ADVENTURES IN BLACK HOLE BINARY WAVEFORM MODELING

Siddharth Mahesh

Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Astrophysics

Sean T. McWilliams, Ph.D., Chair
Zachariah Etienne, Ph.D.
Aldo Romero, Ph.D.
Tudor Stanescu, Ph.D.

Department of Physics and Astronomy

Morgantown, West Virginia
2026

Keywords: Black Holes, Gravitational Waves, Waveform Modeling

ABSTRACT

Adventures in Black Hole Binary Waveform Modeling

by Siddharth Mahesh

The modeling of black hole binaries requires accurate descriptions of both their relativistic vacuum dynamics and their interaction with surrounding astrophysical environments. Although these problems are often treated separately, this dissertation develops a unified view: key observables of black hole binary systems can be understood by identifying the dynamical instabilities that control their evolution. In the vacuum problem, the relevant unstable structure is associated with the remnant spacetime and its role in shaping the merger-ringdown gravitational waveform. In the environmental problem, the relevant instability is the epicyclic instability of orbits in the binary potential, which controls the location and evolution of the circumbinary gap.

The first part of this dissertation develops the Spinning Effective-to-Backwards One Body framework, a hybrid waveform model that combines an Effective-One-Body inspiral with a Backwards-One-Body description of the merger and ringdown. This construction reduces reliance on phenomenological numerical-relativity calibration while retaining competitive waveform accuracy for aligned-spin black hole binaries. The model provides a physically motivated map from the inspiral to the remnant-dominated regime and offers a transparent path toward more interpretable waveform models for next-generation gravitational-wave astronomy.

The second part investigates the interaction between black hole binaries and circumbinary accretion disks. Using Newtonian hydrodynamic simulations and analytical epicyclic-stability theory, this dissertation shows that circumbinary gaps are maintained on orbital timescales by instabilities driven by the binary potential, rather than solely by long-timescale resonant torque balance. This framework is extended to inclined disk-binary systems, where it predicts an unstable sector near intermediate inclinations and explains persistent time-dependent disk oscillations.

Together, these results show that instability-based modeling provides a common analytical language for understanding gravitational-wave signals and electromagnetic environments of black hole binaries. This perspective strengthens the physical interpretability of waveform models, clarifies the dynamics of circumbinary disks, and contributes to the theoretical foundation needed for multi-messenger observations with future detectors.

To my parents, Sudha and Mahesh, and to Maggie, for their unwavering support and encouragement throughout this endeavour. To Dottie, for tolerating my all-nighters and Marmie, for all the proofreading.

Acknowledgments

This dissertation is the product of the excellent mentorship I received from my advisors, Dr. Sean T. McWilliams and Dr. Zachariah Etienne. I am grateful to them for their guidance, support, and encouragement throughout my graduate studies. I would also like to thank them for providing me with the opportunities to present my research at multiple conferences and workshops. I would also like to thank my committee members, Dr. Aldo Romero and Dr. Tudor Stanescu, for their valuable feedback and insights.

I am also immensely grateful to my current and former colleagues Tyler Knowles, Jacob Turner, Regis John, Ashok Choudhary, Michal Pirog, Anuj Kankani, Matthew Cerep, Suchindram Dasgupta, Shay Reddy, and Rithvik Ediyola for scientific discussions and contributing to my growth as a researcher during my time at WVU.

The research in this dissertation would not have been possible without the resources provided by the WVU High Performance Computing Center, particularly the Thorny Flat cluster.

Table of Contents

Acronyms

BHB	Black Hole Binary
BHPT	Black Hole Perturbation Theory
BOB	Backwards One-Body
CBD	Circumbinary Disk
EM	Electromagnetic
EOB	Effective One-Body
GR	General Relativity
GW	Gravitational Wave
GWTC	Gravitational-Wave Transient Catalog
HJE	Hamilton-Jacobi Equation
ICO	Innermost Circular Orbit
IMR	Inspiral-Merger-Ringdown
ISCO	Innermost Stable Circular Orbit
LAL	LIGO Algorithm Library
LIGO	Laser Interferometer Gravitational-Wave Observatory
LISA	Laser Interferometer Space Antenna
LVK	LIGO-Virgo-KAGRA
NR	Numerical Relativity
NRPY	Python/SymPy-Based Code Generation for Numerical Relativity... and Beyond!
PM	Post-Minkowskian
PN	Post-Newtonian
PTA	Pulsar Timing Array
QNM	Quasi-Normal Mode
SEBOB	Spinning Effective-to-Backwards One Body
SEOBNRv5	Spin-aligned Effective One-Body Numerical Relativity version 5
SMBHB	Supermassive Black Hole Binary
STF	Symmetric, Trace-Free
SXS	Simulating eXtreme Spacetimes
TT	Transverse-Traceless

Nomenclature

Symbol	Units	Description
$g_{\mu\nu}$		Spacetime metric
$\eta_{\mu\nu}$		Minkowski metric
$h_{\mu\nu}$		Metric perturbation (gravitational wave strain)
$\Box$		d'Alembert operator
G		Gravitational constant
c		Speed of light
$T_{\mu\nu}$		Stress-energy tensor
t		Time
$\mathbf{x}$		Spatial coordinates
r		Radial distance
I_{ij}		Raw source quadrupole second moment
Q_{ij}		Symmetric trace-free source quadrupole moment
Q_{ij}^{TT}		Transverse-traceless projection of the radiative quadrupole
M		Mass of the system
R		Characteristic size of the system
R_s		Schwarzschild radius
a		Semi-major axis of the binary
ν		Symmetric mass ratio
ϵ		Post-Newtonian expansion parameter $(1/c^2)$
J		Total angular momentum
Ω		Orbital frequency
E		Binding energy
$\mathbf{q}$		Relative separation vector
$\mathbf{p}$		Relative momentum vector
μ		Reduced mass
v		Orbital velocity
Ω_{prec}		Spin precession frequency
$T_{\mu\nu}$		Stress-energy tensor
h_{lm}		Spherical harmonic modes of the strain

Chapter 1

Introduction

The observation of gravitational waves (GWs) from merging black hole binaries (BHBs) by the LIGO-Virgo-KAGRA (LVK) collaboration has revolutionized our understanding of the gravitationally driven merger of compact binaries [1]. These observations provide a unique laboratory for testing General Relativity in the strong gravitational regime and probing the formation and evolution of compact objects. However, fully realizing the potential of GW astronomy requires two critical ingredients: high-fidelity, computationally efficient waveform models and a comprehensive understanding of the environments in which these binaries reside. Within the scope of this thesis, we will limit our scope of study to black hole binaries and circumbinary accretion disks.

The science of gravitational waveform modeling involves the production of accurate template GW signals from the predictions of general relativity (GR). To build a successful template, one requires an accurate solution to the black hole binary spacetime and a computationally efficient reduced solution that maps the intrinsic parameters of the binary system to the observed waveform. Due to the nonlinear nature of GR, the knowledge of individual black hole spacetimes does not directly lead to the knowledge of the spacetime of a black hole binary. Analytically, this solution can be approximated using techniques such as post-Newtonian (PN) and post-Minkowskian (PM) expansions [2], and black hole perturbation theory (BHPT). Computationally, the field of numerical relativity (NR) has emerged as a powerful tool for solving the equations of GR using advanced high-performance computing techniques. Waveform models that transform the results of these approximations to GW templates are called approximants, and there are broadly three families of approximants that handle the full inspiral-merger-ringdown (IMR) signal: Phenomenological, Surrogate, Effective One Body (EOB) families.

The dynamics of black hole binaries can be broadly divided into three stages: the inspiral, the merger, and the ringdown. The inspiral stage is characterized by the bound orbital motion of the two separate black holes, with orbits decaying due to emission of gravitational waves. The ringdown stage is characterized by the decay of small perturbations about the merged remnant. The merger stage is the complex, intermediate stage where the spacetime transitions from two distinct black holes to the final merger remnant. The PN and PM approximations are valid in the early to late inspiral limit while the BHPT approximation is valid in the ringdown. While NR provides a unified picture of inspiral-merger-ringdown (IMR), it is computationally expensive and must be restricted to a small number of orbital cycles in the inspiral and a finite grid of binary parameters. The three aforementioned waveform families combine the results of these approximations to create models that are smooth across the domain of binary parameters.

It is important to note, at this point, that the combination of analytical and numerical approximations of GR is physically achievable if there exist overlapping domains of the binary evolution where two or more solutions are valid. In the early inspiral and ringdown phases, it is easy to translate NR results into the language of PN and BHPT approaches respectively. In the PN limit, there is a natural correspondence between the orbital frequency of the binary and the frequency of the gravitational waves. In the BHPT limit, the gravitational wave strain behaves like a superposition of damped sinusoids, indicative of the emission of quasi-normal modes (QNMs) from a remnant black hole. No such correspondence exists during the late inspiral and merger phases, where the physical description of the system transitions, in a highly non-linear manner, from two distinct compact objects to a merged remnant. Traditionally, waveform models have ignored the analytical study of this regime in favor of using NR results to interpolate between the analytically well-defined domains. The lack of a viable analytical solution presents a unique challenge in waveform modeling since there is no independent method to validate NR simulation results in this domain. This motivates the need for the development of analytical models for the merger component and a scheme for attaching them to inspiral and ringdown models.

