R) = \gamma_p(t)/\alpha$.

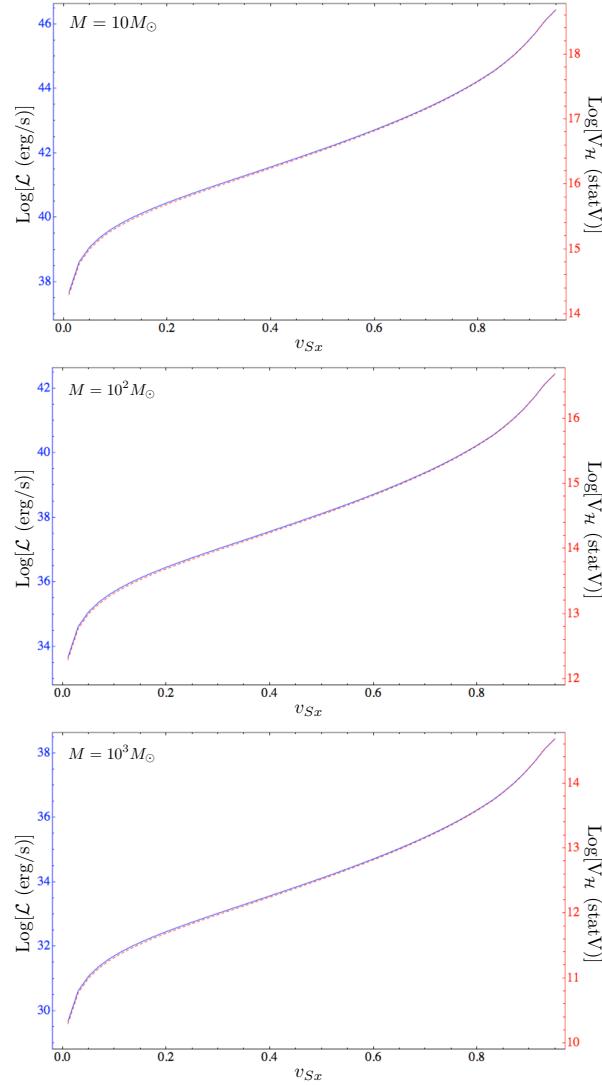


Figure 6.17: Log-luminosity computed from Eq. (7.5) (blue, solid line and leftmost y-axis labels) and representative log-voltage drop on the horizon (red, dashed line and rightmost y-axis labels). Luminosities and voltages are computed for $M = 10M_\odot$, 10^2M_\odot , and 10^3M_\odot with the dipole at Rindler height $z_S = 3M$ as a function of $v_{S,x}$ varying from 0.01 to 0.95. The last stable circular orbit in the Schwarzschild spacetime would have $v_{S,x} = 0.5$.

z coordinate of emission.

$$\epsilon_\infty = \alpha \frac{3hc}{4\pi} \frac{\gamma_p^3}{\eta R_{LC}} \quad (6.72)$$

Keep in mind however, that far enough from the horizon, where the Rindler limit to the Schwarzschild spacetime breaks down ($\alpha \gtrsim 1$), Rindler- α does not predict the correct gravitational redshift. So, we can only use Eq. (6.72) for emission that originates close to the horizon, consistent with the regime in which we are working.

We next solve for the values of γ_p . For a given BH mass and horizon distance z_S as a function of dipole boost $v_{S,x}$, we estimate the maximum γ_p in the radiation reaction limit, in which the rate of energy gain from the horizon battery is balanced by the rate of energy loss due to curvature radiation.

A Rindler observer at the instantaneous location of an accelerating plasma particle will measure the following energy per unit proper time being radiated from the particle due to dipole radiation,

$$\mathcal{P} = \frac{d\epsilon_R}{d\tau_R} = \frac{2}{3} e^2 c \frac{\gamma_p^4}{(\eta R_{LC})^2}. \quad (6.73)$$

Eq. (6.73) is the standard relativistic Larmor formula for the power.

Then for a plasma particle moving on the path $\mathbf{s}(t)$, the radiation reaction limited γ_p is given by,

$$e \frac{dV}{ds^i} \frac{ds^i}{d\tau_R} = \frac{2}{3} e^2 c \frac{\gamma_p^4}{(\eta R_{LC})^2}, \quad (6.74)$$

where use of the locally observed potential, $V = V_H/\alpha$, is justified since we are only considering an infinitesimal potential difference, not a global value.

Upon inspection of the currents in Figure 6.15, it is apparent that representatively large horizon

electric fields exist at $y = 0$ in the $\pm x$ direction. Thus we choose $ds = dx\mathbf{e}_x$ which allows us to write

$$|E_R^x| \left(1 - \frac{1}{\gamma_p^2}\right)^{1/2} = \frac{2}{3} e \frac{\gamma_p^4}{(\eta R_{LC})^2} \quad (6.75)$$

where we have written the 3-velocity of the particle (assumed to be only in the x direction) in terms of γ_p . Note that it is E_R^x and not αE_R^x which should be on the LHS of (6.75) because qE_R^x is the rate of change of momentum as viewed by Rindler observers and we are asking the Rindler observer to locally balance the competing sources of momentum loss and gain.

Solving the above equation for γ_p then gives the radiation reaction limited Lorentz factor, as observed by Rindler observers, as a function of time. Since, however, the fields are stationary in the frame which drags along with the horizon charges, we need only find the maximum γ_p at any time and choose that as our fiducial maximum γ_p . Substituting this into (6.72) and evaluating α at the same z position as we evaluated (6.75), gives the maximum energy due to curvature radiation that the horizon battery can produce at a given z_S , $v_{S,x}$, and M , according to observers at infinity. In practice we find the largest values of ϵ_∞ when evaluating (6.75) and (6.72) at the stretched horizon, although varying the point of evaluation from $z = z_H$ up to $z = 4M$ changes the result for ϵ_∞ by less than an order of magnitude. Figure 6.18 plots the maximum γ_p and ϵ_∞ as a function of $v_{S,x}$ at dipole height of $z_S = 3GM/c^2$ for BH masses $M = 10M_\odot$, $10^2 M_\odot$, and $10^3 M_\odot$.

We estimate maximum γ_p 's of our BH battery to be $\sim 10^{10.2}$ for a $10M_\odot$ BH, $\sim 10^{9.5}$ for a $10^2 M_\odot$ BH, and $\sim 10^{8.7}$ for a $10^3 M_\odot$ BH when $v_{S,x} \sim 0.5$. However, recall that these are the Lorentz factors measured by the Rindler observers at the stretched horizon and do not correspond to the tremendous energies which (6.71) would imply. It is the energy measured at infinity given by (6.72) which carries the only physical relevance here. These γ_p 's correspond to maximum curvature radiation energies at infinity of approximately 30 TeV, 100 GeV, and 1 GeV respectively at

$v_{S,x} = 0.5$. Because of the decrease in horizon voltage for larger mass BH's, as for the luminosity, the curvature radiation energies are smaller for larger mass BH's.

We reiterate that the radiation energies plotted here represent the highest energies of radiation that could be emitted by the NS-BH circuit. Firstly this is because we are not accounting for any plasma effects which may act to screen the maximum fields quoted here. In addition to this, we are using the radiation reaction limited γ_p computed for plasma particles with velocities aligned with the largest values of the electric field across the horizon. There will also be a spectrum of lower energy synchro-curvature radiation not calculated here.

Although we have not computed timescales or detailed spectra of emission we note that with luminosities reaching up to 10^{42} erg/s (10^{48} erg/s for magnetars) and with the capability of producing photons with energies reaching into the TeV range, the mechanism discussed here for a BH mass of $10M_\odot$ could be capable of producing bursts of gamma-rays. Further investigation of this mechanism and the timescale, as well as any variability, of emission is needed in order to say whether the BH-NS circuit is responsible for previously detected high-energy bursts, or rather, if it is responsible for an as of yet unobserved phenomenon.

We now look closer at the larger BH case, where the Rindler limit is an even better proxy for the physical situation.

6.8.2 NS plummet into a SMBH

Although motivated by an interest in stellar mass BHs, the solutions we have found in the Rindler limit well approximate the end of a NS's plummet into an intermediate mass or super-massive black hole (IMBH, SMBH). We have included analysis for IMBH's in the previous section, here we consider a SMBH. Recall that for the mechanism to operate, the BH horizon must be within the magnetosphere of the NS; the distance of the NS from the horizon must be less than the light cylinder radius (6.70) of the NS. In the previous subsection we always had that $R_{lc} \geq z_S = 3M$.

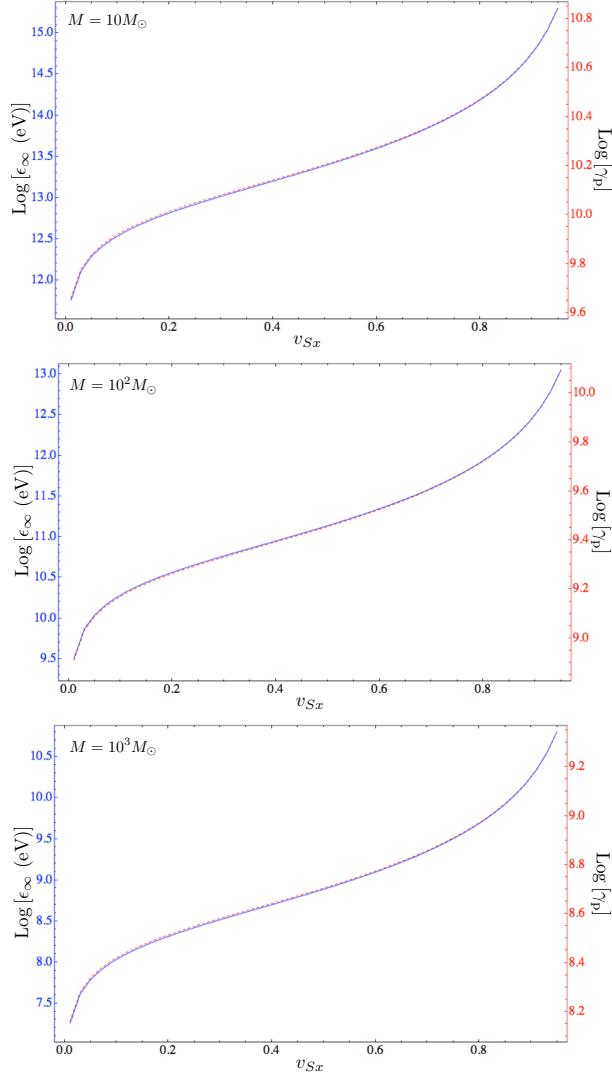


Figure 6.18: Maximum curvature radiation energies computed from Eq. (6.72) (blue, solid line and leftmost y-axis labels) and corresponding maximum γ_p (Eq. 6.75) to which electrons/positrons can be accelerated (red, dashed line and rightmost y-axis labels). Both are computed for $M = 10M_\odot$, $10^2 M_\odot$, and $10^3 M_\odot$ with the dipole at Rindler height $z_S = 3M$ as a function of $v_{S,x}$ varying from 0.01 to 0.95.

For the SMBH case however, the NS light cylinder is smaller than $3M$. Thus we locate the dipole at $z_s = R_{lc}$ so that the circuit is connected. In this case the Rindler approximation is good for the entire time that luminosity can be generated by the BH battery.

In Figure 6.19, we plot the luminosities and energy of curvature radiation that could be generated by the dipole at a distance R_{lc} from a $10^6 M_\odot$ BH horizon. We find that for $v_{S,x} = 0.5$, luminosities of order 10^{26} erg/s can be achieved by the SMBH-NS circuit. The maximum γ_p 's are still rather large reaching values of $\sim 10^{7.4}$ at $v_{S,x} = 0.5$. However, recall that these are the maximum γ_p 's as measured by Rindler observers at the stretched horizon and so observed energies at infinity are reduced by a factor of $\alpha_H = 10^{-4}$ from what would be inferred from γ_p alone. For $v_{S,x} = 0.5$ the SMBH-NS circuit could generate energies of curvature radiation peaking in the X-ray at ~ 100 keV.

At luminosities of $\sim 10^{26}$ erg/s and peak radiation energies of ~ 100 keV, even if the SMBH were in our own Galactic Center, this signal would be difficult to detect, as it emanates from a noisy galactic nucleus and could be beamed in a direction not guaranteed to intersect Earth. Note however, that in the optimal case of a magnetar with $B_p \simeq 10^{15} G$, and a slower spin period of $\sim 10 s$, the circuit would be connected at a $10\times$ greater distance from the horizon and emit at a peak luminosity of $\sim 10^{32}$ erg/s. Such events, if beamed in our direction may produce a short, if faint, X-ray burst coming from the Galactic-Center. We can put a type of upper limit on the length of such a magnetar-SMBH X-ray burst by noting that the infall time observed at infinity for the NS falling from R_{lc} to a R_{NS} at the speed of light is of order a minute. However, the energetics of the magnetosphere could limit any emission to a much shorter interval. For comparison, X-ray flares at the Galactic Center are observed with durations of order an hour and X-ray luminosities of $\sim 10^{35}$ erg/s in the energy range $2 - 10$ keV (Degenaar et al. 2012).

NS-BH systems with BH mass in the range $10^3 - 10^5 M_\odot$ can also be accurately described by the Rindler limit and could reside nearby within globular clusters in the halo of our galaxy (see

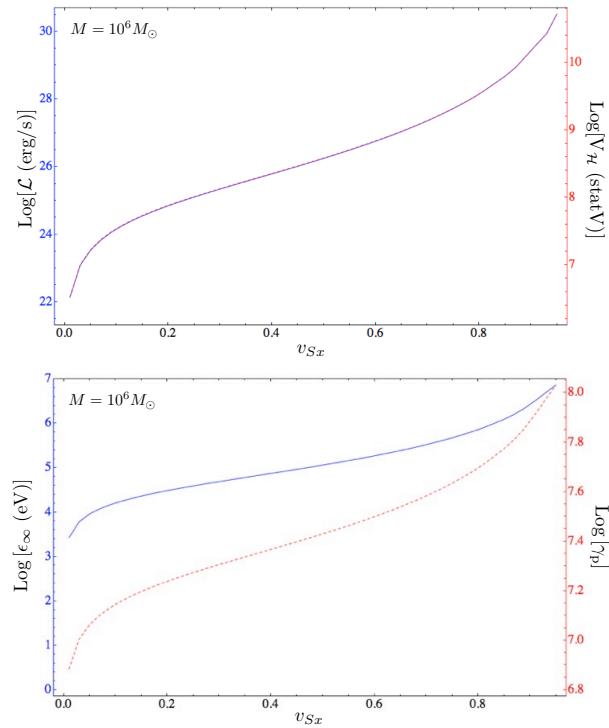


Figure 6.19: Luminosity, voltage (top) and energy, Lorentz-factor (bottom) plots identical to those portrayed in Figures 6.17 and 6.18 respectively. Here we have plotted both panels for a $10^6 M_\odot$ BH and at a much smaller horizon distance than in the previous Figures: $z_S = R_{lc}$, corresponding to the maximum separation where the BH-NS circuit remains connected.

e.g. (Lützgendorf et al. 2012)). The bottom panels of Figures 6.17 and 6.18 show that for a $10^3 M_\odot$ BH the BH-NS system could generate luminosities of order 10^{34} erg/s peaking at a maximum achievable energy of radiation at a few GeV.