Beyond the vacuum dynamics of the binary itself, next generation detectors like LISA [3], will detect binaries that realistically interact with their astrophysical environments. For supermassive black hole binaries (SMBHBs) detectable by the upcoming LISA mission, the interaction between the binary and the gas in its vicinity can lead to periodic EM signatures and may even accelerate the binary's evolution via gas torques. A central question in disk-binary interaction is the formation of the "circumbinary gap", a region of low density around the binary. Traditionally, this gap was thought to be maintained over large viscous timescales, leading to a "decoupling" stage in the

inspiral followed by a delayed post-merger "afterglow" [4]. However, numerical simulations of disk-binary systems have shown a dynamically rich picture of gap formation and evolution, with the circumbinary gap shrinking proportionately to the binary's inspiral and accreting gas into accretion disks around the individual black holes [5, 6]. This motivates the need for an alternative explanation for the dynamics of the circumbinary gap.

In the last two paragraphs, I summarized the motivation to explore two broad open problems in the modeling of the gravitational evolution of black hole binary systems. These problems manifest at completely different stages of the binary evolution and address two different paradigms within the field of *multi-messenger astrophysics*. Even in the absence of decoupling, the effect of astrophysical environments on binaries is limited to the inspiral since the characteristic speeds of the binary systems approach the relativistic limit as they transition to the merger phase. Thus, defining the goals of my dissertation as reconciling the analytical and numerical predictions for the circumbinary gap and developing an analytical model for the merger component of black hole binaries may seem like working on two disjoint areas of science. However, the underlying language required to resolve these problems share a common mathematical and analytical foundation - the impacts of instabilities in dynamical systems.

As will be shown in this dissertation, both these problems are solved by formulating a dynamical system - geodesics in black hole spacetimes and the restricted two-body problem. The key observable, the amplitude of the merger-ringdown strain or the location of the circumbinary gap, can be linked to a key dynamical instability, that of the black hole light ring or bound epicyclic orbits around Newtonian binaries. Not only are these resolutions physically sound and tied to the necessary timescales and mechanisms in the system, but they also provide much needed predictive power for future scientific research. This commonality was not immediately obvious to the broader scientific community and was a key discovery during the performance of the doctoral research. Thus, the dissertation has been devoted towards exploring a breadth of topics in multimessenger astrophysics while developing a depth of rigorous, analytical and numerical tools towards solving problems with the mathematics of dynamical systems. The central claim of this dissertation is that apparently distinct problems in black hole binary astrophysics can be clarified by identifying the dynamical instability that controls the relevant observable. For waveform modeling, this means connecting the merger-ringdown strain to the remnant-centered dynamics that emerge near merger. For circumbinary disks, this means connecting the gap location and disk response to the stability of perturbed orbits in the binary potential. This instability-based perspective provides a common analytical framework for modeling both gravitational-wave and electromagnetic signatures of black hole binaries.

Chapter 2 applies this instability-based perspective to the vacuum relativistic dynamics of black hole binaries. After reviewing the weak-field origin of gravitational radiation and the Effective-One-Body formalism, I present the SEBOB waveform model, which connects EOB inspiral dynamics to a BOB description of the merger-ringdown.

Chapter 3 applies the same perspective to binary environments. I first review the traditional resonant-torque picture of circumbinary gap formation and then show why it does not fully explain the short-timescale disk behavior seen in hydrodynamic simulations. I then present an alternative account based on epicyclic stability in the binary potential and extend this account from coplanar disks to inclined disk-binary systems.

Chapter 4 summarizes these results, provides an overview of the insights gained, and discusses broader implications and future research directions stemming from this instability-based perspective.

Chapter 2

First principles black hole binary waveform model

Following the broad introduction to gravitational waves, the necessity of accurate waveform models and the challenges in their development, this chapter dives into the mathematical formulation of one such class of models. Gravitational waves emerge as one possible solution to the equations of general relativity in the absence of matter. Physically, this solution also manifests as the description of the regions of spacetime far from a compact, localized distribution of matter[7]. Mathematically, the spacetime metric is written as a small perturbation about flat spacetime,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1,$$

where $\eta_{\mu\nu}$ is the Minkowski metric and $h_{\mu\nu}$ is the metric perturbation. It is useful to define the trace

$$h = \eta^{\alpha\beta} h_{\alpha\beta}$$

and the trace-reversed perturbation

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h.$$

After imposing the Lorenz gauge condition

$$\partial^{\mu} \bar{h}_{\mu\nu} = 0,$$

the linearized Einstein equations reduce to the sourced wave equation

$$\Box\bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu}. \tag{2.1}$$

Here, $\Box$ is the flat-spacetime d'Alembertian, G is Newton's gravitational constant, c is the speed of light, and $T_{\mu\nu}$ is the stress-energy tensor of the matter distribution. Inverting the wave operator with retarded boundary conditions yields

$$\bar{h}_{\mu\nu}(t,\mathbf{x}) = \frac{4G}{c^4}\int_\Omega d^3x' \frac{T_{\mu\nu}\left(t-\frac{|\mathbf{x}-\mathbf{x}'|}{c},\mathbf{x}'\right)}{|\mathbf{x}-\mathbf{x}'|}. \tag{2.2}$$

The integration domain Ω is assumed to be a compact volume occupied by the matter distribution, with the stress-energy tensor being zero outside this domain. The physical gravitational-wave strain is obtained from the spatial components after applying the transverse-traceless projection. Taking the far-zone limit $|\mathbf{x}| \gg |\mathbf{x}'|$, expanding the retarded solution, using stress-energy conservation, and projecting into TT gauge gives the leading quadrupole result.

The source quadrupole second moment is defined as

$$I_{ij}(t) = \int_\Omega d^3x'\, x'_i x'_j T_{00}(t,\mathbf{x}'). \tag{2.3}$$

where T_{00} specifically refers to the energy density component of the stress-energy tensor. This object is not trace-free. Its symmetric trace-free part is

$$Q_{ij}(t) = I_{ij}(t) - \frac{1}{3}\delta_{ij}I_{kk}(t). \tag{2.4}$$

The observable radiative field is obtained by applying the transverse-traceless projection in the wave zone. Thus the leading-order far-zone strain may be written schematically as

$$h_{ij}^{TT}(t,r) = \frac{2G}{c^4 r}\frac{d^2}{dt^2}Q_{ij}^{TT}(t-r/c) + O\left(\frac{1}{r^2}\right). \tag{2.5}$$

Here, I_{ij} denotes the raw source quadrupole moment, Q_{ij} denotes its symmetric trace-free part, and Q_{ij}^{TT} denotes the transverse-traceless projection of the radiative quadrupole in the wave zone. The quadrupole moment is one of the many *multipole* moments that describe the symmetries, or lack thereof, in the distribution of the source. The above derivation is the famous *quadrupole formula* for gravitational waves that was derived by Einstein et al. in 1918[8]. While this derivation may seem like

an oversimplification, it is a physically insightful map between the nature of the observable quantity and the source. Making no assumptions of the source, the pre-factor, $\frac{2G}{c^4} \approx 10^{-49}\mathrm{s}^2/(\mathrm{g}\cdot\mathrm{cm})$, is incredibly weak. In order to generate gravitational waves, a system must be spatially asymmetrical (i.e., possess at least one non-zero multipole moment) and dynamical (i.e., possess at least one multipole moment that is non-linearly dependent on time). For an order-of-magnitude estimate, the trace-free quadrupole scales as $Q \sim MR^2$, where M is the mass of the system, R is the characteristic size of the system. For a bound system, the characteristic timescale of the system is given by Kepler's laws $T^2 \sim R^3/GM$, giving us a characteristic strain of

$$h \sim \frac{2G}{rc^4}\frac{MR^2}{T^2} \sim \frac{GM}{c^2 r}\frac{2GM}{c^2 R}. \tag{2.1}$$

The characteristic strain is inversely proportional to the ratio of the characteristic size of the system to the *Schwarzschild radius*, $R_s = 2GM/c^2$. This implies that the strength of the characteristic strain is maximized when the characteristic size of the system is on the order of the Schwarzschild radius.

Astrophysically, this severely limits the types of systems that can generate gravitational waves and, among them, those that can be realistically observed. Ideal gravitational wave sources are compact, asymmetrical and highly dynamical. The most common sources of gravitational waves detected by the LIGO-Virgo-KAGRA collaboration have been binary systems of compact objects, such as black holes and neutron stars [1]. The analysis of individual GW signals can be used to test the validity of general relativity, models of the interior structure and evolution of compact binaries. The analysis of the observed population of compact binary mergers provides a significant window into the formation channels of compact binaries as well as the growth of black hole populations through hierarchical mergers. It therefore becomes imperative to develop accurate waveform models that can be used to analyze these signals.

One such family of models are the Effective One Body (EOB) models. The EOB formalism rewrites the post-Newtonian (PN) 2-body problem as the motion of a test particle in a deformed spacetime. This strategy was proposed by Buonanno and Damour [9] to provide a geometric description of the PN dynamics. The strategy also had the advantage of asymptoting to geodesics around a Schwarzschild black hole in the test particle limit. The current generation of EOB-based models have provided a common language for the multiple analytical and numerical methods employed to solve the relativistic black hole binary problem. The resulting waveforms have sufficient accuracy for data analysis for the current generation of gravitational wave detectors. While they are not as

accurate as numerical relativity simulations, and surrogate models derived from them, their intrinsic analyticity provides a significant advantage in terms of extending to parts of the parameter space that are not sufficiently explored by numerical relativity simulations.

2.1 Building EOB waveforms

The EOB formalism, introduced by Buonanno and Damour [9], provides a map from the conservative (absence of GW radiation) 2PN two-body problem to the more geometrical problem of geodesics in some effective spacetime. Since the mapping is purely conservative, it does not account for the energy and angular momentum radiated away by gravitational waves, and is thus only a *linguistic* map, rather than some representation of the spacetime felt by the binary masses or observers at infinity. Here, I will provide a pedagogical example for the 1PN case.

We begin by stating the Hamiltonian for the 1PN 2-body problem in reduced (centre-of-mass) coordinates [9]:

$$\mathcal{H} = \mathcal{H}_0 + \epsilon\mathcal{H}_1 \tag{2.2}$$

$$\mathcal{H}_0 = \frac{p^2}{2} - \frac{1}{q} \tag{2.3}$$

$$\mathcal{H}_1 = \frac{1}{8}(3\nu - 1)p^4 - \frac{1}{2}[(3+\nu)p^2 + \nu(\mathbf{n}\cdot\mathbf{p})^2]\frac{1}{q} + \frac{1}{2q^2} \tag{2.4}$$

Where $\mathbf{q}$ represents the relative separation vector of the binary (with magnitude $r \equiv q$), $\mathbf{p}$ is the corresponding momentum vector, $\epsilon \equiv \frac{1}{c^2}$ is a PN *bookkeeping* or order parameter that we will use to Taylor expand, and $\mathbf{n} = \frac{\mathbf{q}}{q}$ is the unit separation vector. It should also be noted that the non-boldface p and q are the magnitudes of the momentum and separation vectors respectively. We will henceforth use r interchangeably with q to denote the radial separation.

The corresponding Hamilton-Jacobi equation (HJE) then reads

$$\mathcal{R} = E(2 + \epsilon(1-3\nu)E) + \frac{2(1+\epsilon(4-\nu)E)}{r} + \frac{-l^2 + \epsilon(6+\nu)}{r^2} - \frac{\epsilon\nu l^2}{r^3}. \tag{2.5}$$

where $\mathcal{R}$ is the *radial action*, E is the energy, l is the angular momentum, and $\nu = \frac{m_1 m_2}{(m_1+m_2)^2}$ is the symmetric mass ratio. The expression for $\mathcal{R}$ is not unique due to the quadratic nature of the HJE. However, a single valid solution can be obtained by requiring that $\mathcal{R}$ reduces to the Keplerian value in the non-relativistic limit ($\epsilon \to 0$). For actions of the form

$$\mathcal{R} = A + \frac{2B}{r} + \frac{C}{r^2} + \frac{D_1}{r^3}, \tag{2.6}$$

$$A = 2E + (1 - 3\nu)\epsilon E^2, \tag{2.7}$$

$$B = 1 + (4 - \nu)\epsilon E, \tag{2.8}$$

$$C = -l^2 + (6 + \nu)\epsilon, \tag{2.9}$$

$$D_1 = -l^2 \nu \epsilon, \tag{2.10}$$

We can use Goldstein's trick (see Page 470, Eq. 10.138-9 of [10]) to evaluate the corresponding action variable

$$I_r = \frac{1}{2\pi} \oint dr \sqrt{\mathcal{R}(r, E, l)}. \tag{2.11}$$

The limits on this integral are typically the inflection points, or turning points, for orbits corresponding to choices of conserved quantities. Goldstein's trick eliminates the need to compute inflection points by using the language of complex contour integration and the Cauchy residue theorem [11]. Instead of solving the integral twice about the branch of the real line connecting the inflection points, we invert the contour and excise the branch of the real line so as to include complex infinity and the complex plane. Thus, the value of the above integral is the negative of the contour integral with contours about the poles of the integrand which in this case are 0 and ∞. Calculating the residues at the aforementioned poles and Taylor-expanding them by ϵ, we get

$$I_r = -l + \frac{3}{l}\epsilon + \frac{1}{\sqrt{-2E}} \left(1 + \left(\frac{15}{4} - \frac{\nu}{4}\right) \epsilon E\right) \tag{2.12}$$

Inverting the above equation, the binding energy up to 1PN order can be expressed in terms of $J = l$ and $N = J + I_r$ as:

$$E = -\frac{1}{2N^2} \left(1 - \left(\frac{6}{NJ} - \frac{15 - \nu}{N^2}\right) \epsilon\right) \tag{2.13}$$

The above binding energy is non-relativistic, in that it does not contain the rest energy of the system. The complete relativistic version up to 1PN order is

$$E^R_{1PN,real} = \frac{1}{\epsilon\nu} - \frac{1}{2N^2} \left(1 - \left(\frac{6}{NJ} - \frac{15 - \nu}{N^2}\right) \epsilon\right). \tag{2.14}$$

In order to map this to the dynamics of a geodesic in an effective spacetime, the Hamiltonian for a particle evolving in a (non-spinning) metric with up to 1PN relativistic effects is given by:

$$\mathcal{H} = \sqrt{\epsilon\left(1+\epsilon\frac{a_1}{r}+\epsilon^2\frac{a_2}{r^2}\right)\left(\frac{p_r^2}{1+\epsilon\frac{b_1}{r}}+\frac{p_\phi^2}{r^2}+\frac{1}{\epsilon}\right)} \tag{2.15}$$

Here, a_1, a_2, and b_1 are constants that are to be determined by the EOB matching. It is important to note that for the Schwarzschild metric, $a_2 = 0$, $a_1 = -2$ and $b_1 = 2$. As it happens, the action angle integral for the above Hamiltonian can also be solved by Goldstein's trick to obtain

$$I_r = -l + \frac{a_1^2 - a_2 - \frac{a_1 b_1}{2}}{2l}\epsilon + \frac{-\frac{a_1}{2}+\left(b_1-\frac{7a_1}{8}\right)\epsilon E_{\rm NR}}{\sqrt{-2E_{\rm NR}}}. \tag{2.16}$$

Inverting the above equation and adding the rest energy provides the relativistic binding energy in terms of the action angle variables $N = I_r + l$ and $J = l$:

$$E^R_{\rm eff} = \frac{1}{\epsilon} - \frac{a_1^2}{8N^2} + \frac{a_1^2(2a_2+a_1(b_1-2a_1))}{16N^3J}\epsilon + \frac{a_1^3(7a_1-8b_1)}{128N^4}\epsilon \tag{2.17}$$

In order to map the above binding energy to the 1PN result, we introduce the energy map from the effective spacetime to the real relativistic energy:

$$E^R_{\rm real} = \frac{1}{\nu\epsilon}\sqrt{1+2\nu\left(E^R_{\rm eff}\epsilon-1\right)} \tag{2.18}$$

Finally, matching the real and effective relativistic energy up to 1PN order gives the following equations:

$$a_1^2 = 4, \tag{2.19}$$

$$-\frac{a_1^4 + a_1^2 a_2 - \frac{a_1^3 b_1}{2}}{8} = 3, \tag{2.20}$$

$$-\frac{a_1^3(b_1+\frac{\nu-7}{8}a_1)}{2} = 15-\nu, \tag{2.21}$$

which can be solved to give

$$a_2 = 0, \tag{2.22}$$

$$a_1 = \pm 2, \tag{2.23}$$

$$b_1 = \mp 2. \tag{2.24}$$

Once again, a unique solution can be chosen such that it yields the GR solution for the Schwarzschild metric in the test particle limit $\nu \to 0$. Interestingly, at 1PN order, this solution is exactly the 1PN Schwarzschild solution

$$A_{\text{eff}} = 1 - \frac{2\epsilon}{r} \tag{2.25}$$

$$B_{\text{eff}} = 1 + \frac{2\epsilon}{r} \tag{2.26}$$

This completes the EOB formalism for binary dynamics at the 1PN order. While this is a pedagogical introduction with the first PN effect, the strategy above is used consistently to compute the state-of-the-art EOB models we see in the literature.

The second important ingredient to build a waveform model is a corresponding physical description of the gravitational wave strain. The final ingredient is a prescription for the *radiation-reaction* force that modifies the conservative dynamics to account for the energy and angular momentum radiated away by gravitational waves. The strategy developed by Damour and Nagar [12] is called the *factorized-resummed* approach.

Here, the PN-expanded gravitational wave strains are expanded using the multipolar decomposition and re-written as a multiplicative combination [12]

$$h_{lm} = h_{lm}^{(\text{N})} S_{\text{eff}} T_{lm} f_{lm} e^{i\delta_{lm}}, \tag{2.27}$$

where $h_{lm}^{(\text{N})}$ is the multipolar strain for a Newtonian binary, S_{eff} is the effective source factor, T_{lm} is the tail factor, f_{lm} is the residual PN amplitude contribution, and δ_{lm} is the corresponding residual phase contribution. This factorization is motivated by matching the PN expanded strain to the strain from a test particle in a circular Schwarzschild orbit. The source and tail terms are Schwarzschild contributions that correspond to the curved background of the Schwarzschild spacetime. To obtain the correct matching, as in the EOB Hamiltonian, f and δ are written as

polynomials in the PN expansion parameter $x \equiv v/c$, where v is the orbital velocity of a circular orbit, and the coefficients are treated as free parameters. The $S_{\rm eff}$ and T_{lm} terms are also Taylor expanded and the corresponding expansion is matched order by order to the PN expanded gravitational wave strain. The corresponding radiation-reaction flux is incorporated, for circular binaries, in the form of an angular momentum flux

$$\frac{dp_\phi}{dt} = \mathcal{F}_\phi \propto -\sum_{l,m} m\Omega |h_{lm}|^2, \tag{2.28}$$

where Ω is the instantaneous orbital frequency of the EOB trajectory, and the proportionality hides standard geometric normalization constants dependent on the strain convention.