To determine whether such a signal from a NS-IMBH binary would be detectable with currently operating instruments we consider flux sensitivities of the SWIFT Burst Alert Telescope (BAT) (Barthelmy et al. 2005) and the FERMI Gamma-ray Burst Monitor (GBM) (FER 2009) which are well suited for observing such transient high-energy events. From the BAT flux sensitivity, $\sim 10^{-8}$ ergs $\text{cm}^{-2}\text{s}^{-1}$ and the GBM trigger rate 0.6 photons $\text{cm}^{-2}\text{s}^{-1}$ respectively, we may compute the minimum flux over the instrument energy range needed to detect a NS plunge event with luminosity computed from our model in the previous section. With this we can calculate a maximum observable distance for which our NS-BH circuit signal would be detectable. Assuming that the radiation is beamed into a solid angle $\Delta\Omega = 100 \text{ deg}^2$, taking a photon index of 5/3 for curvature radiation, and integrating over the energy range of the instrument (15 to 150 KeV for the BAT and 150 KeV to 40 MeV for the GBM) we find

$$D_{\max} \simeq 3.8 \text{ Kpc} \sqrt{\left(\frac{\mathcal{L}}{1.3 \times 10^{34} \text{erg/s}} \right) \left(\frac{100 \text{deg}^2}{\Delta\Omega} \right)}$$

SWIFT BAT 15 → 150 KeV

or

$$D_{\max} \simeq 0.4 \text{ Kpc} \sqrt{\left(\frac{\mathcal{L}}{1.3 \times 10^{34} \text{erg/s}} \right) \left(\frac{100 \text{deg}^2}{\Delta\Omega} \right)}$$

FERMI GBM 150 KeV → 40 MeV

where for the luminosity we have used the $v_{S,x} = 0.5$ value for a $10^3 M_\odot$ mass black hole system (See Figure 6.17). Note that the above distances scale directly with the NS magnetic field strength.

For a magnetar, maximum observable distances are on order a Mpc. Since Galactic globular clusters exist within a few Kpc of Earth such NS-IMBH inspirals could be observationally interesting events if the mechanism for EM radiation discussed here operates and if IMBH's exist in globular clusters.

6.9 Conclusions

When a magnetized NS and a BH approach within the NS light cylinder, an electromagnetic circuit is established. In the Rindler limit, this corresponds to a magnetic dipole boosted parallel to the flat-wall horizon. The power supplied to the circuit will increase as the pair draws closer, reaching a maximum just before merger. The maximum voltage the battery attains and the maximum luminosities powered at this final stage scale roughly as

$$\begin{aligned} V_{\mathcal{H}}^{\max} &\simeq 3.3 \times 10^{16} \left(\frac{B_p}{10^{12} \text{ G}} \right) \left(\frac{M}{10M_{\odot}} \right)^{-2} \text{ statvolts} \\ \mathcal{L}^{\max} &\simeq 1.3 \times 10^{42} \left(\frac{B_p}{10^{12} \text{ G}} \right)^2 \left(\frac{M}{10M_{\odot}} \right)^{-4} \frac{\text{erg}}{\text{s}} \end{aligned} \quad (6.76)$$

$$(z_S = 3M, \quad v_{S,x} = 0.5, \quad \mathcal{R}_{\text{plasma}} = \mathcal{R}_{\mathcal{H}}, \quad \mathcal{R}_{NS} = 0) .$$

The scaling changes if the NS does not maintain the fixed height of $3M$ above the horizon and depends as well on the unknowns $\mathcal{R}_{\text{plasma}}$ and \mathcal{R}_{NS} . The estimated maximum could be higher when BH and NS spins are included. NS spin can be thought of as increasing the effective $v_{S,x}$. BH spin adds extra power from the analogue of the BZ effect.

There are many caveats to consider when formulating observational features of an event-horizon battery, such as potential short circuits in the system. Charges from the NS and its surrounding magnetosphere can act to screen the induced electric fields. In addition to these charges, if both the horizon voltage and the magnetic field strength are large enough, pair production could

become an important source of screening charges. Hence, the structure of the NSBH magnetosphere needs to be investigated further in order to determine the viability of an event-horizon battery powered electromagnetic signal. Another concern is that at such high voltages the current generated along the magnetic field lines would be so great that the magnetic fields induced exceed those of the original dipole.^h However, (Lai 2012) has shown that this effect should not be large enough to short out the circuit for a NS-BH system due to the large resistance of the horizon.

It would be essential to pin down the timescales of the various emission mechanisms associated with this phenomenon, although the solutions presented here give us no special advantage in doing so. Numerical results are needed to carefully characterize this EM signal in greater detail – although we can conjecture that there are potentially several distinct channels: 1) a brief jet, 2) beamed synchrotron and curvature radiation that sweeps across the sky, 3) a faint hot spot as charged particles hit the NS pole.

These caveats aside, in light of this analysis we can say that BH-NS binaries with BH's of order 10's of M_{\odot} could conceivably produce luminosities of order 10^{42} erg/s (10^{48} erg/s if the NS is a magnetar) and emit high-energy gamma rays, possibly consistent with a sub-class of gamma ray bursts. Therefore, stellar mass BH-NS binaries detectable by AdLIGO, could power high-energy electromagnetic radiation, possibly into the TeV range, detectable moments prior to the gravitational radiation burst at merger. Discovery of these important pairs could probe NS properties as well as population rates in the pre-AdLIGO era. Also intriguing is the possibility of an IMBH in a binary with a highly magnetized NS. Although less energetic, their emissions may nonetheless be detectable.

Acknowledgements

We would like to thank Jules Halpern and Sean McWilliams as well as participants of the KITP

^hShort circuits could turn the mechanism off temporarily until the current builds up again. A possible signature of this short circuit transient might be repeated spikes in the emissions

“Rattle and Shine” conference (July 2012) for useful discussions. We also thank the anonymous referee for useful suggestions on improving the manuscript. 3D visualizations of field lines were computed using Mayavi (?). This research was supported by an NSF Graduate Research Fellowship Grant No. DGE1144155 (DJD), NSF grant AST-0908365 (JL), a KITP Scholarship under Grant no. NSF PHY05-51164 (JL), and a Guggenheim Fellowship (JL).

6.10 Detailed Solution to the Field Equations

In Minkowski spacetime, Maxwell's equations for the 4-potential A^α are,

$$\square A^\alpha(x) - \partial^\alpha (\partial_\beta A^\beta) = \frac{4\pi}{c} J^\alpha(x) \quad (6.77)$$

where $J^\alpha(x)$ is the 4-current as a function of the coordinates and we retain factors of c in the appendix. Working in the Lorentz Gauge $\partial_\beta A^\beta = 0$, Maxwell's equations become sourced wave-equations,

$$\square A^\alpha(x) = \frac{4\pi}{c} J^\alpha(x). \quad (6.78)$$

The solution for A^α can be written in terms of the retarded (or advanced) Green's function given by,

$$\square_x G(x, \bar{x}) = \delta^{(4)}[x - \bar{x}] \quad (6.79)$$

Where x is the observer spacetime coordinates, and \bar{x} is the spacetime position 4-vector to be integrated over. The above equation shows that $G(x, \bar{x})$ must depend on x and \bar{x} only via the 4-vector $x - \bar{x}$, so we write the retarded Green's function as $G(x - \bar{x})$ and solve Eq. (6.79) to find (Jackson (1991)),

$$G(x - \bar{x}) = \frac{1}{2\pi} H(x^0 - \bar{x}^0) \delta[(x - \bar{x})^2] \quad (6.80)$$

where the Heaviside function H picks out the retarded as opposed to the advanced Green's function. We may then write the solution to Eq. (6.78),

$$A^\alpha = \frac{4\pi}{c} \int G(x - \bar{x}) J^\alpha(\bar{x}) \sqrt{-g} d^4 \bar{x} \quad (6.81)$$

where, in Cartesian coordinates, the metric determinant $g = -1$. A choice of source distribution for the 4-current in (6.81) gives the 4-potential from which the fields may be computed.

6.10.1 Point Charge

Although the derivation of an electric point charge can be found in a standard text on electrodynamics (we follow [Jackson \(1991\)](#) below), we include the derivation here to better elucidate, and put in context, the derivation for the dipole to follow.

The 4-current in terms of the 4-position $x_S^\mu(\tau)$ of a point charge q with arbitrary 4-velocity V^α is

$$J^\alpha(x) = qV^\alpha(\tau)\delta^{(4)}[x - x_S(\tau)]d\tau \quad (6.82)$$

where τ is the proper time of the source charge. Substituting Eq. (6.82) and (6.80) into (6.81) yields,

$$A^\alpha = 2q \int H(x^0 - \bar{x}^0) \delta[(x - \bar{x})^2] V^\alpha(\tau) \delta^{(4)}[\bar{x} - x_S(\tau)] d^4\bar{x} d\tau \quad (6.83)$$

Integrating over the volume,

$$A^\alpha = 2q \int H(x^0 - x_S^0(\tau)) \delta[(x - x_S(\tau))^2] V^\alpha(\tau) d\tau. \quad (6.84)$$

To evaluate this we use the rule,

$$\delta[f(x)] = \sum_i \frac{\delta(x - x_i)}{|(\partial f / \partial x)_{x=x_i}|} \quad (6.85)$$

where the sum is over the i^{th} root of $f(x)$. This allows us to write,

$$\delta \left[(x - x_S(\tau))^2 \right] = \frac{\delta(\tau - \tau_*)}{\left| -2 r_\mu(\tau) V^\mu(\tau) \right|_*} \quad (6.86)$$

where, as in the text, $r^\mu = x^\mu - x_S^\mu(\tau)$ and τ_* is the proper time of the point charge given by the light cone condition,

$$r_\mu r^\mu|_* = 0 \quad (6.87)$$

Using (6.86) to simplify (6.84), we find,

$$A_{LW}^\alpha = \left. \frac{qV^\alpha(\tau)}{V_\mu(\tau)r^\mu(\tau)} \right|_{\tau_*}. \quad (6.88)$$

which are the Lienard-Wiechert Potentials for a moving point charge. Note that all quantities are evaluated at the retarded point $\tau_*(x)$, the proper time at which the source coordinates are coincident with the past light cone of the observer at x . Transforming these to Rindler space (lower case or primed), we find

$$\begin{aligned} A^{\alpha'} &= L_\alpha^{\alpha'} A_{LW}^\alpha = L_\alpha^{\alpha'} \left[\frac{qL_{\beta'}^\alpha V^{\beta'}}{r^\beta L_\beta^{\mu'} V_{\mu'}} \right]_{\tau_*} \\ &= \frac{qL_{\alpha'}^{\alpha'} L_{*\beta'}^\alpha V_*^{\beta'}}{V_{*\mu'} L_{*\beta'}^{\mu'} (x^\beta - x^{\beta}_*)}. \end{aligned} \quad (6.89)$$

as found in [MacDonald & Suen \(1985\)](#). Here $L_\alpha^{\alpha'}$ denotes the transformation matrix $\frac{\partial x^{\alpha'}}{\partial x^\alpha}$. The potentials are written in terms of the vectors at the retarded point, (subscript * above) and the observer point coordinates (no subscript). To write them in terms of only the observer coordinates, and thus obtain the full solution, we use the light cone condition to solve for the intersection of the

past light cone of the observer, and the trajectory of the source. In Minkowski coordinates,

$$(X^i - X_*^i)^2 = (T - T_*)^2, \quad (6.90)$$

and in Rindler coordinates

$$(x - x_*)^2 + (y - y_*)^2 + z^2 + z_*^2 - 2zz_*\cosh[g_H(t - t_*)] = 0 \quad (6.91)$$

6.10.2 A General Dipole Solution

We now derive the analog of the Lienard-Wiechert potential for a pure dipole source with arbitrary 4-velocity. We start again from Eq. (6.81) but write the 4-current for a point dipole source

$$J^\alpha(x) = c\nabla_\mu \int Q^{\alpha\mu}(\tau)\delta^{(4)}[x - x_S(\tau)] d\tau \quad (6.92)$$

where the antisymmetric dipole tensor,

$$Q^{\alpha\mu}(\tau) = V^\alpha p^\mu - p^\alpha V^\mu + \epsilon^{\alpha\mu}_{\rho\sigma} V^\rho m^\sigma, \quad (6.93)$$

is the antisymmetric decomposition of electric and magnetic parts given by [Ribaric & Sustersic \(1995\)](#). We discuss this decomposition further in the next appendix. Eq. (6.93) is a general decomposition of any antisymmetric rank two tensor given vectors V , p and m such that $p \cdot V = m \cdot V = 0$. Such vectors p and m can be chosen in terms of the dipole moments as measured in

the instantaneous rest frame of the source, \mathbf{p}_S and \mathbf{m}_S ,

$$\begin{aligned} p^\alpha &= \left(\gamma_S \beta_j p_S^j, p_S^i + \frac{\gamma_S - 1}{\beta^j \beta_j} \beta_j p_S^j \beta^i \right) \\ m^\alpha &= \left(\gamma_S \beta_j m_S^j, m_S^i + \frac{\gamma_S - 1}{\beta^j \beta_j} \beta_j m_S^j \beta^i \right) \end{aligned} \quad (6.94)$$

In Minkowski coordinates we set $\sqrt{-g} = 1$ and substitute (6.92) into (6.81) to get,

$$A^\alpha = 2 \int \left\{ H(x^0 - \bar{x}^0) \delta[(x - \bar{x})^2] \bar{\nabla}_\mu \left(\int Q^{\alpha\mu}(\tau) \delta^{(4)}[\bar{x} - x_S(\tau)] d\tau \right) \right\} d^4 \bar{x}. \quad (6.95)$$

Since $Q^{\alpha\mu}$ depend on source coordinates while $\bar{\nabla}_\mu$ is taken with respect to observer coordinates, we may take $Q^{\alpha\mu}$ out of the derivative,

$$A^\alpha = 2 \int \int \left\{ H(x^0 - \bar{x}^0) \delta[(x - \bar{x})^2] Q^{\alpha\mu}(\tau) \bar{\nabla}_\mu (\delta^{(4)}[\bar{x} - x_S(\tau)]) \right\} d\tau d^4 \bar{x}. \quad (6.96)$$

To evaluate the above integrals we use the notion of a generalized derivative and employ integration by parts to write,

$$\int_b^a \partial_{\bar{x}} \delta(x - \bar{x}) f(\bar{x}) d\bar{x} = \delta(x - \bar{x}) f(\bar{x}) \Big|_b^a - \int_b^a \delta(x - \bar{x}) \partial_{\bar{x}} f(\bar{x}) d\bar{x} = -[\partial_{\bar{x}} f(\bar{x})]_{\bar{x}=x} \quad (6.97)$$

for a continuous, once differentiable function $f(x)$. We assume $x \subset (a, b)$ so that the boundary terms disappear. Generalizing to multiple dimensions,

$$\int_b^a \int_b^a \partial_{\bar{x}} [\delta(x - \bar{x}) \delta(y - \bar{y})] f(\bar{x}, \bar{y}) d\bar{x} d\bar{y} = -[\partial_{\bar{x}} f(\bar{x}, y)]_{\bar{x}=x}, \quad (6.98)$$

and to the case at hand,

$$\int \partial_\alpha [\delta^n(x - \bar{x})] F^{\alpha\beta\dots}(\bar{x}^0, \bar{x}^1 \dots \bar{x}^n) \sqrt{-\bar{g}} d^n \bar{x} = - [\partial_\alpha F^{\alpha\beta\dots}(\bar{x}^0, \bar{x}^1 \dots \bar{x}^n) \sqrt{-\bar{g}}]_{\bar{x}^\alpha = x}. \quad (6.99)$$

Integrating by parts (*i.e.* using 6.99)

$$\begin{aligned} A^\alpha &= -2 \int [\bar{\nabla}_\mu H(x^0 - \bar{x}^0) \delta[(x - \bar{x})^2] Q^{\alpha\mu}(\tau)]_{\bar{x}=x_S(\tau)} d\tau \\ &= -2 \int \nabla_\mu^S (H(x^0 - x_S^0(\tau)) \delta[(x - x_S(\tau))^2]) Q^{\alpha\mu}(\tau) d\tau \end{aligned} \quad (6.100)$$

where $\nabla_\mu^S = \frac{\partial}{\partial x_S^\mu(\tau)}$ and the second line follows because $Q^{\alpha\mu}(\tau)$ does not depend on \bar{x} , and the function being differentiated and then evaluated at $x_S(\tau)$ does not depend anywhere on $x_S(\tau)$ before being evaluated. Applying the product rule to the derivative, we obtain two terms to evaluate

$$A_1^\alpha = -2 \int \nabla_\mu^S [H(x^0 - x_S^0(\tau))] \delta[(x - x_S^0(\tau))^2] Q^{\alpha\mu}(\tau) d\tau \quad (6.101)$$

$$A_2^\alpha = -2 \int H(x^0 - x_S^0(\tau)) \nabla_\mu^S [\delta[(x - x_S^0(\tau))^2]] Q^{\alpha\mu}(\tau) d\tau \quad (6.102)$$

Using

$$[\bar{\nabla}_\mu H(x^0 - \bar{x}^0)]_{\bar{x}=x_S} = -\delta_\mu^0 \delta(x^0 - x_S^0) \quad (6.103)$$

and

$$\delta[(x - x_S(\tau))^2] = \frac{\delta(\tau - \tau_*)}{| -2 r_\nu(\tau) V^\nu(\tau) |_*} \quad (6.104)$$

we find

$$\begin{aligned} A_1^\alpha(x) &= \int d\tau Q^{\alpha 0} \delta(x^0 - x_S^0) \frac{\delta(\tau - \tau_*)}{[r \cdot V]_*} \\ &= \left. \frac{Q^{\alpha 0}}{(r \cdot V)} \delta(x^0 - x_S^0) \right|_* \end{aligned} \quad (6.105)$$

which is only non-zero at the retarded point, so for observers at the location of the dipole. The second term can be written,

$$A_2^\alpha = -2 \int Q^{\alpha\mu}(\tau) H(x^0 - x_S^0(\tau)) \nabla_\mu^S \tau \frac{d}{d\tau} \left[\frac{\delta(\tau - \tau_*)}{| -2(r \cdot V)|_*} \right] d\tau \quad (6.106)$$

where we have used the chain rule to rewrite ∇_μ^S . Integrating by parts we find,

$$A_2^\alpha = \int \frac{d}{d\tau} [Q^{\alpha\mu}(\tau) H(x^0 - x_S^0(\tau)) \nabla_\mu^S \tau] \frac{\delta(\tau - \tau_*)}{(r \cdot V)_*} d\tau \quad (6.107)$$

which, upon integration over τ , we may write as

$$A_2^\alpha = \frac{d}{d\tau_*} [Q^{\alpha\mu}(\tau_*) H(x^0 - x_S^0(\tau_*)) \nabla_\mu^S \tau_*] \frac{1}{(r \cdot V)_*} \quad (6.108)$$

where we have again exploited the fact that the function being differentiated in square brackets does not depend anywhere on τ_* . This allows us to first evaluate the function at τ_* and then take the derivative wrt τ_* , instead of differentiating first and then evaluating.

Now, we can use

$$\nabla_\mu^S \tau|_* = \nabla_\mu \tau|_* = \left. \frac{r_\mu}{r \cdot V} \right|_* \quad (6.109)$$

which follows from writing out the gradient of the null condition with respect to source coordinates,

$$\begin{aligned} r_\nu \nabla_\mu^S r^\nu \Big|_* &= r_\nu (\nabla_\mu^S x^\nu - \delta_\mu^\nu) \Big|_* = 0 \\ r_\nu \nabla_\mu^S x^\nu \Big|_* &= r_\mu \Big|_*, \end{aligned} \quad (6.110)$$

so that we may write

$$\nabla_\mu^S \tau \Big|_* = \nabla_\mu^S x^\gamma \nabla_\gamma \tau \Big|_* = \frac{r_\mu}{r \cdot V} \Big|_*. \quad (6.111)$$

Using this, we can combine terms and include A_1^α to obtain,

$$A^\alpha(x) = \nabla_\mu \left[\frac{Q^{\alpha\mu}}{(r \cdot V)} \right]_* - \left[\gamma_S \frac{Q^{\alpha\mu} r_\mu}{(r \cdot V)^2} - \frac{Q^{\alpha 0}}{(r \cdot V)} \right]_* \delta(\mathbf{r}) \quad (6.112)$$

The first term is our desired solution for observers off of the source worldline. The terms which turn on at the position of the dipole fall off one factor of $-(r \cdot V)$ more slowly than the worldline terms. Still however, the off-worldline terms and their curl blow up at the position of the dipole. A very different means to the first term in Eq. (6.112) can be found in [Peter Rowe & Rowe \(1987\)](#).

For completeness we write out the transformed Rindler potential (for observers not at the position of the dipole) in terms of positions, velocities, accelerations, and dipole moments in the Rindler frame analogous to Eq. (6.89),

$$A^{\alpha'}(x') = L_\alpha^{\alpha'} \epsilon^{\alpha\mu}_{\rho\sigma} \left[\frac{L_{\rho'}^\rho a^{\rho'} L_{\sigma'}^\sigma m^{\sigma'} L_\mu^{\mu'} r_{\mu'}}{(r \cdot V)^2} + \frac{L_{\rho'}^\rho V^{\rho'} L_{\sigma'}^\sigma \dot{m}^{\sigma'} L_\mu^{\mu'} r_{\mu'}}{(r \cdot V)^2} - \frac{L_{\rho'}^\rho V^{\rho'} L_{\sigma'}^\sigma m^{\sigma'} L_\mu^{\mu'} r_{\mu'}}{(r \cdot V)^3} (1 + r \cdot a) \right]_* . \quad (6.113)$$

6.11 Dipole Moments

To clarify the interpretation of the dipole moment 4-vectors, we will work in direct analogy to the electric and magnetic field vectors for which the antisymmetric Maxwell tensor is

$$F^{\alpha\mu} = u^\alpha E^\mu - E^\alpha u^\mu + \epsilon^{\alpha\mu}_{\rho\sigma} B^\rho u^\sigma \quad (6.114)$$

The fields as measured by an observer with 4-velocity u^μ are then given by

$$\begin{aligned} E^\alpha &= F^{\alpha\mu} u_\mu \\ B^\alpha &= \frac{1}{2} \epsilon^{\alpha\mu\gamma\delta} F_{\gamma\delta} u_\mu \end{aligned} \quad (6.115)$$

The electric field of one observer is *not* a coordinate transformation of the electric field of another observer, although the two electric fields can be related.

To relate the fields of two different observers, consider first a Minkowski observer. Having 4-velocity $u_M^\mu = (1, 0, 0, 0)$. She sees

$$\begin{aligned} \mathbf{E}_M &= -F^{i0} \\ \mathbf{B}_M &= \frac{1}{2} \epsilon^{ijk} F_{jk} \end{aligned} \quad (6.116)$$

so that we can build the Maxwell tensor in Minkowski coordinates:

$$F^{\alpha\mu} = \begin{pmatrix} 0 & -\mathbf{E}_M \\ \mathbf{E}_M & \epsilon_k^{ij} \mathbf{B}_M^k \end{pmatrix}. \quad (6.117)$$

On the other hand, the EM fields measured by an observer that moves with generic 4-velocity $u^\mu = \gamma(1, \beta)$ according to our Minkowski observer can be related to the EM fields measured by

our Minkowski observer through:

$$\begin{aligned} E_o^{\alpha'} &= \frac{\partial x^{\alpha'}}{\partial x^\alpha} F^{\alpha\mu} u_\mu \\ B_o^{\alpha'} &= \frac{1}{2} \frac{\partial x^{\alpha'}}{\partial x^\alpha} \epsilon^{\alpha\mu\gamma\delta} F_{\gamma\delta} u_\mu. \end{aligned} \quad (6.118)$$

Expanding gives the relation

$$\begin{aligned} \mathbf{E}'_o &= \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}} \cdot \mathbf{E}_M)(1 - \gamma) + \gamma \mathbf{E}_M + \gamma(\boldsymbol{\beta} \times \mathbf{B}_M) \\ \mathbf{B}'_o &= \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}} \cdot \mathbf{B}_M)(1 - \gamma) + \gamma \mathbf{B}_M - \gamma(\boldsymbol{\beta} \times \mathbf{E}_M) \end{aligned} \quad (6.119)$$

(equivalent to Eqs. (6.34).) For emphasis, \mathbf{E}_M are the components of the electric field measured by a Minkowski observer in her inertial frame basis while by contrast \mathbf{E}'_o are the components of the electric field measured by an observer boosted (relative to the Minkowski observer) expressed in the (boosted) observer's coordinate basis. The fields \mathbf{E}_M and \mathbf{E}'_o are not related solely by a coordinate transformation.

To construct the relevant objects and interpretations for the EM dipole moments we note that

$$\begin{aligned} Q &\leftrightarrow F \\ p &\leftrightarrow E \\ m &\leftrightarrow B \\ V &\leftrightarrow u \end{aligned} \quad (6.120)$$

Working in analogy with the above, we begin with the antisymmetric dipole tensor $Q^{\alpha\mu}$. Any antisymmetric rank two tensor can be decomposed given a timelike unit vector V and two vectors

p and m which are orthogonal to V

$$Q^{\alpha\mu} = V^\alpha p^\mu - p^\alpha V^\mu + \epsilon_{\rho\sigma}^{\alpha\mu} V^\rho m^\sigma , \quad (6.121)$$

The identities

$$\begin{aligned} p^\alpha &= Q^{\alpha\mu} V_\mu \\ m^\alpha &= \frac{1}{2} \epsilon^{\alpha\mu\gamma\delta} F_{\gamma\delta} V_\mu \end{aligned} \quad (6.122)$$

provide definitions for the covariant dipole moments. We will use $V^\alpha = V_S^\alpha = \gamma_S(1, \beta_S)$, the 4-velocity of the source. An observer at rest in the coordinate basis in which Q is expressed measures dipole moments (6.122) for a source moving with velocity V_S^μ with respect to that basis.

In direct analogy to the EM field tensor, the dipole moment tensor $Q^{\alpha\mu}$ is the physically significant entity while, in direct analogy to the electric and magnetic fields, p and m are observer dependent.

In the rest frame of the source, $V_S^{\mu'} = (1, 0, 0, 0)$, and we can define the rest-frame moments:

$$\begin{aligned} p^{\alpha'} &\equiv (0, \mathbf{p}_S) = -Q^{i'0'} \\ m^{\alpha'} &\equiv (0, \mathbf{m}_S) = \frac{1}{2} \epsilon^{i'j'k'} Q_{j'k'} \end{aligned} \quad (6.123)$$

so that we can build the dipole tensor in Minkowski coordinates:

$$Q_{\mu'}^{\alpha'} = \begin{pmatrix} 0 & -\mathbf{p}_S \\ \mathbf{p}_S & \epsilon^{i'j'}_{k'} \mathbf{m}^{k'} \end{pmatrix} . \quad (6.124)$$

In general, the source may be moving with respect to the natural basis. If we want to express the

components of Q in a basis with respect to which the source is moving, we Lorentz transform

$$Q^{\alpha\mu} = \frac{\partial x^\alpha}{\partial x^{\alpha'}} \frac{\partial x^\mu}{\partial x^{\mu'}} Q^{\alpha'\mu'} . \quad (6.125)$$

The Lorentz boosted source velocity is $V_S^\mu = \gamma_S(1, \beta_S)$ and the Lorentz boosted dipole moments are

$$\begin{aligned} p^\alpha &= \left(\gamma_S \beta_S \cdot \mathbf{p}_S, \mathbf{p}_S + (\gamma_S - 1)(\hat{\beta}_S \cdot \mathbf{p}_S)\hat{\beta}_S \right) \\ m^\alpha &= \left(\gamma_S \beta_S \cdot \mathbf{m}_S, \mathbf{m}_S + (\gamma_S - 1)(\hat{\beta}_S \cdot \mathbf{m}_S)\hat{\beta}_S \right) \end{aligned} \quad (6.126)$$

which is identical to Eqs. (6.31) with $\mathbf{p}_S \neq 0$. Expressions (6.126) are the moments as measured by an observer that sees the source boosted.

There is a subtlety to be noted. V determines the basis in which you are expressing Q , unlike EM fields where u does not determine the basis in which you are expressing F . In other words, you can choose a basis and write the components of F in that basis. However, there are no restrictions on which observer you consult in that basis and therefore no restrictions on which u to contract with in the definitions of E, B . By contrast, once you choose a basis and write out the components of Q in that basis, you have fixed the source velocity V_S . There is one and only one V_S with respect to a given basis, and so one and only one V_S to contract with in the definitions in p and m .

Chapter 7

Bright Transients from Strongly Magnetized Neutron Star - Black Hole Mergers

7.1 introduction

Black holes are dark dead stars. Neutron stars are giant magnets. As the neutron star (NS) whips around the black hole (BH) in the final stages in the life of a pair, an electromotive force (emf) is generated that is powerful enough to light a beacon, which conceivably we might observe at cosmological distances ([McWilliams & Levin 2011](#); [D’Orazio & Levin 2013](#)). The battery could power synchro-curvature radiation, a blazing fireball, or relativistic jets.

Famously, tidal disruption of a NS is expected to generate a gamma-ray burst after merger ([Narayan et al. 1992b](#)). However, it is under-appreciated that most BHs should be large enough ($\gtrsim 6M_{\odot}$) to swallow their NSs whole and so no gamma-ray burst is expected from typical pairs ([Özel et al. 2010](#)). Therefore, our BH battery, which operates with the NS intact, may be one

of the only significant sources of electromagnetic luminosity for coalescing BHNS binaries.^a An observation of such a transient would be exciting in its own right. Advanced gravitational-wave detectors (*e.g.*, Harry & LIGO Scientific Collaboration 2010), with the prospect of multi-messenger astronomy, provide added incentive for the more detailed predictions of the EM signatures we present here.

Even with the benefit of nearly fifty years of observations, common NS pulsars require theoretical attention. If the decades of pulsar research offer a sociological lesson, it would be that the details of the electromagnetic processes are not easy to model, that the mechanisms at work are not obvious. Without the benefit of observations, we would not presume to offer a definitive or complete electromagnetic portrait of the BHNS engine. But we can sketch plausible emission mechanisms to encourage first searches for these potentially important transients.

As already argued in the original references (McWilliams & Levin 2011; D’Orazio & Levin 2013), curvature radiation is a natural channel for luminosity. We examine the spectrum of curvature radiation here. (We mention that another intriguing channel for some fraction of the battery power could be radio emission through coherent processes, providing the correct timescales and energetics for a subclass of the fast radio bursts (Mingarelli et al. 2015).) We conclude that, just before merger, when the power is greatest, curvature radiation results in copious pair production which fuels a fireball. The fireball expands under its own pressure until the photosphere radiates as a blackbody peaking in the hard X-ray to γ -ray range for milliseconds (msec) to seconds depending on NS magnetic-field strength.

If the merger were to happen in our own galaxy, we might watch the spectrum of curvature radiation ramp up followed by the brighter fireball. At cosmological distances, the high-energy lead up in curvature radiation will be too faint to detect, but the fireball could be observable at a rate of at least a few per year with the FERMI Gamma-Ray Burst Monitor (GBM), for NS’s with

^aResonant shattering of the NS crust could also generate an interesting electromagnetic signature for non-disrupting systems (Tsang et al. 2012; Tsang 2013).