To generate waveforms using the EOB formalism, the test particle is initialized on a circular orbit. The equations of motion are then evolved numerically until the test particle reaches a prescribed end condition such as the innermost circular orbit (ICO) or the innermost stable circular orbit (ISCO) of the EOB Hamiltonian. The trajectory is used to obtain the gravitational wave strain as a function of time using the factorized-resummed strain formula.

The next section provides a more detailed strategy to generating waveforms in the EOB formalism. It uses a more state-of-the-art EOB model, SEOBNRv5 [13], which also includes the effects of the black hole spins and higher order corrections from numerical relativity. Since the post-Newtonian formalism breaks down during the late inspiral, the EOB model must be augmented with a merger-ringdown waveform model. The next section introduces the backwards one body (BOB) formalism for binary merger-ringdowns and details the incorporation of the BOB model into the SEOBNRv5 framework to create the spinning effective-to-backwards one body (SEBOB) waveform model.

2.2 The SEBOB model

All material relevant to this chapter is contained within my publication [14]. The numerical implementation can be found in the open source repository `https://github.com/nrpy/nrpy/` as part of the `seobnrv5_aligned_spin_inspiral` and `sebobv2` examples.

The publication in its entirety, including all figures and tables is provided below. My scientific contribution to this work involved the numerical implementation of the SEOBNRv5 waveform model [13], the derivation of the inspiral to merger matching conditions, the accuracy and systematic analysis of the waveform model against numerical relativity results. The end product of the study is the SEBOB waveform model.

In addition to the scientific formulation of the waveform model, I contributed substantially to the long term maintenance and documentation of the model within the `NRPy` codebase. This included the development of exhaustive code quality tests, robust numerical consistency checks, and continuous integration pipelines.

Combining effective one-body inspirals and backwards one-body merger-ringdowns for aligned spin black hole binaries

Siddharth Mahesh[*,‡] and Sean T. McWilliams[†,‡]

Department of Physics and Astronomy, West Virginia University, Morgantown, West Virginia 26506, USA

Zachariah Etienne

Department of Physics, University of Idaho, Moscow, Idaho 83843, USA

(Received 9 September 2025; accepted 10 November 2025; published 22 December 2025)

High-fidelity gravitational waveform models are essential for realizing the scientific potential of next-generation gravitational-wave observatories. Highly accurate, state-of-the-art models often rely on extensive phenomenological calibrations to numerical relativity (NR) simulations for the late-inspiral and merger phases, which can limit physical insight and extrapolation to regions where NR data is sparse. To address this, we introduce the spinning effective-to-backwards one-body (`SEBOB`) formalism, a hybrid approach that combines the well-established effective-one-body (EOB) framework with the analytically driven backwards-one-body (BOB) model, which describes the merger-ringdown from first principles as a perturbation of the final remnant black hole. We present two variants building on the state-of-the-art `SEOBNRv5HM` model: `seobnrv5_nrnqc_bob`, which retains standard NR-calibrated nonquasicircular (NQC) corrections and attaches a BOB-based merger-ringdown, and a more ambitious variant, `seobnrv5_bob`, which uses BOB to also inform the NQC corrections, thereby reducing reliance on NR fitting and enabling higher-order ($\mathcal{C}^2$) continuity by construction. Implemented in the open-source NRPy framework for optimized C-code generation, the `SEBOB` model is transparent, extensible, and computationally efficient. By comparing our waveforms to a large catalog of NR simulations, we demonstrate that `SEBOB` yields accuracies comparable to the highly calibrated `SEOBNRv5HM` model, providing a viable pathway toward more physically motivated and robust waveform models for precision gravitational-wave astronomy.

DOI: 10.1103/67c8-1rv3

I. INTRODUCTION

The last decade of gravitational-wave (GW) astronomy, driven by observations from the LIGO and Virgo detectors, has produced a catalog of $\mathcal{O}(100)$ compact binary coalescences (CBCs) [1]. This has inaugurated an era of gravitational-wave astrophysics, yielding unprecedented insights into the populations and properties of stellar remnants, enabling novel tests of general relativity [2], new measurements of the cosmic expansion rate [3], and a deeper understanding of compact object populations [4]. Looking ahead, next-generation observatories such as the space-based Laser Interferometer Space Antenna [5], and the planned third-generation ground-based detectors Einstein Telescope [6] and Cosmic Explorer [7], promise order-of-magnitude improvements in sensitivity. These instruments will regularly produce signals with signal-to-noise ratios (SNRs) in the hundreds or thousands, demanding GW waveform models of correspondingly higher fidelity. Indeed, even at current detector sensitivities, systematic errors in waveform modeling can appreciably bias inferred source parameters [8], and reducing such modeling errors will be absolutely critical for future high-SNR science.

The increasing sensitivity and anticipated event rates necessitate highly accurate and computationally efficient gravitational waveform models. These models are the cornerstone of GW astronomy, essential for detecting signals, accurately estimating source parameters, and performing robust astrophysical inference [9]. Current data analysis pipelines, particularly those based on Bayesian inference, often require millions of waveform evaluations for a single event, demanding models that are both precise and fast.

Several approaches model CBC waveforms. Numerical relativity (NR) simulations provide the most accurate solutions to Einstein's equations for the strong-field merger phase [10], serving as crucial benchmarks. Post-Newtonian (PN) theory offers analytical approximations for the early inspiral [11] but breaks down near merger. Phenomenological models [12] and surrogate models [13]

[*]Contact author: sm0193@mix.wvu.edu

[†]Contact author: Sean.McWilliams@mail.wvu.edu

[‡]Also at Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, West Virginia 26505, USA.

TABLE I. Summary of NR inputs for the (2,2) mode of the SEOBNRv5 waveform model. The first three parameters are calibration parameters, tuned via Bayesian inference against NR waveforms. The remaining parameters are derived from NR fits and are specified, prior to calibration, for NQC corrections or the merger-ringdown model. Nominal values for these parameters are found in [15], except for M_f and a_f, which are given in the listed citations in the Method column. The abbreviation NBI stands for nested Bayesian inference.

Parameter	Description	Precalibration	Method	In seobnrv5_bob
a_6	Effective 5 PN nonspinning inspiral coefficient		NBI	✓
$d_{\rm SO}$	Effective 4.5 PN inspiral spin-orbit coefficient		NBI	✓
$\Delta t^{\rm ISCO}$	Time shift between EOB ISCO passage and (2,2) peak strain		NBI	✓
$\lvert h_{22}^{\rm NR}\rvert^{\rm peak}$	(2,2) mode peak strain amplitude	✓	NR fit	✓
$\lvert\ddot{h}\rvert_{22}^{\rm peak}$	(2,2) mode strain amplitude second derivative at peak strain time	✓	NR fit	
$\omega_{22}^{\rm peak}$	(2,2) mode frequency at peak strain time	✓	NR fit	✓
$\dot{\omega}_{22}^{\rm peak}$	(2,2) mode frequency derivative at peak strain time	✓	NR fit	
$M_{\rm f}$	Remnant mass	✓	NR fit [17]	✓
$a_{\rm f}$	Remnant spin	✓	NR fit [18]	✓
$c_{1,f}$	Merger-ringdown amplitude coefficient	✓	NR fit	⋯
$c_{2,f}$	Merger-ringdown amplitude coefficient	✓	NR fit	⋯
$d_{1,f}$	Merger-ringdown phase coefficient	✓	NR fit	⋯
$d_{2,f}$	Merger-ringdown phase coefficient	✓	NR fit	⋯

combine analytical insights with fits to NR data. The effective-one-body (EOB) formalism [14] maps the two-body problem to an effective particle in a deformed metric, incorporating PN information, resummation techniques, and NR calibration. State-of-the-art EOB models like SEOBNRv5HM [15] (which includes higher-order modes and gravitational self-force information) and TEOBResumS [16] are widely used. While the ultimate goal includes precessing spins (e.g., via SEOBNRv5PHM development), accurate aligned-spin models are a crucial foundation.

This paper focuses on advancing aligned-spin EOB models. While SEOBNRv5HM [15] is highly accurate, its merger-ringdown and nonquasicircular (NQC) corrections often rely on phenomenological fits to NR data (see Table I). This empirical tuning can obscure underlying physics and limit extrapolation, especially to regions like extreme mass ratios or nearly extremal spins where NR calibration is sparse.

To address this, we combine EOB with the backwards-one-body (BOB) formalism [19]. BOB describes the late inspiral, merger, and ringdown by working backwards from the final remnant black hole, effectively treating the late-stage binary as a linear perturbation of the stationary remnant. It leverages the idea that outgoing GWs are driven by perturbations near the remnant's light ring [20], with an underlying orbital frequency smoothly transitioning to the remnant's quasinormal mode (QNM) frequencies [21–23]. Notably, the original BOB model reproduced NR merger-ringdown waveforms to within NR uncertainty *without* phenomenological tuning parameters, suggesting an accurate analytic description of merger physics from first principles.