$\gtrsim 10^{14}$ G surface magnetic fields. Such events could possibly be a subclass of short gamma-ray bursts. Since the fireball takes at least ~ 0.2 ms to 0.02s to expand and release the light, the burst from the fireball would lag just behind the peak gravitational-wave emission. Post-merger, the transfer of magnetic flux on to the black hole might lead to a brief jet and afterglow. Pre- and post-merger triggered events could be observed to occur very close to each other in timing. We hope the predicted transient discussed here encourages observational interest.

7.1.1 The power of the battery

Firstly, we review the estimate of the energy budget for the BH battery. The NSBH system behaves analogously to a unipolar inductor, which has been investigated in application to a number of other astrophysical systems, *e.g.* Jupiter and its moon Io (Goldreich & Lynden-Bell 1969), planets around white dwarfs (Li et al. 1998) and main sequence stars (Laine et al. 2012; Laine & Lin 2012), binary neutron stars (Vietri 1996; Piro 2012; Lai 2012; Palenzuela et al. 2013), compact white dwarf binaries (Wu et al. 2002; Dall’Osso et al. 2006, 2007; Lai 2012), BHs boosted through magnetic fields (Lyutikov 2011; Penna 2015b), and the Blandford-Znajek (BZ) mechanism (Blandford & Znajek 1977) for a single BH spinning in a magnetic field (for recent numerical work on the BZ mechanism see *e.g.* Palenzuela et al. 2011; Kiuchi et al. 2015). The calculation for NSBH systems, already presented in Ref. McWilliams & Levin (2011) and confirmed in the detailed relativistic analysis of Ref. D’Orazio & Levin (2013), as well as the numerical calculations of Ref. Paschalidis et al. (2013), gives the scaling of power available for conversion into electromagnetic luminosity. In the next section we will consider the implications of throwing this power into luminous elements in the BHNS circuit.

For observers which have not fallen through, the BH horizon is well approximated, electromagnetically, as a conducting sphere (Thorne et al. 1986). The relative motion of the BH through the magnetic field of the NS induces an emf. We visualize the circuit which generates this emf

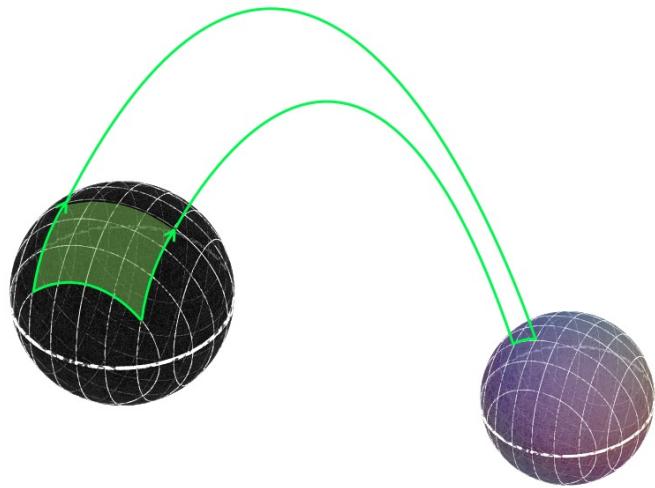


Figure 7.1: Schematic of a Faraday loop as seen by an observer external to the horizon. The black sphere depicts the BH horizon orbiting out of the page. In green is a schematic of the instantaneous closed loop defining one of infinitely many circuits made up of electrons and positrons moving along magnetic fields lines which trace the BH horizon.

in Fig. 7.1. Because charged particles are bound to a given field line, we imagine that one set of field lines forms one set of wires in a closed circuit. In conceptualizing the circuit it is important to distinguish between field lines that act as wires at a given instant and those that contribute to the changing magnetic flux through the circuit. The circuit is closed by connecting the wires along the surface of the horizon, as in the snapshot of Fig. 7.1. As the BHNS pair orbits, the circuit sweeps through the dipole field. The changing magnetic flux through a surface bounded by the changing circuit corresponds to an emf. There are an infinite number of such circuits as different field lines intersect the BH.

Following Ref. [McWilliams & Levin \(2011\)](#), the voltage generated is given by

$$\begin{aligned}
 V_{\mathcal{H}} &= \int \alpha \mathbf{E} \cdot d\mathbf{s} = -\frac{1}{c} \frac{d}{dt} \int \alpha \mathbf{B} \cdot d\mathbf{A} \\
 &= - \oint \alpha \left(\frac{\mathbf{v}}{c} \times \mathbf{B} \right) \cdot d\mathbf{s},
 \end{aligned} \tag{7.1}$$

where \mathbf{v} is the relative velocity of the BH horizon with respect to magnetic field lines and we add a factor of the lapse function for a spinning BH α by hand to account for the gravitational redshifts.^b Given a dipole magnetic field, which drops off with distance from the NS as r^{-3} , anchored on the NS with radius R_{NS} (taken to be 10 km throughout) and surface magnetic-field strength B_{NS} ,

$$B(r) = B_{\text{NS}} \left(\frac{R_{\text{NS}}}{r} \right)^3, \quad (7.3)$$

the voltage (7.1) acquires a contribution only from the integral along the horizon in the direction of the line connecting the BH and NS, and so evaluates to

$$V_{\mathcal{H}} = 2R_H \left[\frac{r(\Omega_{\text{orb}} - \Omega_{\text{NS}})}{c} + \frac{S}{4\sqrt{2}} \right] B_{\text{NS}} \left(\frac{R_{\text{NS}}}{r} \right)^3, \quad (7.4)$$

where R_H is the radius of the horizon and where we have included a factor to account for the spin, $0 \leq S \leq 1$, of the BH McWilliams & Levin (2011). Notice that in Eq. (7.3), B_{NS} drops off with distance from the NS, so the voltage varies across the horizon for small binary separations. In the limit in which we ignore the finite size of the compact objects, we interpret r as the binary separation.

The total power that can be liberated by the battery is,

$$\mathcal{P}(t) = \frac{V_{\mathcal{H}}^2(t)}{(\mathcal{R}_{\mathcal{H}} + \mathcal{R}_{\text{NS}})^2} \mathcal{R}_{\text{NS}}, \quad (7.5)$$

^bIn Boyer-Lindquist coordinates for a Kerr BH,

$$\begin{aligned} \alpha &= \frac{\rho}{\Sigma} \sqrt{\Delta} \\ \rho &= (\varepsilon^2 + S^2 \cos^2 \theta)^{1/2} \\ \Sigma &= ([\varepsilon^2 + S^2] - S^2 \Delta \sin^2 \theta)^2. \end{aligned} \quad (7.2)$$

for BH spin $S \leq 1$. Here we use ε for the distance from the BH to be distinguished from the distance from the neutron star r .

The resistance across the horizon of the BH is $\mathcal{R}_{\mathcal{H}} = 4\pi/c \text{ cm}^{-1}\text{s}$. Since the effective resistance of the NS and its magnetosphere (\mathcal{R}_{NS}) is unknown, we choose $\mathcal{R}_{\text{NS}} = \mathcal{R}_{\mathcal{H}}$ to give the largest possible luminosities. This impedance matching condition is the same as that imposed to derive the Blandford-Znajek power (Blandford & Znajek 1977), in which case the angular velocity of magnetic-field lines at infinity are set to one half of the BH horizon angular velocity (Thorne et al. 1986; Penna 2015a).

The power scales roughly as

$$\mathcal{P} \sim M^2 B_{\text{NS}}^2 r^{-6} v^2 . \quad (7.6)$$

At large separations $v^2 \sim M/r$ is small, climbing to near the speed of light at merger. Measuring length in units of M , the power scales as

$$\mathcal{P} \sim B_{\text{NS}}^2 M^{-4} v^2 . \quad (7.7)$$

For a fixed number of gravitational radii between the NS surface and the BH horizon, a larger BH boosts the power as M^2 , but the larger implied distance between the two decreases the magnetic-field strength at the horizon by M^{-6} .

We discuss briefly when these scalings break down. In the limit that the NS and BH are close, and their finite sizes are important, the NS surface can come arbitrarily close to the BH horizon in which case $B_{\text{NS}}^2 r^{-6} \rightarrow B_{\text{NS}}^2$. Placing the NS surface at the horizon and spinning it with velocity v would generate power which increases with BH mass as $\mathcal{P} \sim M^2 B_{\text{NS}}^2 v^2$. If however, the BH mass was very large, the variation of the magnetic field across the BH horizon would become important. For very large BHs, the NS light cylinder will not span the horizon.^c In these cases, our assumption that the voltage drop is across the entire horizon breaks down and the power will scale more weakly than M^2 . In the present work, we ignore finite-size effects and take Eqs. (7.3)-(7.5) to be a good

^cWhen the BH event horizon is larger than the size of the NS light cylinder, $M \gtrsim c^3 G^{-1} \Omega_{\text{NS}}^{-1} \sim 10^4 M_{\odot} 2\pi/\Omega_{\text{NS}}$, the full voltage drop of Eq. (7.4) cannot be realized.

estimate of the average power available via the BH battery.

Here and throughout the rest of the paper we treat the NS surface magnetic-field strength as an unknown parameter. Because there are no observations of NSBH binaries, and hence no measurements of NS field strengths near merger with a BH, we have chosen a range in accordance with the observed NS fields (see *e.g.* [Kaspi & Kramer 2016](#)). We consider fields ranging from those of the radio pulsar population 10^{12} G up to the observed magnetar field strengths of a few times 10^{15} G ([Olausen & Kaspi 2014](#)) and beyond to larger, but not impossible field strengths of 10^{16} G,^d in order to probe the full range of energies available to the NSBH system. Conversely and as we discuss in Section §[7.5](#), our models can constrain the NS field strength at merger.

In Figure [7.2](#), we plot the total power available for liberation by the binary as a function of time for varying NS magnetic field strengths and a maximally spinning BH of mass $10M_{\odot}$.^e Importantly, over the range of possible magnetic-field strengths, the energy liberated through the BH battery mechanism is many orders of magnitude lower than that liberated by gravitational radiation ([McWilliams & Levin 2011](#)), hence the orbital inspiral timescales are set by gravitational radiation loss and are robust despite different possible channels for the electromagnetic power. The time dependent separation $r(t)$ decays due to gravitational radiation losses [Peters \(1964\)](#),

$$r(t) = \left(r^4(0) - 4 \frac{64}{5} \frac{G^3}{c^5} M_{\text{NS}} M (M + M_{\text{NS}}) t \right)^{1/4}, \quad (7.8)$$

where M_{NS} is the NS mass taken to be $1.4M_{\odot}$ throughout. Over the final second, the power available climbs by ~ 8 orders of magnitude. For a 10^{12} G dipole field, the power rises from pulsar

^dNS field strengths above $\sim 10^{16}$ G would generate EM power through the BH-battery mechanisms which would rival the emission due to gravitational radiation and hence invalidate our use of Eq. (7.8) for evolution of the binary orbital separation.

^eDepending on the NS equation of state, the choice of a maximally spinning BH could cause the NS to be partially disrupted (*e.g.* [Foucart 2012](#)). However the exact line between partial and non-disruption is not known and a small change in BH spin to $S = 0.95$, which brings the fiducial system safely within the non-disrupting regime, only weakly affecting our results. Hence we choose $S = 1$ for simplicity.

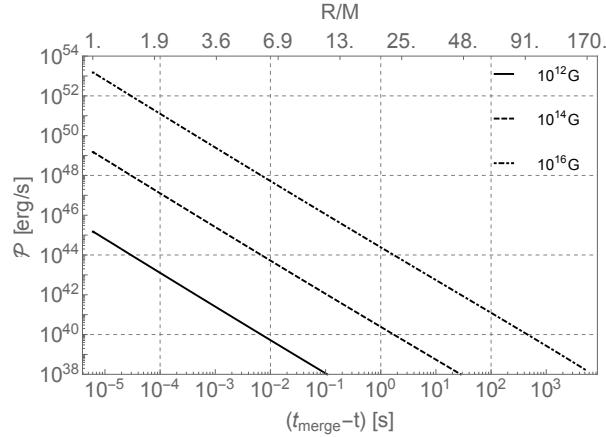


Figure 7.2: Total possible power supplied by the BH battery via Eq. (7.5) as a function of time until merger for two point masses undergoing orbital decay via gravitational radiation reaction (Eq. 7.8). The solid, dashed, and dot-dashed lines indicate NS surface magnetic-field strengths of 10^{12}G , 10^{14}G , and 10^{16}G respectively, for a BH mass of $10M_{\odot}$. The plot extends to a binary separation of GM/c^2 , the size scale of the event horizon for the maximally spinning BH we consider. We have dropped factors of G and c in the axis labels.

scales $\sim 10^{36}\text{erg s}^{-1}$ in that second, to $\sim 10^{44}\text{erg s}^{-1}$ in the final millisecond (at $r = 2GM/c^2$). The power scales as B^2 reaching 10^{52}erg s^{-1} for a magnetar with $B \sim 10^{16}\text{G}$. For a maximally spinning BH, the horizon is at $r = GM/c^2$, so we extend the luminosity scaling in Figure 7.2 down to this separation (noting that we still have $GM/c^2 > R_{\text{NS}}$ for $M \geq 7M_{\odot}$) where the luminosity peaks at $\sim 10^{45}\text{erg s}^{-1}(B/10^{12}\text{G})^2$.

Eqn. (7.5) gives an estimate of the power the battery could generate. Whether or not this power is available to light up the pair is the question at hand. We describe the most straightforward vehicles to convert the power into luminosity in the following sections.

7.2 Curvature radiation

The voltage drop will accelerate charges across magnetic-field lines connecting the NS to the BH. Basic physics suggests that these accelerated charges will provide a sensible channel for luminos-

ity. The charges spiral around and are pushed along the magnetic fields when there is a parallel component of electric field, $\mathbf{E} \cdot \mathbf{B} \neq 0$. The result is a primary spectrum of curvature radiation.^f

The extent to which the BH battery can act as a particle accelerator is mitigated by the conducting properties of the surrounding magnetosphere. The NS sustains a magnetosphere by pulling charges from the NS and through various pair production channels in the magnetosphere ([Goldreich & Julian 1969](#); [Ruderman & Sutherland 1975](#)). The plasma acts as a conductor and will screen the NS's electric fields until force-free conditions are established, that is, until $\mathbf{E} \cdot \mathbf{B} = 0$.

Once the BH enters the light cylinder of the NS and the battery is established, the electric field configuration changes and the magnetosphere adjusts with those changes. At the large separations of the light cylinder, the plasma is tenuous but in the final stages when the voltage is most powerful, both compact objects should be submerged in the conducting plasma. Consequently, we anticipate that some of the emf generated by the orbital motion is screened and forces are muted. However, as with the pulsar, there must be gaps in which screening is inefficient and across which particles must be accelerated. Additionally, current sheets could act to dissipate the BH battery power.

We currently do not know the degree to which the voltage is reduced by screening. In the future, global particle-in-cell codes could asses the gap structure in a BHNS magnetosphere. To make simple estimates, we continue to use the full power of the battery in the calculation of the curvature radiation, aware that screening could significantly reduce the estimates.

To obtain the primary curvature radiation spectrum, we assume a distribution in energy of the magnetosphere electrons and positrons. The spectrum of curvature radiation is given by integrating the one-electron spectrum multiplied by the number distribution of charged particles.

$$P_C(\nu, t) = \int_{\gamma_{\min}}^{\gamma_{\max}} N(\gamma) \frac{dP_C}{d\nu} d\gamma \quad (7.9)$$

^fWhen the energy of curvature photons is great enough, they will interact with the magnetosphere magnetic and electric fields and produce electron-positron pairs. As the curvature photons are not locked to move along magnetic-field lines, the secondary pairs can have a non-negligible component of motion transverse to the magnetic field, resulting in a secondary synchrotron spectrum.

where $dP_C/d\nu$ represents the curvature radiation power per unit frequency (e.g, Cheng & Zhang 1996). We model the population as a power law in the relativistic Lorentz factor γ ,

$$N(\gamma)d\gamma = N_0\gamma^{-p}d\gamma. \quad (7.10)$$

The normalization constant N_0 is chosen so that the total bolometric luminosity matches Eq. (7.5)

$$N_0 = \frac{\mathcal{P}}{\int \int \gamma^{-p} \frac{dP_C}{d\nu} d\gamma d\nu}, \quad (7.11)$$

so that the magnetosphere number density ($\sim N_0/r^3$) is set by the physics of curvature radiation and the requirement that the magnetosphere maximally radiates the BH-battery power.

The spectrum then depends on the energy distribution of electrons and positrons through the exponent p , and the time-dependent minimum and maximum Lorentz factors of particles in the magnetosphere $\gamma_{\max}(t)$ and $\gamma_{\min}(t)$ that we must input from the physical model of the BHNS battery. As the spectrum is not greatly dependent on the minimum γ or the power law index p (see the Appendix), we leave these as free parameters. The shape of the spectrum will depend on the choice of $N(\gamma)$, but, for what follows, the most important consideration will be where the high energy end of the spectrum is cut off. This is set by the maximum electron Lorentz factor in the magnetosphere.

We approximate the maximum γ as the largest radiation reaction limited Lorentz factor in the magnetosphere. Electrons and positrons are accelerated along magnetic-field lines to radiation reaction limited velocities given by solving,

$$ec|\mathbf{E}_{||}|(1 - \gamma^{-2})_{\max}^{1/2} = \frac{2}{3} \frac{ce^2\gamma_{\max}^4}{\rho_c^2} \quad (7.12)$$

for the Lorentz factor γ_{\max} . Here ρ_c is the radius of curvature of magnetic-field lines. We evaluate

ρ_c for a dipole magnetic field in the binary equatorial plane, $\rho_c = R_{\text{NS}}/3\sqrt{r/R_{\text{NS}}}$. We use the horizon electric field sourced by the potential drop Eq. (7.4) to estimate a maximum value of the accelerating electric fields, $|\mathbf{E}_{||}| \approx |\mathbf{E}| \sim \frac{V_H}{R_H}$ where R_H is the radius of the BH horizon.

Then the radiation reaction limited Lorentz factor of electrons/positrons, at the BH horizon is

$$\gamma_{\max} \approx 4.2 \times 10^7 \left(\frac{r}{6GM/c^2} \right)^{-5/8} \left(\frac{B_{\text{NS}}}{10^{12}G} \right)^{1/4}, \quad (7.13)$$

choosing fiducial parameters $R_{\text{NS}} = 10^6$ cm and $M_{\text{BH}} = 10M_{\odot}$. Electrons and positrons will emit curvature radiation with characteristic energy

$$\epsilon_{\gamma} = \frac{3hc}{4\pi\rho_c} \gamma^3 \approx 1.8 \text{ TeV} \left(\frac{\gamma}{4.2 \times 10^7} \right)^3. \quad (7.14)$$

We plot a representative curvature radiation spectrum for a fiducial $10M_{\odot}$ BH with maximal spin. The dependence of the curvature spectrum on γ_{\min} and p is explored in the Appendix.

In agreement with previous works ([McWilliams & Levin 2011](#); [D’Orazio & Levin 2013](#)), Figure 7.3 shows that the BHNS curvature radiation can be very high energy, >TeV, near merger. In the following section, we point out that this curvature radiation will be prone to copious pair production through interaction with the strong electromagnetic fields of the magnetosphere as well as photon-photon collisions. The pair production will further populate the electron-positron plasma surrounding the binary. Depending on the efficiency at which pairs are produced from the available energy of the BH battery, the magnetosphere will become optically thick to curvature photons. This trapped radiation can power a fireball, which we now characterize.

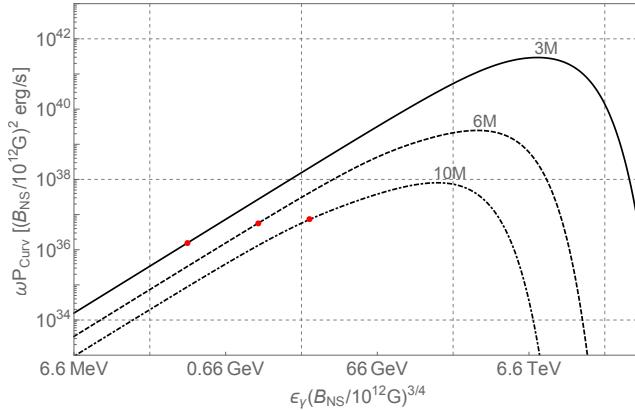


Figure 7.3: The spectra of primary curvature radiation at times corresponding to binary separations $10GM/c^2$, $6GM/c^2$, and $3GM/c^2$ (dot-dashed, dashed, solid) scaled to $B_{\text{NS}} = 10^{12}G$ (factors of G and c are omitted in the labels). We use an electron energy power law index of $p = 2.0$ and a minimum Lorentz factor set by radiation reaction in the outer magnetosphere. Dependence on both parameters is minimal (see the Appendix). The red dots indicate photon energies above which the magnetosphere is opaque to pair production via $\gamma + B$ interactions.

7.3 Fireball

As the BH and NS draw closer, the energy available to accelerate particles increases as $r^{-3}v$, resulting in a higher density of higher energy curvature photons. A consequence is pair production through the interaction of the magnetic field and high-energy photons ($\gamma + B \rightarrow e^+ + e^-$) and through photon collisions ($\gamma + \gamma \rightarrow e^+ + e^-$), preventing the highest energy curvature photons from escaping the magnetosphere. The result is an optically thick pair+radiation fluid, which will expand outwards under its own pressure until pair production becomes disfavored and radiation can escape; the result is a fireball.

7.3.1 pair production

The optical depth to $\gamma + B \rightarrow e^+ + e^-$, at binary separation r is

$$\begin{aligned}\tau_{\gamma B} &= r \left[\frac{4.4}{e^2/(\hbar c)} \frac{\hbar}{m_e c} \frac{B_q}{B_\perp} \exp\left(\frac{4}{3\xi}\right) \right]^{-1} \\ \xi &\equiv \frac{\hbar\omega}{2m_e c^2} \frac{B_\perp}{B_q} \\ B_q &\equiv \frac{m_e^2 c^3}{e\hbar} \approx 4.4 \times 10^{13} \text{G} \\ B_\perp &\equiv \text{Min} \left\{ x/(R_{\text{NS}}/3\sqrt{r/R_{\text{NS}}}), 1 \right\} B(r)\end{aligned}\quad (7.15)$$

for photons with $\hbar\omega \gtrsim 2m_e c^2$. The quantity in brackets is the mean free path for pair production given by Refs. [Erber \(1966\)](#); [Ruderman & Sutherland \(1975\)](#), B_q is a natural quantum mechanical measure of magnetic-field strength, and B_\perp is the component of magnetic field perpendicular to the photon trajectory. The quantity in curly brackets in the last line of Eq. (7.15) is the sine of the angle between a photon trajectory and the magnetic field direction, which is simply the distance x a photon has traveled in direction initially tangent to a field line, divided by the radius of curvature of field lines. As a characteristic value, we take the radius of curvature to be that of a dipole field line which goes through the center of the BH at binary separation r . This approximation assumes that $\xi \ll 1$, which is always true initially when $x = 0$ and $B_\perp = 0$. In practice we cap $\xi \leq 1$ because we are only interested in when $\tau_{\gamma B} \rightarrow 1$. After this point the $\gamma + \gamma \rightarrow e^+ + e^-$ process will also become important, so we need not rely solely on the above calculation (see below).

For very high energy photons, the optical depth limits to very large values but drops exponentially for lower energy photons, generated earlier in the binary inspiral. To capture the steep dependence of the $\gamma + B \rightarrow e^+ + e^-$ optical depth on photon frequency, we evaluate $\tau_{\gamma B}$ at a frequency near the peak of the time dependent curvature radiation spectrum (see Figure 7.3).

The red dots plotted on top of the spectra of Figure 7.3 show the frequency at which the $\gamma +$

$B \rightarrow e^+ + e^-$ optical depth (Figure 7.4) becomes unity for three different snapshots during the inspiral. Above the frequency indicated by the red dots in Figure 7.3, photons pair produce with the magnetic field before escaping the magnetosphere.

The optical depth for $\gamma + \gamma \rightarrow e^+ + e^-$ at binary separation r is,

$$\tau_{\gamma\gamma} \approx r n_{\gamma*} \sigma_{\gamma\gamma} \quad (7.16)$$

where we use a collision cross section $\sigma_{\gamma\gamma} = 11/180\sigma_T$ (Lithwick & Sari 2001; Svensson 1987) averaged over photon energy and written in terms of the Thomson scattering cross section σ_T .

Once the magnetosphere becomes optically thick to $\gamma + B$ pair production, we assume that the radiation plus pair-plasma thermalizes. Then we may approximate $n_{\gamma*}$ as the portion of the Planck spectrum with sufficient energy to produce pairs

$$n_{\gamma*} = \frac{8\pi}{c^3} \int_{2m_e c^2/h}^{\infty} \frac{\nu^2 d\nu}{e^{h\nu/kT} - 1}. \quad (7.17)$$

which is an underestimate as any two photons with energies $\sqrt{\epsilon_1 \epsilon_2} \geq 2m_e c^2$ are favored to create pairs upon collision, not just those above $2m_e c^2$.

Figure 7.4 shows the optical depth of the magnetosphere to both $\gamma + B$ and $\gamma + \gamma$ pair production as a function of time during inspiral for NS magnetic-field strengths which bracket the expected range. The $\gamma + B$ process becomes important first, when curvature-photon energies surpass a critical value (see the red dots plotted on the spectra of Figure 7.3). Much closer to merger, $\gamma + \gamma \rightarrow e^- + e^+$ also becomes an important source of pair production and hence photon opacity.

The high optical depths in Figure 7.4 suggest copious pair production due to $\gamma + B$ earlier in the inspiral. If this process thermalizes the radiation and pairs, then our assumption of a Planck gas in the computation of the subsequent $\gamma + \gamma$ optical depth is warranted. The important point is that, with the large magnetic-field strengths and energy densities present in the NSHB magnetosphere

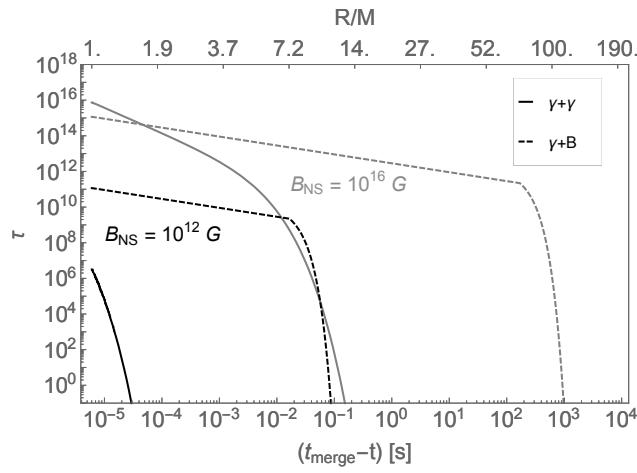


Figure 7.4: The optical depth to the different pair producing processes. The magnetosphere curvature photons are trapped by $\gamma + B$ early on, $\gamma + \gamma$ also becomes relevant for magnetosphere photons just before merger. The $\gamma + B$ optical depth is computed at a time dependent frequency near the peak of the primary curvature spectrum. Factors of G and c are omitted in the upper x-axis label.

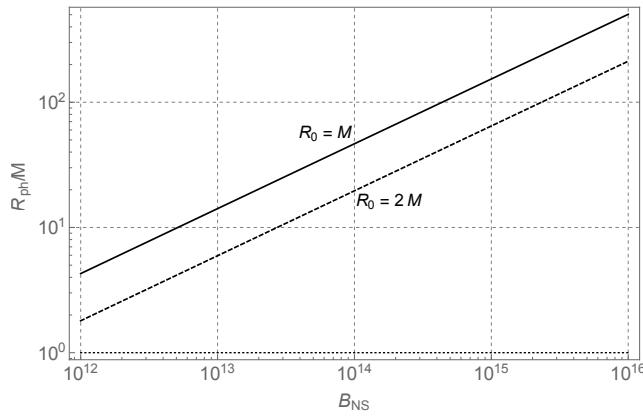


Figure 7.5: The radius of the photosphere as a function of NS magnetic-field strength, for different assumed radii of energy injection $R_0 = GM/c^2, 2GM/c^2$. Factors of G and c are omitted in the figure labels.

near merger, both pair production processes will be favored. Hence we reason that pair production traps and thermalizes the power generated by the BH battery.

We can conclude from this section that the era of curvature radiation gives way to a hot fireball in the final moments before merger. Curvature radiation becomes trapped when $\tau_{\gamma B} = 1$ (Figure 7.4), from which we find that high energy curvature radiation will no longer escape for the final 0.1s ($B/10^{12}\text{G}$) of inspiral. Figure 7.2 shows that at 0.1s ($B/10^{12}\text{G}$) before merger the BH-battery luminosity, and thus the maximum power in curvature radiation, is $\sim 10^{38}\text{erg s}^{-1} (B/10^{12}\text{G})^{1/2}$, a factor of $\sim 10^7 (B/10^{12}\text{G})^{3/2}$ lower than the BH battery peak power at merger. Consequently, at $P_C \lesssim 10^{38}\text{erg s}^{-1} (B/10^{12}\text{G})^{1/2}$, the ramp up in high-energy curvature radiation will likely only be observable within the galaxy.

The subsequent fireball however, could be observable at cosmological distances. We characterize the emission from the fireball in the following section.

7.3.2 Expansion and Emission

The optically-thick pair plus radiation fluid – the fireball – will expand under its own pressure. The alternative is that the fireball falls right down into the BH, although we argue this won't happen. To determine if the fireball will expand, we consider the imbalance of gravity and the mechanical pressure P of the fluid. The condition for expansion is

$$\frac{dP}{dr} \sim \frac{P}{R_0} > \rho \frac{GM}{R_0^2}. \quad (7.18)$$

where R_0 is the initial scale over which energy is injected by the battery. For a radiation dominated fluid $P = \rho c^2/3$ and then

$$R_0 \gtrsim \frac{GM}{c^2}, \quad (7.19)$$

dropping all numerical factors. Radiation pressure alone can cause the fireball to expand. We note that the force balance is marginal at small size scales and will depend on the density distribution in addition to magnetic pressure, both of which will likely increase the outward pressure of the fireball and should be treated in a more detailed calculation. Considering the high temperature at merger, the pressure may be dominated by pairs, not radiation. In this limit, $kT > m_e c^2$, the total pair pressure is $7/4$ the radiation pressure and the fireball will still expand.

After merger, the magnetic fields responsible for $\gamma + B$ pair production will decay without the NS to anchor them (see however §7.4). This means that, after merger, only $\gamma + \gamma$ pair production and electron scattering will trap photons in the expanding fireball. To track the expansion of the fluid from this point, we estimate its properties during and after merger.