Our primary contribution is the spinning effective-to-backwards one body (SEBOB) model. We present two distinct strategies for this integration. The first, termed seobnrv5_nrnqc_bob, uses the standard NR-calibrated NQC corrections from SEOBNRv5HM for the inspiral-plunge phase, then smoothly attaches a BOB model to analytically model the plunge, merger, and ringdown. The second, more ambitious strategy, seobnrv5_bob utilizes BOB not only for the merger-ringdown but also to inform the NQC corrections themselves, thereby reducing reliance on direct NR fits and facilitating, by construction, higher-order continuity (e.g., $\mathcal{C}^2$) at the transition point. SEBOB aims to be an analytically driven model for the most nonlinear stage, enhancing robustness where NR calibrations are unreliable. In comprehensive comparisons to Simulating eXtreme Spacetimes (SXS) NR waveforms [focusing on the dominant $(\ell, m) = (2, 2)$ mode], both SEBOB variants achieve median noise-free mismatches of $\mathcal{M} \simeq 2 \times 10^{-4}$, competitive with SEOBNRv5 and TEOBResumS, even as the BOB-informed variant reduces direct NR calibration (see Sec. V).

We implement SEBOB within NRPy [24], an open-source Python package for symbolic manipulation and optimized C-code generation. NRPy translates EOB Hamiltonians, fluxes, and waveform quantities from a human-readable symbolic form into highly efficient, automatically generated C code, using advanced common subexpression elimination (CSE) on Hamiltonian, flux, and strain expressions. This offers the transparency and flexibility advantages of developing in Python (making it an "experimental playground"), while retaining the performance benefits of C. In wall-time benchmarks (Sec. V), our NRPy-based C implementation yields ∼3× speedups over pySEOBNR [25] for equal and moderate mass ratios, while remaining competitive with other state-of-the-art generators, including LALSuite's C-based SEOBNRv5_ROMs [26] (when taking into account the ROM is already in frequency domain). Our analysis

also identifies the main sources of residual error near the peak amplitude, pointing to improvements in the treatment of amplitude curvature that we defer to future work.

In summary, SEBOB addresses several needs for next-generation gravitational-wave modeling. It aims for improved accuracy in the late merger-ringdown through analytic insight, which reduces dependence on empirical calibration to NR. Furthermore, it is designed for computational efficiency and scalability suitable for the data rates and parameter-space demands of future observatories. Finally, it promotes enhanced reproducibility and transparency through its open-source implementation. The remainder of this paper is organized as follows: Sec. II details the SEOBNRv5HM formalism. Section III introduces BOB and our SEBOB integration prescriptions. Section IV discusses the NRPy implementation. Section V presents accuracy and performance comparisons. Section VI concludes with a discussion of the results and future directions.

Throughout this paper, we employ geometrized units ($c = G = M = 1$). Time derivatives are denoted by an overdot (e.g., $\dot{X} \equiv dX/dt$). The symmetric mass ratio of a binary is given by $\eta = m_1 m_2/(m_1 + m_2)^2$, where m_1 and m_2 are the masses of the component black holes. The dimensionless spin angular momenta of the individual black holes are represented by $\chi_i \equiv \vec{S}_i/m_i^2$. Waveform angular frequency i.e., derivative of the waveform phase, is denoted by ω and orbital frequency i.e, derivative of the azimuthal angle, is denoted by Ω.

II. SPINNING EOB WAVEFORMS

The original EOB formalism [14] canonically mapped the nonspinning, second-order PN two-body problem to the geodesic motion of an effective test particle in a modified Schwarzschild spacetime. Subsequent advances, including higher PN orders, spins, resummation techniques, and calibration against NR data, have yielded state-of-the-art models such as SEOBNRv5HM [15] and TEOBResumS [27]. While these models share foundational EOB principles, this work focuses specifically on SEOBNRv5HM. This section details the analytical foundations of SEOBNRv5HM, discusses the incorporation of NR data into the formalism, and outlines the NR parameters essential for constructing inspiral-merger-ringdown (IMR) waveforms.

Within the EOB formalism, the conservative Hamiltonian $\mathcal{H}_{\rm real}$ of the PN dynamics is canonically mapped to an effective Hamiltonian $\mathcal{H}_{\rm eff}$ via the relation,

$$\mathcal{H}_{\rm real} = \frac{1}{\eta}\sqrt{1 + 2\eta(\mathcal{H}_{\rm eff} - 1)}. \tag{1}$$

The effective Hamiltonian $\mathcal{H}_{\rm eff}$ describes the geodesic motion of an effective test particle in a modified Kerr metric $g_{\mu\nu}(x^j)$, with canonical coordinates and momenta denoted by $\{x^j, p_j\}$. It is derived from the relativistic mass-shell constraint, $g^{\mu\nu}p_\mu p_\nu = -1$, supplemented by a potential $\mathcal{Q}(x^j, p_j)$ that accounts for quartic $[\mathcal{O}(p_j^4)]$ and higher-order corrections to geodesic motion. The effective Hamiltonian is given by,

$$g^{00}\mathcal{H}_{\rm eff}^2 + g^{0i}\mathcal{H}_{\rm eff}p_i + g^{ij}p_i p_j + \mathcal{Q}(x^j, p_j) = -1. \tag{2}$$

Expressed in spherical polar coordinates (r, ϕ) with associated momenta (p_r, p_ϕ), the effective Hamiltonian also takes the form,

$$\mathcal{H}_{\rm eff} = H_{\rm odd} + H_{\rm even}, \tag{3}$$

where $H_{\rm odd}$ encodes spin-orbit interactions, $H_{\rm even}$ describes orbital motion, and p_ϕ represents the system's canonical angular momentum. For a detailed expansion of the Hamiltonian, readers are referred to Sec. II and Appendix A of Ref. [15]. In the nonspinning limit ($a_i \to 0$, denoted by the subscript noS), the orbital component $H_{\rm even}$ simplifies to,

$$H_{\rm even}|_{a_i\to 0} = \sqrt{A_{\rm noS}\left(1 + A_{\rm noS}D_{\rm noS}p_{r_*}^2 + \frac{p_\phi^2}{r^2} + Q_{\rm noS}\right)}. \tag{4}$$

The functions $A_{\rm noS}$, $D_{\rm noS}$, and $Q_{\rm noS}$ incorporate the canonically mapped PN information in this limit. To improve computational efficiency and numerical stability closer to the effective horizon, the coordinates (r, p_r) are transformed to tortoise coordinates (r_*, p_{r_*}) via,

$$\xi = A_{\rm noS}\sqrt{D_{\rm noS}}, \qquad p_{r_*} = \xi p_r, \qquad \frac{dr_*}{dr} = \frac{1}{\xi}. \tag{5}$$

Inspiral gravitational waveform modes h_{lm} are constructed using a factorized representation, $h_{lm}^{\rm F}$, derived from PN waveform modes. These modes depend on the binary parameters (η, χ_1, χ_2), the canonical phase space variables $(\vec{r}, \vec{p})$, the EOB Hamiltonian $\mathcal{H}_{\rm real}$, the instantaneous orbital frequency $\Omega = d\phi/dt$, and $\Omega_{\rm circ}$, which is the corresponding frequency for a circular orbit ($p_{r_*} \to 0$) at radius r. Gravitational radiation reaction is incorporated by resumming factorized modes $h_{lm}^{\rm F}$ into a dissipative flux term that modify Hamilton's equations. The angular momentum flux $\mathcal{F}_\phi$ and the corresponding radial flux $\mathcal{F}_r$ are given by,

$$\mathcal{F}_\phi = -\frac{\Omega}{8\pi}\sum_{l=2}^{8}\sum_{m=1}^{l} m^2|h_{lm}^{\rm F}|^2, \tag{6}$$

$$\mathcal{F}_r = \frac{p_{r_*}}{p_\phi}\mathcal{F}_\phi. \tag{7}$$

The canonical coordinates and momenta are evolved numerically using Hamilton's equations of motion, with the fluxes $\mathcal{F}_\phi$ and $\mathcal{F}_r$ included as external force terms,

$$\frac{dr}{dt} = \xi \frac{\partial \mathcal{H}_{\rm real}}{\partial p_{r_*}}, \tag{8}$$

$$\frac{d\phi}{dt} = \frac{\partial \mathcal{H}_{\rm real}}{\partial p_\phi} \equiv \Omega, \tag{9}$$

$$\frac{dp_{r_*}}{dt} = -\xi \frac{\partial \mathcal{H}_{\rm real}}{\partial r} + \mathcal{F}_r, \tag{10}$$

$$\frac{dp_\phi}{dt} = \mathcal{F}_\phi. \tag{11}$$

The equations of motion are typically integrated until the orbital frequency Ω or the time derivative of the radial momentum dp_{r_*}/dt reaches an extremum. At each integration time step, auxiliary quantities ($\mathcal{H}_{\rm real}$, Ω, and $\Omega_{\rm circ}$) that are necessary for computing the factorized waveform modes $h^{\rm F}_{lm}$ are evaluated alongside the canonical variables.

A. Numerical relativity in the EOB formalism

The EOB formalism, as presented thus far, relies primarily on analytical PN approximations of general relativity. To enhance the accuracy of inspiral waveforms for gravitational-wave signal parameter estimation, corrections from NR are essential. This subsection details the three primary NR inputs used to construct the full IMR waveform: NQC corrections, the merger-ringdown (MR) waveform model, and inspiral calibration parameters.