Because the optically-thick, pair plus radiation fluid is assumed to be in thermal equilibrium, we can estimate the temperature of the fluid as,

$$T(r) = \left(\frac{\mathcal{P}(r)}{4\pi r^2 \sigma} \right)^{1/4}, \quad (7.20)$$

as a function of the binary separation throughout inspiral, where $\mathcal{P}(r)$ is the power emitted by the BH battery at separation r , and σ is the Stefan-Boltzmann constant. Then the initial temperature of the fireball T_0 is the final temperature before the magnetic fields are swallowed/dissipated and the pair plus radiation fluid is released to expand. Evaluating this temperature at a final binary separation of $R_0 \sim GM/c^2$ gives an initial injection temperature of

$$kT_0 = 85 \text{ keV} \left(\frac{B_{\text{NS}}}{10^{12} G} \right)^{1/2}. \quad (7.21)$$

We treat the fireball as an adiabatically expanding, relativistic fluid. As the fluid expands to a

radial size scale R , it cools as $T = T_0(R/R_0)^{-1}$. At a large enough R , $\gamma + \gamma$ pair production and electron scattering will no longer trap photons, and radiation escapes.

The $\gamma + \gamma$ optical depth is given by Eq. (7.16) and the optical depth to electron/positron scattering is,

$$\tau_{\text{es}} \sim rn_{\pm}\sigma_T, \quad (7.22)$$

where σ_T is the Thomson scattering cross section, and n_{\pm} is the rest-frame, pair number density in thermal equilibrium. We estimate n_{\pm} as the electron number density (*e.g.*, Paczynski 1986), true for $kT \ll m_e c^2$, which is always the case in the photosphere for $B_{\text{NS}} \lesssim 10^{16}$ G. Then,

$$n_{\pm} \approx \frac{4\pi^{3/2}}{h^3} (2m_e kT)^{3/2} \exp\left(-\frac{m_e c^2}{kT}\right) \quad (7.23)$$

Eventually the fireball expands until the temperature has dropped sufficiently for both $\tau_{\gamma\gamma} \leq 1$ and $\tau_{\text{es}} \leq 1$. We call this radius the photosphere radius R_{ph} . We find that the fireball first becomes transparent to $\gamma + \gamma$ pair production and then to electron scattering at a larger, but similar radius (within a factor of a few). Hence the photosphere is defined where $\tau_{\text{es}}(R_{\text{ph}}) \equiv 1$. The photosphere radius as a function of NS magnetic-field strength is plotted in Figure 7.5 for two choices of the initial size of the fireball, GM/c^2 and $2GM/c^2$ (we assume a fiducial $R_0 = GM/c^2$ throughout).

We estimate the Lorentz factor of the adiabatically expanding fluid as $\gamma = R/R_0$ (Paczynski 1986) for $R \gg R_0$. Then emission from the photosphere will be that of a blackbody boosted at Lorentz factor $\gamma_{\text{ph}} = R_{\text{ph}}/R_0$. Such a boosted blackbody looks like the rest frame blackbody but with an effective temperature

$$T_{\text{eff}} = \frac{T_{\text{ph}}}{\gamma_{\text{ph}}(1 - v_{||}/c)} \equiv DT_{\text{ph}} \quad (7.24)$$

where D is the doppler factor, T_{ph} is the temperature in the rest frame of the photosphere, and

$v_{||} = v \cos \theta$ is the line-of-sight velocity, where θ is the angle from observer line of sight. Because the shell is expanding spherically, each patch of the expanding photosphere will have a different effective temperature and the observed, time-dependent spectrum will be a sum of the spectra of all patches on equivalent light travel time surfaces (e.g., Pe'er & Ryde 2011). We do not include such details here; in §7.5 we integrate the line-of-sight dependent blackbody spectra over the photosphere to find a composite spectrum, but for now we make a simple estimate for the peak energy of blackbody emission.

The total photospheric emission will not deviate greatly from blackbody, and the majority of emission will come from the portion of the expanding sphere for which the Doppler factor is positive, where the angle to the line of sight is less than $1/\gamma$. For highly relativistic expansion, the blue-shifted temperature Eq. (7.24) becomes $T = \gamma T_{\text{ph}}$ at $\theta = 1/\gamma$ and $T = 2\gamma T_{\text{ph}}$ at $\theta = 0$. For simplicity we use that the photosphere emission is a blackbody with temperature $T \sim \gamma T_{\text{ph}}$. Then because the photosphere temperature is related to the initial temperature as $T_{\text{ph}} = T_0(R_0/R) = T_0/\gamma_{\text{ph}}$, the observed blackbody temperature is simply $T = T_0$ (*see also Paczynski 1986*); the observed temperature is the same as the initial injection temperature of Eq. (7.21) (the effects of gravitational redshift are negligible for $R_{\text{ph}} \gg R_0$). For a fiducial energy-injection size scale of $R_0 = GM/c^2$, the *observed* photosphere emission will peak at

$$h\nu_{\text{peak}} = 0.24 \text{ MeV} \left(\frac{B_{\text{NS}}}{10^{12}G} \right)^{1/2}, \quad (7.25)$$

ranging from hard X-Rays to γ -rays.

From the pair density at the photosphere we estimate the plasma frequency to be,

$$\nu_{\text{pl}} = \sqrt{\frac{n_{\pm} e^2}{\pi m_e}} \lesssim 4.4 \times 10^{12} \text{ Hz} \left(\frac{B_{\text{NS}}}{10^{12}G} \right)^{-0.26}. \quad (7.26)$$

The blackbody emission is not shorted out by the pair plasma, however, emission in the far-infrared

and at longer wavelengths does not escape the photosphere.

Because the photosphere is generated due to a decrease in pair density, there will be no detectable signal from blue shifted pair annihilation (*see also Paczynski 1986; Goodman 1986*). The ratio of energy in pairs to that in radiation at the photosphere is small,

$$\frac{E_{\pm}}{E_{\gamma}} \simeq \frac{m_e c^2 n_{\pm} c}{\sigma T_{\text{ph}}^4} < 10^{-8}. \quad (7.27)$$

Finally we note that, because the fireball must expand out to its photosphere size before it can radiate, the EM transient predicted here will occur at least $R_{\text{ph}}/c \sim 0.2 \text{ msec} \sqrt{B/10^{12}\text{G}}$ after the initial energy injection. If energy injection is associated with merger, then this EM signature will occur shortly after peak gravitational wave emission. Hence gravitational waves from the inspiral stage, which will trigger a LIGO detection, will also warn of this EM counterpart.

To summarize, we predict that, as the binary nears the final few GM/c^2 in binary separation, high energy curvature radiation will produce pairs by interacting with other photons and also the magnetic field. The BHNS magnetosphere becomes optically thick to pair production, trapping the energy injected by the BH battery. This energy injection causes the optically thick pair plus radiation fluid to expand outwards until the temperature drops below that which favors a high pair density. At this point pair production and electron scattering no longer contain the photons and they escape. For initial NS field strengths of $10^{12} \rightarrow 10^{16}\text{G}$, the observable radiation is characterized as:

- Black-body radiation with a peak photon energy $h\nu \sim 0.24 \text{ MeV} \sqrt{B_{\text{NS}}/10^{12}\text{G}}$.
- A bolometric luminosity of up to $10^{45} \text{ erg s}^{-1} (B_{\text{NS}}/10^{12}\text{G})^2$.
- Defining $\Delta t_{42}(B_{\text{NS}})$ as the time before merger over which the BH is supplying power above $10^{42} \text{ erg s}^{-1}$, and associating this with the emission timescale, the the burst times to the

closest order of magnitude are $\Delta t_{42}(10^{12}\text{G}) \sim 10^{-3}\text{ s}$, $\Delta t_{42}(10^{14}\text{G}) \sim 0.1\text{ s}$, $\Delta t_{42}(10^{16}\text{G}) \sim 10\text{ s}$.

We next consider a post-merger signal and the observability of both merger and post-merger events.

7.4 Post Merger

When the BH swallows the NS, a magnetic flux is deposited onto the BH, magnetizing the hole. The no-hair theorem suggests the BH, in vacuum, must shed the absorbed B -field on order the BH light crossing time, in very long-wavelength, $\sim R_H$, radiation (e.g., [Baumgarte & Shapiro 2003](#)). However, [Lyutikov & McKinney \(2011\)](#) have argued, in the context of NS collapse to a BH, that because the BH is immersed in magnetosphere plasma, the no-hair theorem is not applicable and the BH may retain a magnetic field anchored in a remnant magnetosphere for longer. The situation is similar to our case where the BH swallows the NS. In the limit of a non-resistive plasma, magnetic field lines are frozen into the plasma of the magnetosphere. Because of the frozen-in condition, field lines which connect the NS surface to infinity before merger must also connect the BH horizon to infinity after merger, while closed field lines are swallowed along with the NS. Hence a magnetic field is anchored onto the BH merger remnant. For a resistive plasma, the field will decay on the resistive timescale of the magnetosphere. As a consequence, the remnant BH could generate an electromagnetic signature through the Blandford-Znajek (BZ) mechanism ([Blandford & Znajek 1977](#); [Lyutikov & McKinney 2011](#)).

The initial BZ power can be written in terms of the magnetic flux deposited onto the BH horizon

as

$$P_{\text{BZ}} \sim \frac{\phi^2}{4\pi c} \left(\frac{Sc}{R_H(S)} \right)^2 \quad (7.28)$$

$$\sim 3 \times 10^{42} \text{ ergs}^{-1} \text{S}^2 \left(\frac{B_{\text{NS}}}{10^{12} \text{G}} \right)^2 \left(\frac{2\pi/\Omega_{\text{orb}}}{1 \text{msec}} \right)^{-2} \left(\frac{R_H(S)}{GM/c^2} \right)^{-2},$$

where S is the dimensionless BH spin related to the BH angular momentum by $J = SGM^2/c$, $R_H(S)$ is the spin dependent horizon radius, and $2\pi/\Omega_{\text{orb}}$ is the binary orbital period. In the second line we have approximated the magnetic flux thrown onto the BH as the flux of open magnetic-field lines at the NS polar caps (Goldreich & Julian 1969; Lyutikov & McKinney 2011),

$$\phi = 2\pi B_{\text{NS}} R_{\text{NS}}^2 \sin^{-1} \left(\frac{R_{\text{NS}} \Omega}{c} \right), \quad (7.29)$$

where, in the single NS case, Ω is the NS spin angular frequency, but here the light cylinder, and hence the footprint of open field lines on the NS surface, is determined by the orbital velocity in addition to the NS spin. Approximating Ω as the orbital angular frequency near merger, Figure 7.6 plots the initial power available to the post merger BH as a function of BH spin.

Notice that the post merger BZ power scales as M^{-2} through $R_H(S)$ whereas the usual BZ power scales as M^2 . The BZ power depends on the square of the magnetic flux deposited onto the BH, which in the standard case, scales with the squared black hole surface area M^4 ; adding also the dependence on horizon angular velocity, which scales as M^{-2} , gives the usual M^2 scaling. In the BHNS merger case however, the magnetic flux is set not by the BH size, but by the available flux brought in by the NS, so indeed larger BHs emit less BZ power.

Such a post-merger event will likely generate a relativistically beamed jet which peaks at maximum luminosity given by Figure 7.6 and then decays with the decaying BH magnetosphere. If the BH can hold onto the magnetosphere for a long enough time, such an event might generate a

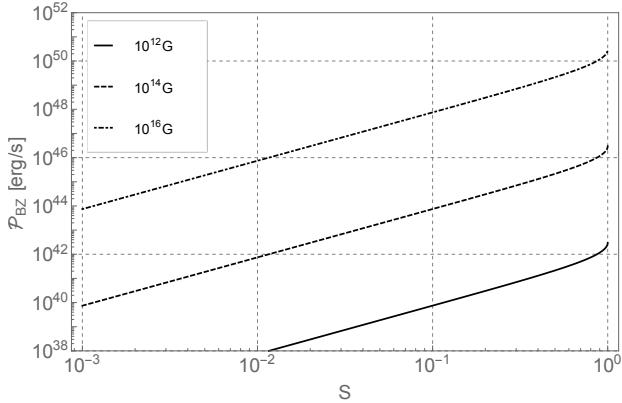


Figure 7.6: The power available to the post-merger, spinning BH remnant as a function of remnant spin and NS magnetic-field strength. This power is generated from the Blandford-Znajek process and the flux of open NS magnetic-field lines, Eq. (7.28). This maximal power will decay as the remnant magnetosphere decays on the resistive timescale.

type of afterglow to the BHNS merger. Assuming that the post merger signal begins at the same time as fireball expansion, at merger, then the peak luminosity of the post merger signal would be observed $R_{\text{ph}}/c \sim 0.2 \text{ msec} \sqrt{B/10^{12}\text{G}}$ before the blackbody fireball emission. We mention this as it is of observational interest and an avenue to pursue in developing the full portrait of the BH battery.

7.5 Observability

The Fermi Gamma-Ray Burst Monitor (GBM [FER 2015](#)) is well suited for detecting the transients described above. It has an energy range of $0.008 \rightarrow 30 \text{ MeV}$, capturing the peak of emission predicted for binaries with 10^{12} G to $\sim 10^{16} \text{ G}$ NS magnetic-field strengths (Eq. 7.25). It has a $2\mu\text{s}$ timing resolution, sufficient to resolve the $\gtrsim 1\text{msec}$ bursts. The Fermi GBM also operates with a nearly full-sky field of view (currently operating at 9.5 sr with a 10 sr goal), important for catching such possibly rare transients.

We estimate the photon flux at the instrument by assuming emission from a blackbody with

Doppler boosted (Eq. 7.24) and cosmologically redshifted temperature. The photon flux at the GBM is

$$\begin{aligned}
F_{\text{obs}} &= 2\pi \int_0^{\theta_c} \int_{\nu_{\min}}^{\nu_{\max}} \frac{2\nu^2}{c^2} \frac{\cos \theta \sin \theta d\nu d\theta}{\exp \left[\frac{h\nu(1+z)}{kT_{\text{eff}}(\theta)} \right] - 1} \\
\theta_c &= \frac{R_{\text{ph}}}{d_A(z)} \\
d_A(z) &= \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}} \\
T_{\text{eff}}(\theta) &= T_{\text{ph}} \left[\gamma \left(1 - \frac{v}{c} \cos \left(\frac{\pi}{2} \frac{\theta}{\theta_c} \right) \right) \right]^{-1}
\end{aligned} \tag{7.30}$$

where d_A is the angular diameter distance in the 2015 Planck cosmology with $\Omega_M = 0.308$, $\Omega_\Lambda = 1 - \Omega_m$, and $H_0 = 67.8 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration et al. 2015), and where integration is over the solid angle of the photosphere at redshift z , and over the frequency limits of the GBM. We use the minimum detectable flux for the GBM to solve $F_{\text{obs}}(z) = F_{\text{min}}$ for the maximum observable redshift to which BHNS transients could be observed. Using the GBM on-board trigger sensitivity, $F_{\text{min}} = 0.71 \text{ cm}^{-2} \text{ s}^{-1}$ (FER 2015), we find,

$$\begin{aligned}
d_M^{\max}(B_{\text{NS}} = 10^{12}\text{G}) &\sim 9 \text{ Mpc}; & z^{\max} &= 0.002 \\
d_M^{\max}(B_{\text{NS}} = 10^{13}\text{G}) &\sim 49 \text{ Mpc}; & z^{\max} &= 0.011 \\
d_M^{\max}(B_{\text{NS}} = 10^{14}\text{G}) &\sim 270 \text{ Mpc}; & z^{\max} &= 0.064 \\
d_M^{\max}(B_{\text{NS}} = 10^{15}\text{G}) &\sim 1.3 \text{ Gpc}; & z^{\max} &= 0.339 \\
d_M^{\max}(B_{\text{NS}} = 10^{16}\text{G}) &\sim 5.1 \text{ Gpc}; & z^{\max} &= 1.886,
\end{aligned} \tag{7.31}$$

which we have quoted in terms of the co-moving radial distance d_M and the corresponding redshift. The $\gtrsim 10^{13}\text{G}$ binaries are detectable out to beyond the initial LIGO volume, while only the $\gtrsim 10^{14.5}\text{G}$ binaries are detectable out to approximately the advanced LIGO volume for NSBH mergers

B_{NS} [G]	Minimum	Expected	Maximum
10^{12}	1.4×10^{-6}	6.9×10^{-5}	2.3×10^{-3}
10^{13}	2.4×10^{-4}	1.2×10^{-2}	0.4
10^{14}	3.9×10^{-2}	2.0	66
10^{15}	5.0	248	8.3×10^3
10^{16}	267	1.3×10^4	4.5×10^5

Table 7.1: Expected number of Fermi GBM events in units of [yr^{-1}] $f_{\text{fb}}(B_{\text{NS}})$ where $f_{\text{fb}}(B_{\text{NS}})$ is the fraction of BHNS coalescences with NS magnetic-field strength B_{NS} and which will not tidally disrupt the NS and will generate the signal predicted here. B_{NS} is the NS surface magnetic-field strength.