First, NQC corrections are applied to improve waveform accuracy in the *plunge regime*, where the quasicircular orbit approximation used for $h^{\rm F}_{lm}$ becomes invalid. After obtaining the EOB trajectory and factorized waveform, the NQC-corrected inspiral-plunge waveform $h^{\rm insp\text{-}plunge}_{22}$ is computed as,

$$h^{\rm insp\text{-}plunge}_{22} = h^{\rm F}_{22}\left[1 + \frac{p^2_{r_*}}{r^2\Omega^2}\left(a_1 + \frac{a_2}{r} + \frac{a_3}{r^{3/2}}\right)\right] e^{i\left(b_1 \frac{p_{r_*}}{r\Omega} + b_2 \frac{p^3_{r_*}}{r\Omega}\right)}. \tag{12}$$

Coefficients $\{a_1, a_2, a_3\}$ and $\{b_1, b_2\}$ are determined by matching $h^{\rm insp\text{-}plunge}_{22}$'s amplitude and frequency (defined as the phase derivative) to corresponding NR values at a specific time, $t^{\rm peak}_{22}$. This $t^{\rm peak}_{22}$ corresponds to the peak amplitude of the NR (2,2) mode strain, related to EOB dynamics via the tunable parameter $\Delta t^{\rm ISCO}$:

$$t^{\rm peak}_{22} = t^{\rm ISCO} - \Delta t^{\rm ISCO}. \tag{13}$$

Here, $t^{\rm ISCO}$ denotes the time the EOB trajectory crosses the innermost stable circular orbit (ISCO) radius of the remnant black hole. The remnant's mass M_f and dimensionless spin a_f, essential for ISCO radius calculation, are obtained from NR fitting formulas [17,18]. To determine the amplitude coefficients $\{a_1, a_2, a_3\}$, the factorized EOB waveform and trajectory are interpolated near $t^{\rm peak}_{22}$. These coefficients are then found by enforcing agreement between the NQC-corrected EOB amplitude $|h^{\rm insp\text{-}plunge}_{22}|$ (and its first two time derivatives) and the corresponding NR amplitude $|h^{\rm NR}_{22}|$ (and its first two time derivatives, with the first derivative being zero at peak amplitude), all evaluated at $t^{\rm peak}_{22}$,

$$\begin{pmatrix} \mathcal{Q}_1 & \mathcal{Q}_2 & \mathcal{Q}_3 \\ \dot{\mathcal{Q}}_1 & \dot{\mathcal{Q}}_2 & \dot{\mathcal{Q}}_3 \\ \ddot{\mathcal{Q}}_1 & \ddot{\mathcal{Q}}_2 & \ddot{\mathcal{Q}}_3 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} |h^{\rm NR}_{22}| - |h^{\rm F}_{22}| \\ \frac{d}{dt}|h^{\rm NR}_{22}| - \frac{d}{dt}|h^{\rm F}_{22}| \\ \frac{d^2}{dt^2}|h^{\rm NR}_{22}| - \frac{d^2}{dt^2}|h^{\rm F}_{22}| \end{pmatrix}_{t^{\rm peak}_{22}}, \tag{14}$$

$$\mathcal{Q} = \left[|h^{\rm F}_{22}|\left\{\frac{p^2_{r_*}}{r^2\Omega^2}, \frac{p^2_{r_*}}{r^3\Omega^2}, \frac{p^2_{r_*}}{r^{7/2}\Omega^2}\right\}\right]_{t=t^{\rm peak}_{22}}. \tag{15}$$

Similarly, phase correction coefficients $\{b_1, b_2\}$ are determined by matching the NQC-corrected frequency $\omega_{22} = d\phi_{22}/dt$ and its time derivative $\dot{\omega}_{22}$ of the NQC-corrected EOB waveform to corresponding NR values at $t^{\rm peak}_{22}$,

$$\begin{pmatrix} \mathcal{P}_1 & \mathcal{P}_2 \\ \dot{\mathcal{P}}_1 & \dot{\mathcal{P}}_2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \omega^{\rm NR}_{22} - \omega^{\rm F}_{22} \\ \dot{\omega}^{\rm NR}_{22} - \dot{\omega}^{\rm F}_{22} \end{pmatrix}_{t^{\rm peak}_{22}}, \tag{16}$$

$$\mathcal{P} = \left[\left\{\frac{p_{r_*}}{r\Omega}, \frac{p^3_{r_*}}{r\Omega}\right\}\right]_{t=t^{\rm peak}_{22}}. \tag{17}$$

Given the trajectory and the factorized waveform, the quantities $\mathcal{Q}$ and $\mathcal{P}$ can be interpolated, spline-differentiated, and evaluated at $t^{\rm peak}_{22}$. Equations (14) and (16) are then solved numerically to determine the NQC coefficients $\{a_1, a_2, a_3\}$ and $\{b_1, b_2\}$, respectively. Second, the EOB formalism incorporates a separate MR model to specify the full gravitational strain of a black hole binary merger. In `SEOBNRv5`, MR waveform modes are modeled as the emission of quasinormal modes (QNMs) [21] from the remnant black hole, with amplitudes and phases modulated by phenomenological functions fitted to NR data,

$$h^{\rm merger\text{-}ringdown}_{22}(t) = \tilde{A}_{22}(t) e^{i\tilde{\phi}_{22}(t)} e^{-i\sigma^{\rm QNM}_{22}(t - t^{\rm peak}_{22})}, \tag{18}$$

$$\tilde{A}_{22}(t) = \eta c_{1,c} \tanh\left[c_{1,f}(t - t^{\rm peak}_{22}) + c_{2,f}\right] + c_{2,c}, \tag{19}$$

$$\tilde{\phi}_{22} = \phi_0 - d_{1,c} \log\left[\frac{1 + d_{2,f} e^{-d_{1,f}(t - t^{\rm peak}_{22})}}{1 + d_{2,f}}\right], \tag{20}$$

$$\sigma^{\rm QNM}_{22} = -\frac{i}{\tau^{\rm QNM}_{22}} + \omega^{\rm QNM}_{22}. \tag{21}$$

Here, $\omega_{22}^{\rm QNM}$ and $\tau_{22}^{\rm QNM}$ are the frequency and damping time of the dominant (2,2) QNM. These are calculated as functions of the remnant mass M_f and spin a_f (obtained from NR fits) using tools such as the QNM package [28]. Coefficients $\{c_{1,f}, c_{2,f}, d_{1,f}, d_{2,f}\}$ are determined by fitting to a catalog of NR waveforms. The remaining coefficients $\{c_{1,c}, c_{2,c}, d_{1,c}, \phi_0\}$ are fixed by requiring C^1 continuity (continuous value and first derivative) when attaching the MR waveform to the NQC-corrected inspiral-plunge waveform at $t_{22}^{\rm peak}$.

Third, after NQC corrections and the MR model structure are incorporated, the overall accuracy of the resulting IMR waveform is further enhanced by tuning three parameters $\{a_6, d_{\rm SO}, \Delta t^{\rm ISCO}\}$ within the EOB dynamics against a large set of NR waveforms. The parameter a_6 is a pseudo-PN coefficient entering the nonspinning inspiral potential $A_{\rm noS}$ at 5PN order,

$$\begin{aligned} A_{\rm noS} &= 1 - 2u + 2\eta u^3 + \eta\left(\frac{94}{3} - \frac{41\pi^2}{32}\right)u^4 \\ &\quad + \left[\eta\left(\frac{2275\pi^2}{512} - \frac{4237}{60} + \frac{128\gamma_{\rm E}}{5} + \frac{256\ \ln\ 2}{5}\right)\right. \\ &\quad \left. + \eta^2\left(\frac{41\pi^2}{32} - \frac{221}{6}\right) + \frac{64}{5}\eta\ \ln\ u\right]u^5 \\ &\quad + \left[\eta a_6 + \left(-\frac{144\eta^2}{5} - \frac{7004\eta}{105}\right)\ln\ u\right]u^6, \end{aligned} \tag{22}$$

where $u \equiv 1/r$ is the inverse of the radial separation. The parameter $d_{\rm SO}$ is a pseudo-PN parameter that enters the spin-orbit Hamiltonian at the 4.5 PN order,

$$H_{\rm odd} = \frac{p_\phi(g_{a_+}a_+ + g_{a_-}a_-) + \eta d_{\rm SO}u^3 p_\phi a_+ + G_{a^3}^{\rm align}}{r^3 + a_+^2(r+2)}, \tag{23}$$

where the spin combinations $a_\pm = \chi_1 m_1 \pm \chi_2 m_2$, $g_{a_\pm}$ are the corresponding gyrogravitomagnetic factors, and $G_{a^3}^{\rm align}$ contains cubic-in-spin couplings [see Eqs. (26)–(30) of [29]]. The third $\Delta t^{\rm ISCO}$ parameter, defined earlier, encodes the time between ISCO crossing and the peak of the NR waveform. These three parameters are calibrated via Nested Bayesian inference, minimizing mismatch against a set of 446 SXS [10] (and one Einstein Toolkit) waveforms. The inference proceeds in stages: first, nonspinning waveforms constrain a_6 and the spin-independent components of $\Delta t^{\rm ISCO}$; subsequently, $d_{\rm SO}$ and the spin-dependent components of $\Delta t^{\rm ISCO}$ are constrained. Crucially, this calibration is performed on NQC-corrected, merger-attached waveforms. Consequently, NR-derived parameters for NQC corrections ($|h_{22}^{\rm NR}|^{\rm peak}$, $|\ddot{h}|_{22}^{\rm peak}$, $\omega_{22}^{\rm peak}$, $\dot{\omega}_{22}^{\rm peak}$) and MR model fits ($c_{i,f}, d_{i,f}, M_f, a_f$) are considered "precalibration" inputs, fixed prior to the final tuning of $\{a_6, d_{\rm SO}, \Delta t^{\rm ISCO}\}$. The complete set of NR-derived parameters for the SEOBNRv5 model is summarized in Table I.

The SEOBNRv5 waveform model, which combines novel PN approximation approaches with advances in accurate NR simulations, provides highly accurate templates for analyzing gravitational wave signals from black hole binary mergers observed by LIGO, Virgo, and KAGRA. The model [15] is implemented in Cython within the open-source package pySEOBNR [25]. For faster waveform evaluation, a reduced-order surrogate model, SEOBNRv5_ROM, is also implemented in C as part of LALSuite's [26].

III. BACKWARDS ONE-BODY WAVEFORMS

The BOB formalism [19] models the late inspiral and merger-ringdown stages of black hole binary coalescence. It describes these phases as dynamical perturbations propagating outward through null congruences near the remnant's light ring, superposed on the stationary spacetime of the merged remnant black hole. This section first reviews BOB's mathematical formulation, then presents and justifies two distinct prescriptions for matching EOB inspiral waveforms. Throughout this section, $\omega(t)$ and related quantities like ω_0 and $\omega_{\rm QNM}$ refer to the gravitational waveform frequency for a given (l, m) mode.

The BOB formalism models gravitational wave perturbation amplitudes using the *Newman-Penrose scalar* [20] ψ_4, which represents the second derivative of the gravitational wave strain, $\ddot{h}$. Based on the divergence of null congruences near the light ring, the amplitude of each (l, m) mode is given by [cf. Eq. (5) in Ref. [19]]:

$$A_{lm}(t) \equiv |\psi_{4,lm}| = A_{p,lm}\,\mathrm{sech}\left(\frac{t - t_{p,lm}}{\tau_{lm}}\right), \tag{24}$$

where $(t_{p,lm}, A_{p,lm})$ correspond to the time and value of the peak $\psi_{4,lm}$ amplitude for that mode, and τ_{lm} is the damping time of the (l, m) QNM [21] of the remnant black hole.

The waveform phase derivation assumes a single underlying orbital frequency $\Omega(t)$, which governs all multipole frequencies via $\omega_{lm} = m\Omega$. This frequency evolution is inferred from the identity [30,31] relating the news function, $\mathcal{N}_{lm} \equiv \dot{h}_{lm}$, to the frequency evolution,

$$|\mathcal{N}_{lm}|^2 \propto m^2 \frac{d\Omega^2}{dt}. \tag{25}$$

Combined with the adiabatic relation $|\psi_{4,lm}|^2 \approx m^2\Omega^2|\mathcal{N}_{lm}|^2$ (more precisely, $|\psi_{4,lm}|^2 \propto m^4\Omega^3\frac{d\Omega}{dt}$, which leads to the form below) that assumes a negligible contribution from the news amplitude derivative to ψ_4, we obtain an expression for $\omega_{lm} = m\Omega$. Dropping mode labels, the frequency $\omega(t)$ is given by,

$$\omega(t) = \left\{\omega_0^4 + k\left(\tanh\left(\frac{t-t_p}{\tau}\right) - \tanh\left(\frac{t_0-t_p}{\tau}\right)\right)\right\}^{1/4}, \tag{26}$$

with k defined as,

$$k = \frac{\omega_{\rm QNM}^4 - \omega_0^4}{1 - \tanh\left(\frac{t_0-t_p}{\tau}\right)}, \tag{27}$$

where t_0 and ω_0 are the time and waveform frequency at a reference point, and $\omega_{\rm QNM}$ is the fundamental QNM frequency of the remnant black hole. Integrating $\omega(t)$ then yields the waveform phase $\phi(t)$,

$$\begin{aligned}\phi(t) = \phi_0 &+ \omega_{+\infty}\tau(\arctan_+ + \operatorname{arctanh}_+) \\ &- \omega_{-\infty}\tau(\arctan_- + \operatorname{arctanh}_-),\end{aligned} \tag{28}$$

where ϕ_0 is the phase at the reference time t_0, $\omega_{\pm\infty}$ are the frequencies evaluated in the limit $t \to \pm\infty$, and the terms $\arctan_\pm/\operatorname{arctanh}_\pm$ are defined as,

$$\arctan_\pm = \left(\tan^{-1}\left(\frac{\omega}{\omega_{\pm\infty}}\right) - \tan^{-1}\left(\frac{\omega_0}{\omega_{\pm\infty}}\right)\right), \tag{29}$$

$$\operatorname{arctanh}_\pm = \left(\tanh^{-1}\left(\frac{\omega}{\omega_{\pm\infty}}\right) - \tanh^{-1}\left(\frac{\omega_0}{\omega_{\pm\infty}}\right)\right), \tag{30}$$

where $\tan^{-1}$ and $\tanh^{-1}$ are the inverse tangent and inverse hyperbolic tangent, respectively.

To relate ψ_4 to the gravitational wave strain h, we assume the phase evolves linearly at leading order near merger, yielding $\psi_4 = \ddot{h} \approx -\omega(t)^2 h$. Under this approximation, the complex strain is given by,

$$h(t) = \frac{|h_0|\omega_0^2}{\omega(t)^2}\frac{\operatorname{sech}\left(\frac{t-t_p}{\tau}\right)}{\operatorname{sech}\left(\frac{t_0-t_p}{\tau}\right)}\exp\left(i\phi(t)\right), \tag{31}$$

where $|h_0|$ is the strain amplitude at the reference time t_0.

A condition relating t_p to t_0 is required. For matching with `SEOBNRv5`, t_0 is chosen as the time of peak strain amplitude, $t_{22}^{\rm peak}$, the natural end point of the EOB inspiral. Requiring $d|h|/dt = 0$ at $t = t_0$ leads to the following expression:

$$t_p = t_0 - 2\tau\ln\left(\frac{\omega_0}{\omega_{\rm QNM}}\right), \tag{32}$$

where the identity

$$\tanh^{-1}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) \tag{33}$$

is used to invert the hyperbolic tangent function. Equation (32) differs from the t_p condition derived in Eq. (9) of Ref. [19], which imposes $\mathcal{C}^2$ phase continuity and depends on $\dot{\omega}_0$ for an arbitrary matching time t_0. However, within the EOB framework, NQC corrections ensure a peak in the strain at the matching time. This choice for t_0 therefore ensures the matching point corresponds to the peak strain amplitude in both the EOB and BOB models. This approach, implemented in the `seobnrv5_nrnqc_bob` model, utilizes BOB to model the merger-ringdown, attached to the NQC-corrected EOB inspiral.

Alternatively, BOB can directly provide NQC corrections. The conventional `SEOBNRv5` NQC corrections [cf. Eq. (12)] rely on NR fits to the peak amplitude, frequency, and their derivatives at the peak strain time. With the BOB approach, these derivatives are instead computed analytically from the BOB model [Eqs. (26) and (31)] evaluated at the peak strain time. The functional form of the NQC-corrected inspiral-plunge waveform from Eq. (12) is maintained, but the coefficients $\{a_1, a_2, a_3, b_1, b_2\}$ are instead determined by requiring agreement with the BOB-predicted values for the strain amplitude $|h_{\rm BOB}|$, its first time derivative [which is zero at t_0 by construction using Eq. (32)], its second time derivative $d^2|h_{\rm BOB}|/dt^2$, as well as the frequency $\omega_{\rm BOB}$ and its first time derivative $d\omega_{\rm BOB}/dt|$, all evaluated at t_0. This model is implemented in NRPy as `seobnrv5_bob`.

The approach of using BOB to inform NQCs is motivated by the observation that the peak strain occurs close to the perturber crossing the remnant ISCO, a regime particularly apt for BOB's description. It reduces reliance on direct NR fitting for NQC parameters by leveraging BOB's analytical structure. Primary NR inputs are thus limited to the remnant properties (mass and spin, which determine $\omega_{\rm QNM}$ and τ) and the strain amplitude $|h_0|$ and frequency ω_0 at the matching time t_0. Furthermore, while the `SEOBNRv5` merger-ringdown attachment guarantees only $\mathcal{C}^1$ continuity (continuous complex h and $\dot{h}$) at the matching point t_0, employing BOB for NQC computation inherently links inspiral corrections to the merger-ringdown model, thus enabling $\mathcal{C}^2$ continuity (continuous $h, \dot{h}, \ddot{h}$) by construction.

Conversely, NQC corrections are applied to the entire inspiral waveform via their dependence on EOB dynamical variables (p_{r_*}, r, Ω). For generic noncircular orbits, the physically correct dependence may involve PN waveform predictions. In such cases, BOB may not provide a complete description of the inspiral's NQC corrections. Therefore, employing BOB solely as a merger-ringdown attachment (`seobnrv5_nrnqc_bob`) retains the NR-calibrated NQCs for the inspiral and could be advantageous.