(redshift $z \sim 0.1$; [Abadie et al. 2010](#)).

To estimate the number of expected detections out to z^{max} we need to know the rate of BHNS mergers as a function of B_{NS} , and we need to know what fraction of those mergers generate the signal derived here. BHNS coalescence rates are computed by Ref. [Abadie et al. \(2010\)](#). They predict between 6×10^{-4} and 1 BHNS coalescences per Mpc^3 per Myr with a most probable rate of 0.03 per Mpc^3 per Myr. Estimating the number of non-disrupting BHNS mergers with a given NS magnetic-field strength is beyond the scope of the present work. Instead, we parameterize the fraction of BHNS mergers which generate the signal predicted here as $f_{\text{fb}}(B_{\text{NS}})$. Using the calculated maximum detection redshifts we calculate the co-moving detection volume. Using this maximum detection volume, coalescence rates with $f_{\text{fb}} = 1$, and a 10 sr field of view, Table 7.1 lists the expected number of events that FERMI GBM could detect per year.

For BHNS binaries with $B_{\text{NS}} \lesssim 10^{14}\text{G}$, these optimistic, expected rates of detection drop below 1 per year. To probe the binaries with $B_{\text{NS}} \gtrsim 10^{13}\text{G}$ at a rate of $\sim 1.0 f_{\text{fb}} \text{ yr}^{-1}$, future X-ray instruments must have full-sky sensitivities of $\sim 10 \times$ the FERMI GBM. They must have sensitivities $\sim 600 \times$ the GBM to reach $B_{\text{NS}} \gtrsim 10^{12}\text{G}$ binaries at the same rate.

Assuming our model roughly captures the BHNS luminosity and spectrum, there are two options for BHNS mergers with $B_{\text{NS}} \gtrsim 10^{14}\text{G}$. Either we have already observed the high-magnetic

field BHNS fireballs as a subclass of short gamma-ray bursts (sGRBs), or we have not, and the fraction of non-disrupting BHNS binaries with such magnetic fields f_{fb} is very small.

The BHNS fireball could compose a subclass of the sGRB population if a, yet unknown, mechanism saturates NS field strengths to maximal $\geq 10^{15}\text{G}$ values near merger, then the rates predicted here become comparable to the inferred (beaming angle dependent) rates of sGRBs, $8 \rightarrow 1100 \text{ Gpc}^3 \text{ yr}^{-1}$ from Swift measurements ([Coward et al. 2012](#)). The analysis of section §[7.3](#) allows emission from $\sim 10^{15}\text{G}$ fireballs to be of order seconds, consistent with sGRB timescales.

Alternatively, evidence has been found that a class of sGRBs, making up 10 to 25 per cent of the total, may be at a near $z \leq 0.025$ [Tanvir et al. \(2005\)](#). These would be a different class than those sGRBs for which distances can be measured out to a Gpc through afterglows (*e.g.*, [Berger et al. 2005](#)). The implication is that a class of sGRBs has a much lower luminosity engine, which could be powered by the $B_{\text{NS}} \sim 10^{13}\text{G}$ BHNS transients discussed here. This possibility, however, requires an explanation for increased rates of BHNS mergers in the local universe.

If the BHNS fireball is not a subset of the observed GRB population, then, based on the present non-detection, we may place limits on the fraction of binaries which carry $B_{\text{NS}} \gtrsim 10^{14}\text{G}$, to merger. Using the expected rates and the total operation time of the GBM at its current sensitivity (~ 5 years) we find that $f_{\text{fb}}(\geq 10^{15}\text{G}) \lesssim 10^{-3}$ and $f_{\text{fb}}(\geq 10^{16}\text{G}) \lesssim 10^{-4}$. Where the inequalities assume that f_{fb} is a steeply decreasing function of magnetic-field strength for $B_{\text{NS}} > 10^{14}\text{G}$.

Another possibility is that these upper limits for the luminosity of the signal are indeed over-estimates and mechanisms such as screening in the magnetosphere greatly damp power output; continued electromagnetic, as well as future gravitational wave, observations will test this. Concurrently, further modeling of the BHNS magnetosphere would hone the expected signal and the derived rates of detection.

The above analysis relies on a choice of $R_0 = GM/c^2$ for the size scale of energy injection. This is a natural choice, however we discuss briefly the dependence of our results on injection

radius. If we go with a large value of $R_0 = 2GM/c^2$, then less energy is injected over a larger volume and the initial temperature of the fireball drops to $18 \text{ keV } (B_{\text{NS}}/10^{12}G)^{1/2}$ from our fiducial $85 \text{ keV } (B_{\text{NS}}/10^{12}G)^{1/2}$ for $R_0 = GM/c^2$. This corresponds to a peak black body temperature of $52 \text{ keV } (B_{\text{NS}}/10^{12}G)^{1/2}$, down from the fiducial $0.24 \text{ MeV } (B_{\text{NS}}/10^{12}G)^{1/2}$. These lower energies are still within the energy range of the Fermi GBM, but a combination of less injected energy, smaller photosphere sizes (Figure 7.5) (and hence smaller expansion speed at the photosphere) decrease the maximum observable distance of the fireball by a factor of ~ 3 and also decreases the expected rates (Table 7.1) by one to two orders of magnitude.

7.6 Conclusion

We have used BH-battery energetics to argue that near merger, a BHNS will produce an electromagnetic transient. A spectrum of high energy ($\sim \text{TeV}$) curvature radiation will escape the magnetosphere before the last 0.1s ($B/10^{12}\text{G}$) of inspiral. This signature will only reach luminosities of $\sim 10^{38} \text{ erg s}^{-1} (B/10^{12}\text{G})^{1/2}$ before being quenched by pair-production and fueling the more luminous fireball transient. The expanding fireball will become transparent and emit as a blackbody in the X-Ray to γ -ray range for of order $10^{-3} \rightarrow 10$ seconds depending on the NS magnetic-field strength. The observed luminosity can peak at $10^{45} \text{ erg s}^{-1}$ for a 10^{12}G NS magnetic field or up to $10^{53} \text{ erg s}^{-1}$ for magnetar strength fields. If the BH can hold onto the NS magnetic fields after merger through a slow decay of the magnetosphere (Lyutikov & McKinney 2011), a spinning remnant BH could power a relativistic jet with bolometric luminosity up to two orders of magnitude lower than the fireball luminosity, peaking at $\sim 0.2 \text{ msec} \sqrt{B/10^{12}\text{G}}$ before the observed fireball emission, and decaying on the unknown resistive timescale of the magnetosphere.

The prospects for detecting the bright, fireball transient are dependent on the (unknown) distribution of NS magnetic-field strengths B_{NS} at merger. To explore these prospects, we have left

the NS surface magnetic-field strength as a free parameter. Conversely, BHNS merger rates allow our model to put constraints on B_{NS} at merger. Given predicted BHNS merger rates, the majority of BHNS mergers must have $B_{\text{NS}} > 10^{14}\text{G}$ to be detectable by Fermi GBM at the rate of $\sim 1 \text{ yr}^{-1}$. If $B_{\text{NS}} \lesssim 10^{12}$ at merger, as might be expected from the observed pulsar magnetic-field strengths ([Bhattacharya & van den Heuvel 1991](#)), a future X-ray instrument would need a full sky sensitivity of $\gtrsim 600$ the present FERMI GBM capabilities to detect these EM signatures of BHNS coalescence. If ordered magnetic fields are amplified to $\gtrsim 10^{15}\text{G}$ at merger, then expected FERMI GBM detection rates for the signature in this study climb to rival the gamma-ray burst rate, and may be a subclass of sGRBs ([Coward et al. 2012](#)).

Any observation of a BH battery transient would be exciting in its own right. With advanced LIGO now operational, the EM counterpart to BHNS coalescence has additional payout potential, offering unique information to extend the astronomical reach of the gravitational-wave observatories.

Acknowledgements: The authors thank Andrei Beloborodov, Brian Metzger, and Sean McWilliams for useful discussions. DJD acknowledges support from a National Science Foundation Graduate Research Fellowship under Grant No. DGE1144155. JL thanks the Tow Foundation for their support. JL was also supported by a Guggenheim Fellowship and is a Chancellor’s Fellow at Chapman University.

7.7 Parameter Dependence of Curvature Spectra

Figure 7.7 plots the curvature radiation spectra, identical to Figure 7.3, but for different values of the electron-energy power law index p , and the minimum electron Lorentz factor in the magnetosphere, γ_{\min} . We vary p from 1.0 to 3.0. We choose minimum Lorentz factors which bracket the range of plausible values: $\gamma_{\min} = 1$, and a minimum radiation-reaction limited Lorentz factor

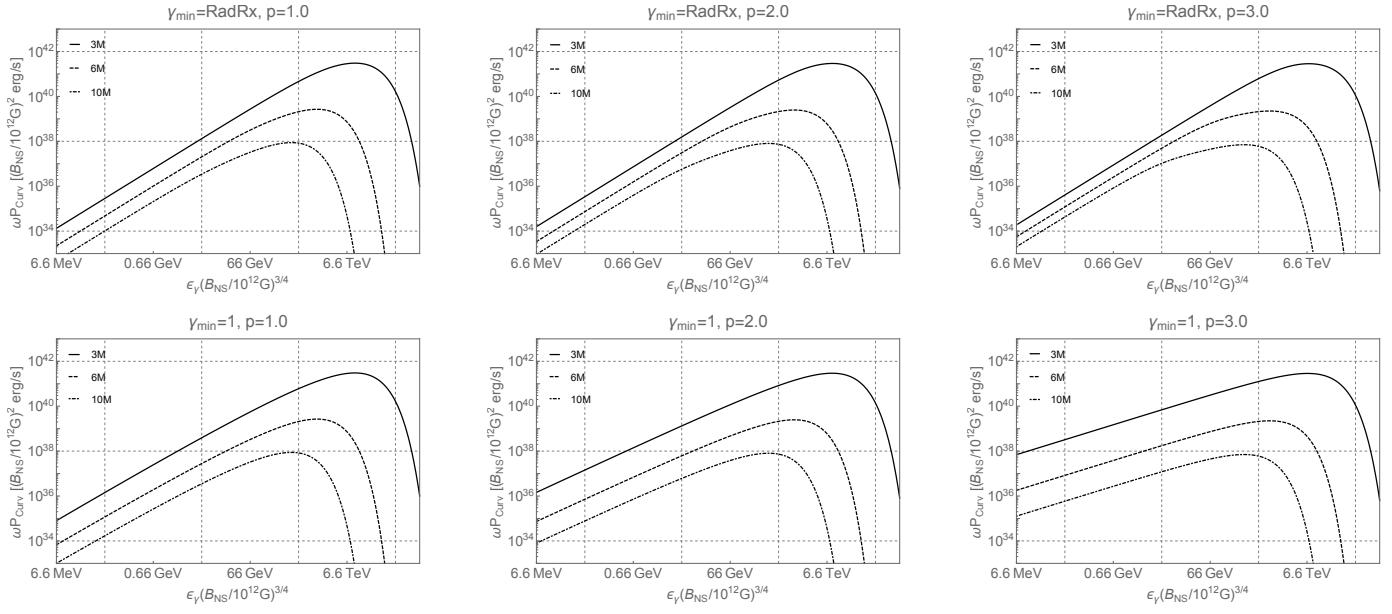


Figure 7.7: The spectrum of primary curvature radiation at times corresponding to binary separations of $10M$, $6M$, and $3M$ (dot-dashed, dashed, solid). Each panel is for the labeled minimum electron Lorentz factor and power law index p of electron energies. $\gamma_{\min} = \text{RadRx}$ refers to the radiation reaction limited Lorentz factor at the point of weakest electric field in the region connecting NS and BH (of order a few to 10 times smaller than the maximum Lorentz factor near merger).

which we compute with Eq. (7.13) but with electric field at the edge of the binary orbital light cylinder (Ω_{orb}/c) that falls off from its horizon value as r^{-2} (Thorne et al. 1986). Near merger this is only a few times smaller than the maximum γ computed from the horizon electric fields.

]

Bibliography

2009, FERMI GBM characteristics

2015, FERMI GBM characteristics

Abadie, J., Abbott, B. P., Abbott, R., et al. 2010, Classical and Quantum Gravity, 27, 173001

Armitage, P. J., & Natarajan, P. 2002, ApJL, 567, L9

—. 2005, ApJ, 634, 921

Artymowicz, P., & Lubow, S. H. 1994, ApJ, 421, 651

Artymowicz, P., & Lubow, S. H. 1996, ApJL, 467, L77

Barnes, J. E., & Hernquist, L. 1992, ARA&A, 30, 705

Barnes, J. E., & Hernquist, L. 1996, ApJ, 471, 115

Barthelmy, S. D., Barbier, L. M., Cummings, J. R., et al. 2005, SSR, 120, 143

Baumgarte, T. W., & Shapiro, S. L. 2003, ApJ, 585, 930

Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, Nature, 287, 307

Berger, E., Price, P. A., Cenko, S. B., et al. 2005, Nature, 438, 988

- Bhattacharya, D., & van den Heuvel, E. P. J. 1991, Phys. Rep., 203, 1
- Blandford, R. D., & Znajek, R. L. 1977, MNRAS, 179, 433
- Bryden, G., Chen, X., Lin, D. N. C., Nelson, R. P., & Papaloizou, J. C. B. 1999, ApJ, 514, 344
- Callegari, S., Mayer, L., Kazantzidis, S., et al. 2009, ApJL, 696, L89
- Chang, P., Strubbe, L. E., Menou, K., & Quataert, E. 2010, MNRAS, 407, 2007
- Chapon, D., Mayer, L., & Teyssier, R. 2013, MNRAS, 429, 3114
- Cheng, K. S., & Zhang, J. L. 1996, ApJ, 463, 271
- Coward, D. M., Howell, E. J., Piran, T., et al. 2012, MNRAS, 425, 2668
- Crida, A., Morbidelli, A., & Masset, F. 2006, Icarus, 181, 587
- Cuadra, J., Armitage, P. J., Alexander, R. D., & Begelman, M. C. 2009, MNRAS, 393, 1423
- Curé, M., Rial, D. F., Cassetti, J., Christen, A., & Boffin, H. M. J. 2015, A&A, 573, A86
- Dall’Osso, S., Israel, G. L., & Stella, L. 2006, A&A, 447, 785
- . 2007, A&A, 464, 417
- D’Angelo, G., & Lubow, S. H. 2008, ApJ, 685, 560
- De Rosa, R. J., Patience, J., Ward-Duong, K., et al. 2014, MNRAS, 445, 3694
- Degenaar, N., Miller, J. M., Kennea, J., Gehrels, N., & Wijnands, R. 2012, ArXiv e-prints, arXiv:1210.7237
- del Valle, L., & Escala, A. 2012, ApJ, 761, 31

- . 2014, ApJ, 780, 84
- Dong, R., Rafikov, R. R., & Stone, J. M. 2011a, ApJ, 741, 57
- . 2011b, ApJ, 741, 57
- Dong, R., Rafikov, R. R., Stone, J. M., & Petrovich, C. 2011c, ApJ, 741, 56
- . 2011d, ApJ, 741, 56
- D’Orazio, D. J., Haiman, Z., Duffell, P., Farris, B. D., & MacFadyen, A. I. 2015a, MNRAS, 452, 2540
- D’Orazio, D. J., Haiman, Z., & MacFadyen, A. 2013, MNRAS, 436, 2997
- D’Orazio, D. J., Haiman, Z., & Schiminovich, D. 2015b, Nature, 525, 351
- D’Orazio, D. J., & Levin, J. 2013, PRD, 88, 064059
- Dotti, M., Colpi, M., Haardt, F., & Mayer, L. 2007, MNRAS, 379, 956
- Dotti, M., Merloni, A., & Montuori, C. 2015, MNRAS, 448, 3603
- Dotti, M., Sesana, A., & Decarli, R. 2012, Advances in Astronomy, 2012, 940568
- Duffell, P. C. 2015, ApJL, 807, L11
- Duffell, P. C., Haiman, Z., MacFadyen, A. I., D’Orazio, D. J., & Farris, B. D. 2014, ApJL, 792, L10
- Duffell, P. C., & MacFadyen, A. I. 2011, ApJS, 197, 15
- . 2012, ApJ, 755, 7
- . 2013, ApJ, 769, 41

Dunhill, A. C., Cuadra, J., & Dougados, C. 2015, MNRAS, 448, 3545

Dürmann, C., & Kley, W. 2015, A&A, 574, A52

East, W. E., Pretorius, F., & Stephens, B. C. 2012, PRD, 85, 124009

Edgar, R. G. 2006, e-print arxiv:astro-ph/0609756

—. 2008, ArXiv e-prints, arXiv:0807.0625

Eracleous, M., Boroson, T. A., Halpern, J. P., & Liu, J. 2012, ApJS, 201, 23

Erber, T. 1966, Reviews of Modern Physics, 38, 626

Erwin, P., & Sparke, L. S. 1999, ApJ, 521, 798

Escala, A., Larson, R. B., Coppi, P. S., & Mardones, D. 2005, ApJ, 630, 152

Etienne, Z. B., Liu, Y. T., Shapiro, S. L., & Baumgarte, T. W. 2009, PRD, 79, 044024

Etienne, Z. B., Paschalidis, V., & Shapiro, S. L. 2012, PRD, 86, 084026

Faber, J. A., Baumgarte, T. W., Shapiro, S. L., & Taniguchi, K. 2006, ApJL, 641, L93

Farris, B. D., Duffell, P., MacFadyen, A. I., & Haiman, Z. 2014, ApJ, 783, 134

—. 2015a, MNRAS, 447, L80

—. 2015b, MNRAS, 446, L36

Farris, B. D., Gold, R., Paschalidis, V., Etienne, Z. B., & Shapiro, S. L. 2012, Physical Review Letters, 109, 221102

Farris, B. D., Liu, Y. T., & Shapiro, S. L. 2011, PRD, 84, 024024

Ferrarese, L., & Ford, H. 2005, SSR, 116, 523

Foucart, F. 2012, PRD, 86, 124007

Fryxell, B., Olson, K., Ricker, P., et al. 2000, The Astrophysical Journal Supplement Series, 131, 273

Fung, J., Shi, J.-M., & Chiang, E. 2014, ApJ, 782, 88

Gergely, L. Á., & Biermann, P. L. 2012, submitted to PRD; eprint arXiv:1208.5251

Giacomazzo, B., Perna, R., Rezzolla, L., Troja, E., & Lazzati, D. 2013, ApJL, 762, L18

Goldreich, P., & Julian, W. H. 1969, ApJ, 157, 869

Goldreich, P., & Lynden-Bell, D. 1969, ApJ, 156, 59

Goldreich, P., & Tremaine, S. 1979, ApJ, 233, 857

—. 1980, ApJ, 241, 425

Goodman, J. 1986, ApJL, 308, L47

Goodman, J. 2003, MNRAS, 339, 937

Goodman, J., & Rafikov, R. R. 2001, ApJ, 552, 793

Gould, A., & Rix, H.-W. 2000, ApJL, 532, L29

Graham, M. J., Djorgovski, S. G., Stern, D., et al. 2015a, Nature, 518, 74

—. 2015b, MNRAS, 453, 1562

Haehnelt, M. G., & Kauffmann, G. 2002, MNRAS, 336, L61

- Haiman, Z., Kocsis, B., & Menou, K. 2009, ApJ, 700, 1952
- Haiman, Z., Kocsis, B., Menou, K., Lippai, Z., & Frei, Z. 2009, Classical and Quantum Gravity, 26, 094032
- Harry, G. M., & LIGO Scientific Collaboration. 2010, Classical and Quantum Gravity, 27, 084006
- Hayasaki, K., Mineshige, S., & Sudou, H. 2007, Publications of the Astronomical Society of Japan, 59, 427
- Heras, J. A. 1998, Physics Letters A, 237, 343
- Hinkley, S., Kraus, A. L., Ireland, M. J., et al. 2015, ApJL, 806, L9
- Ivanov, P. B., Papaloizou, J. C. B., & Polnarev, A. G. 1999, MNRAS, 307, 79
- Jackson, J. D. 1991, Classical Electrodynamics, 3rd edn. (New Jersey, USA: John Wiley & Sons, Inc.)
- Kanagawa, K. D., Tanaka, H., Muto, T., Tanigawa, T., & Takeuchi, T. 2015, MNRAS, 448, 994
- Kaspi, V. M., & Kramer, M. 2016, ArXiv e-prints, arXiv:1602.07738
- Kiuchi, K., Sekiguchi, Y., Kyutoku, K., et al. 2015, PRD, 92, 064034
- Kley, W., & Dirksen, G. 2006, A&A, 447, 369
- Kley, W., & Nelson, R. P. 2012, ARA&A, 50, 211
- Kocsis, B., Frei, Z., Haiman, Z., & Menou, K. 2006, ApJ, 637, 27
- Kocsis, B., Haiman, Z., & Loeb, A. 2012a, MNRAS, 427, 2680
- . 2012b, MNRAS, 427, 2660

- Kocsis, B., Haiman, Z., & Menou, K. 2008, ApJ, 684, 870
- Kocsis, B., Yunes, N., & Loeb, A. 2011, PRD, 84, 024032
- Komossa, S. 2006, Memorie della Società Astronomica Italiana, 77, 733
- Kormendy, J., & Richstone, D. 1995, ARA&A, 33, 581
- Kovács, T. 2013, MNRAS, 430, 2755
- Lai, D. 2012, ApJL, 757, L3
- Laine, R. O., & Lin, D. N. C. 2012, ApJ, 745, 2
- Laine, R. O., Lin, D. N. C., & Tagger, M. 2012, in European Planetary Science Congress 2012, EPSC2012–744
- Landau, L. D., & Lifschitz, E. M. 1959, Fluid Mechanics (Addison-Wesley)
- Lee, W. H., Ramirez-Ruiz, E., & Granot, J. 2005, ApJL, 630, L165
- Li, J., Ferrario, L., & Wickramasinghe, D. 1998, ApJL, 503, L151
- Lin, D. N. C., & Papaloizou, J. 1986, ApJ, 307, 395
- Lin, D. N. C., & Papaloizou, J. C. B. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine, 749–835
- Lippai, Z., Frei, Z., & Haiman, Z. 2009, ApJ, 701, 360
- Lithwick, Y., & Sari, R. 2001, ApJ, 555, 540
- Liu, T., Gezari, S., Heinis, S., et al. 2015, ApJL, 803, L16
- Lodato, G., Nayakshin, S., King, A. R., & Pringle, J. E. 2009, MNRAS, 398, 1392

- Lommen, A. N. 2012, Journal of Physics Conference Series, 363, 012029
- Lubow, S. H. 1991a, ApJ, 381, 259
- . 1991b, ApJ, 381, 268
- Lützgendorf, N., Kissler-Patig, M., Gebhardt, K., et al. 2012, ArXiv e-prints, arXiv:1212.3475
- Lyutikov, M. 2011, PRD, 83, 064001
- Lyutikov, M., & McKinney, J. C. 2011, PRD, 84, 084019
- MacDonald, D., & Thorne, K. S. 1982, MNRAS, 198, 345
- MacDonald, D. A., & Suen, W.-M. 1985, PRD, 32, 848
- MacFadyen, A. I., & Milosavljević, M. 2008, ApJ, 672, 83
- Mayer, L. 2013, Classical and Quantum Gravity, 30, 244008
- Mayer, L., Kazantzidis, S., Madau, P., et al. 2007, Science, 316, 1874
- McKernan, B., Ford, K. E. S., Kocsis, B., & Haiman, Z. 2013, MNRAS, 432, 1468
- McKernan, B., Ford, K. E. S., Lyra, W., & Perets, H. B. 2012, MNRAS, 425, 460
- McWilliams, S. T., & Levin, J. 2011, ApJ, 742, 90
- Milosavljević, M., & Phinney, E. S. 2005, ApJ, 622, L93
- Mingarelli, C. M. F., Levin, J., & Lazio, T. J. W. 2015, ApJL, 814, L20
- Murray, C. D. 1994, Icarus, 112, 465
- Murray, C. D., & Dermott, S. F. 1999, Solar system dynamics

- . 2000, Solar System Dynamics (Cambridge University Press; Cambridge, UK)
- Narayan, R., Paczynski, B., & Piran, T. 1992a, ApJL, 395, L83
- . 1992b, ApJL, 395, L83
- Nelson, R. P., Papaloizou, J. C. B., Masset, F., & Kley, W. 2000, MNRAS, 318, 18
- Nixon, C. J., Cossins, P. J., King, A. R., & Pringle, J. E. 2011, MNRAS, 412, 1591
- Nixon, C. J., King, A. R., & Pringle, J. E. 2011, MNRAS, 417, L66
- Noble, S. C., et al. 2012, ApJ, 755, 51
- Olausen, S. A., & Kaspi, V. M. 2014, ApJS, 212, 6
- Orosz, J. A., , Welsh, W. F., & et. al. 2012, Science, 337, 1511
- Özel, F., Psaltis, D., Narayan, R., & McClintock, J. E. 2010, ApJ, 725, 1918
- Paczynski, B. 1986, ApJL, 308, L43
- Paczynski, B., & Rudak, B. 1980, Acta Astron, 30, 237
- Palenzuela, C., Bona, C., Lehner, L., & Reula, O. 2011, Classical and Quantum Gravity, 28, 134007
- Palenzuela, C., Garrett, T., Lehner, L., & Liebling, S. L. 2010, PRD, 82, 044045
- Palenzuela, C., Lehner, L., Ponce, M., et al. 2013, ArXiv e-prints, arXiv:1301.7074
- Papaloizou, J., & Pringle, J. E. 1977, MNRAS, 181, 441
- Papaloizou, J. C. B., Nelson, R. P., & Snellgrove, M. D. 2004, MNRAS, 350, 829

Paschalidis, V., Etienne, Z. B., & Shapiro, S. L. 2013, PRD, 88, 021504

Payne, M. J., & Lodato, G. 2007, MNRAS, 381, 1597

Pe'er, A., & Ryde, F. 2011, ApJ, 732, 49

Penna, R. F. 2015a, PRD, 92, 084017

—. 2015b, PRD, 91, 084044

Peter Rowe, E. G., & Rowe, G. T. 1987, Phys. Rep., 149, 287

Peters, P. C. 1964, Physical Review, 136, 1224

Pichardo, B., Sparke, L. S., & Aguilar, L. A. 2005, MNRAS, 359, 521

Piro, A. L. 2012, ApJ, 755, 80

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, ArXiv e-prints, arXiv:1502.01589

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical Recipes: The Art of Scientific Computing (Cambridge University Press; Cambridge, UK)

Preto, M., Berentzen, I., Berczik, P., & Spurzem, R. 2011, ApJL, 732, L26

Pringle, J. E. 1981, ARA&A, 19, 137

Rezzolla, L., Giacomazzo, B., Baiotti, L., et al. 2011, ApJL, 732, L6

Ribaric, M., & Sustersic, L. 1995, SIAM Journal of Applied Mathematics, 593

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 20

- Robertson, B., Cox, T. J., Hernquist, L., et al. 2006, ApJ, 641, 21
- Roedig, C., Dotti, M., Sesana, A., Cuadra, J., & Colpi, M. 2011, MNRAS, 415, 3033
- Roedig, C., Sesana, A., Dotti, M., et al. 2012a, A&A, 545, A127
- . 2012b, A&A, 545, A127
- Rudak, B., & Paczynski, B. 1981, Acta Astron, 31, 13
- Ruderman, M. A., & Sutherland, P. G. 1975, ApJ, 196, 51
- Schnittman, J. D. 2010, ApJ, 724, 39
- Sesana, A., Haardt, F., Madau, P., & Volonteri, M. 2005, ApJ, 623, 23
- Sesana, A., Roedig, C., Reynolds, M. T., & Dotti, M. 2012, MNRAS, 420, 860
- Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
- Shapiro, S. L. 2010, PRD, 81, 24019
- Shen, Y., & Loeb, A. 2010, ApJ, 725, 249
- 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
- Shibata, M., & Taniguchi, K. 2008, PRD, 77, 084015
- Shibata, M., & Uryu, K. 2007, Classical and Quantum Gravity, 24, 125
- Springel, V., Di Matteo, T., & Hernquist, L. 2005, ApJL, 620, L79
- Svensson, R. 1987, MNRAS, 227, 403

- Syer, D., & Clarke, C. J. 1995, MNRAS, 277, 758
- Tanaka, T., Menou, K., & Haiman, Z. 2012, MNRAS, 420, 705
- Tanvir, N. R., Chapman, R., Levan, A. J., & Priddey, R. S. 2005, Nature, 438, 991
- Thorne, K. S., Price, R. H., & MacDonald, D. A. 1986, Black holes: The membrane paradigm (Yale University Press)
- Triaud, A. H. M. J., Gillon, M., Selsis, F., et al. 2013, ArXiv e-prints, arXiv:1304.7248
- Tsalmantza, P., Decarli, R., Dotti, M., & Hogg, D. W. 2011, ApJ, 738, 20
- Tsang, D. 2013, ApJ, 777, 103
- Tsang, D., Read, J. S., Hinderer, T., Piro, A. L., & Bondarescu, R. 2012, Physical Review Letters, 108, 011102
- Vietri, M. 1996, ApJL, 471, L95
- Volonteri, M., Haardt, F., & Madau, P. 2003a, ApJ, 582, 559
- . 2003b, ApJ, 582, 559
- Ward, W. R. 1997, Icarus, 126, 261
- Wu, K., Cropper, M., Ramsay, G., & Sekiguchi, K. 2002, MNRAS, 331, 221
- Yunes, N., Kocsis, B., Loeb, A., & Haiman, Z. 2011, Physical Review Letters, 107, 171103
- Zhu, Z., Stone, J. M., & Rafikov, R. R. 2013, ApJ, 768, 143