Equation (32) implies that the merger-ringdown waveform is determined solely by the strain amplitude $|h_0|$ and frequency ω_0 at the matching time t_0, along with the remnant mass and spin used to calculate $\omega_{\rm QNM}$ and τ. Thus, BOB waveforms eliminate all NR-informed degrees of freedom in modeling the merger-ringdown evolution, provided that the remnant mass and spin themselves are predicted from inspiral parameters without direct NR fitting

(e.g., using methods like those in Ref. [32]). This can also aid in calibrating the EOB Hamiltonian to accurately predict remnant mass and spin directly from the inspiral dynamics (see Ref. [32]). Although the BOB formalism can model any gravitational wave mode (l, m), our analysis in this paper is restricted to the dominant (2,2) mode. The combination of BOB's analytical model for the merger-ringdown with the EOB inspiral's analytical foundations into an EBOB framework aims to describe the complete black hole binary coalescence waveform analytically and from physical motivations. In Sec. V, we will assess the fidelity of BOB predictions against numerical relativity data and examine the impact of using BOB-derived NQCs on overall waveform accuracy.

IV. NUMERICAL IMPLEMENTATION

The `SEBOB` waveform model, comprising the `seobnrv5_nrnqc_bob` and `seobnrv5_bob` variants, is implemented in the open-source NRPy framework [24,33]. NRPy, built upon the SymPy library [34] for symbolic manipulation, translates all required EOB and BOB mathematical expressions into highly optimized, modular C code. This symbolic-to-numerical workflow, which extensively applies CSE to reduce runtime costs, facilitates transparency, extensibility, and high performance.

The generated C code is self-contained and can be compiled as an executable or linked as a library. It relies on the GNU scientific library (GSL) [35] for core numerical tasks, including root-finding, ordinary differential equation (ODE) integration, interpolation, and linear algebra. This structure enables seamless integration into gravitational-wave data analysis packages such as LALSuite [26] by avoiding Python-to-C overhead, and supports deployment within infrastructures such as BlackHoles@Home [33] for large-scale model calibration and waveform generation.

A. EOB dynamics: Initial conditions and evolution

1. Initial conditions

EOB evolution begins by determining initial conditions for the canonical phase-space variables $\{r, \phi, p_{r_*}, p_\phi\}$. First, for a user-specified initial orbital frequency Ω_0, the initial orbital separation r_0 and azimuthal momentum $p_{\phi,0}$ for quasicircular orbits are obtained by solving the two-dimensional nonlinear algebraic system,

$$\frac{dp_r}{dt} \equiv \frac{\partial \mathcal{H}_{\rm real}(r, p_{r_*} = 0, p_\phi)}{\partial r} = 0, \tag{34}$$

$$\frac{d\phi}{dt} \equiv \frac{\partial \mathcal{H}_{\rm real}(r, p_{r_*} = 0, p_\phi)}{\partial p_\phi} = \Omega_0. \tag{35}$$

This system is solved using GSL's hybrid Powell multidimensional root-finding algorithm. The first condition is equivalent to no radial motion since $p_{r_*} = 0$ guarantees $dr/dt = 0$ since the Hamiltonian is quadratic in p_{r_*}.

Second, the initial radial momentum $p_{r_{*,0}}$, typically small but nonzero for dissipative inspirals, is determined by solving a one-dimensional root-finding problem using GSL's Brent-Dekker solver. The radial momentum is derived by equating the initial conservative radial velocity $\dot{r}_{\rm dyn} = \xi \partial \mathcal{H}_{\rm real}/\partial p_{r_*}$ to the radial velocity due to radiation reaction $\dot{r}_{\rm rad} = \dot{r}_{\rm dyn}$ in the adiabatic limit, with $\dot{r}_{\rm rad}$ given by [32]

$$\dot{r}_{\rm dyn} \equiv \xi \frac{\partial \mathcal{H}_{\rm real}(r = r_0, p_{r_*}, p_\phi = p_{\phi,0})}{\partial p_{r_*}}, \tag{36}$$

$$\dot{r}_{\rm rad} \equiv \frac{\mathcal{F}_\phi}{dE/dr}, \tag{37}$$

$$\frac{dE}{dr} \equiv -\left[\frac{\frac{\partial \mathcal{H}}{\partial p_\phi}\frac{\partial^2 \mathcal{H}}{\partial r^2}}{\frac{\partial^2 \mathcal{H}}{\partial p_\phi \partial r}}\right]_{(r=r_0, p_{r_*}, p_\phi = p_{\phi,0})}. \tag{38}$$

The initial phase ϕ_0 is set to zero.

2. Orbital evolution

With initial conditions established, Hamilton's equations of motion are integrated forward in time using a GSL adaptive eighth-order Runge-Kutta Prince-Dormand method (or similar) with error control. Tortoise coordinates (r_*, p_{r_*}) are used to enhance numerical stability, particularly near the EOB horizon analog. Integration terminates after a radial separation of $6M$ is reached *and* one of the following conditions is met,

(i) The orbital frequency Ω reaches a maximum and begins to decrease, indicating an unphysical peak in frequency evolution.
(ii) dp_{r_*}/dt becomes positive or peaks, indicating an unphysical outspiral.
(iii) dr/dt becomes positive, indicating an unphysical outspiral.
(iv) The orbital separation r drops below a predefined minimum (e.g., related to $r_{\rm ISCO}$). This is only realized in cases where the parameter $\Delta t^{\rm ISCO}$ is positive, indicating the NR waveform strain peaks before the EOB trajectory reaches the ISCO of the remnant black hole.
(v) The circular orbit frequency $\Omega_{\rm circ}$ exceeds unity at radius $r < 3M$. This is indicative of unphysically high rotation speeds and is used to terminate the integration.

Throughout the integration, quantities required for radiation reaction fluxes $(\mathcal{F}_r, \mathcal{F}_\phi)$ and waveform modes $(h^{\rm F}_{lm})$, such as $\mathcal{H}_{\rm real}$, Ω, and $\Omega_{\rm circ}$, are computed from the phase-space state and binary parameters. These quantities are stored at variable resolution, with coarser sampling during the early, slower inspiral and finer sampling of the last $250M$ in time for performing NQC corrections.

3. Factorized waveform generation

Factorized inspiral waveform modes $h^{\rm F}_{lm}$ are computed at each step of the finely sampled EOB trajectory. These modes are constructed symbolically within NRPy and generated into optimized C code. This symbolic generation significantly facilitates faster computation of waveform modes, particularly by enabling precomputation of coefficients that depend only on constant binary parameters. This benefit is particularly important for precessing binaries, where these coefficients become time dependent due to spin vector evolution and thus cannot be precomputed during integration.

B. Nonquasicircular corrections and merger attachment

To accurately model the late inspiral and plunge, NQC corrections are applied to the dominant $(l=2, m=2)$ factorized EOB mode $h^{\rm F}_{22}$, as described by Eq. (12). This involves determining the NQC amplitude coefficients $\{a_1, a_2, a_3\}$ and phase coefficients $\{b_1, b_2\}$ by solving 3×3 and 2×2 linear systems [Eqs. (14) and (16)], respectively, using GSL routines for lower-upper decomposition and forward/backwards substitution.

These systems require quantities computed at the predicted matching time, t_0, defined as,

$$t_0 = t^{\rm ISCO} - \Delta t^{\rm ISCO}, \tag{39}$$

where $t^{\rm ISCO}$ is the time the EOB trajectory crosses the ISCO radius of the remnant black hole, and $\Delta t^{\rm ISCO}$ is a calibration parameter that sets the time at which the NR waveform strain peaks relative to the ISCO time. Both $t^{\rm ISCO}$ and t_0 are determined from the finely sampled portion of the trajectory by identifying the nearest index in the radial or time arrays, followed by interpolation between the ten nearest sample points at higher resolution.

In a minority of cases, if ODE integration terminates before the perturber crosses the remnant ISCO or before t_0, then t_0 is set to the last sample point of the trajectory. Consistency in NQC coefficient calculation is crucial, as GSL cubic spline interpolation (used for time derivatives at the last sample point) applies natural boundary conditions, while pySEOBNR (also used) employs not-a-knot boundary conditions, leading to significant differences. To ensure consistency, we modified pySEOBNR to use natural boundary conditions for all its cubic spline interpolations.

1. `seobnrv5_nrnqc_bob`: NR-informed NQCs and BOB merger-ringdown

This variant mirrors the `SEOBNRv5HM` approach for NQC corrections. The target values for the right-hand side of Eqs. (14) and (16)—specifically, the amplitude $|h^{\rm NR}_{22}|$, its first and second time derivatives, the frequency $\omega^{\rm NR}_{22}$, and its first time derivative—are derived from fits to NR simulations, evaluated at the peak of the (2,2) mode of the strain. After constructing the NQC-corrected inspiral-plunge waveform $h^{\rm insp\text{-}plunge}_{22}$, the BOB model (Sec. III) is attached at t_0 to describe the merger and ringdown. This attachment procedure ensures $\mathcal{C}^1$ continuity of the complex strain h and its first time derivative $\dot{h}$ at t_0.

2. `seobnrv5_bob`: BOB-informed NQCs and merger-ringdown

This model integrates BOB more deeply into the inspiral by leveraging it to inform the NQC corrections. The functional form of the NQC corrections [Eq. (12)] is maintained, but the target values for the linear systems are sourced analytically from the BOB model itself [Eqs. (26), (31), and (32)], evaluated at $t_0 = t^{\rm peak}_{22}$. Specifically, the required amplitude and frequency derivatives at the matching time, $(d^2|h_{\rm BOB}|/dt^2|_{t_0}$ and $d\omega_{\rm BOB}/dt|_{t_0})$, are computed analytically from the BOB equations, with NRPy performing symbolic differentiation and generating C code for their evaluation. Remnant parameters M_f and a_f (essential for computing $\omega_{\rm QNM}$ and τ within the BOB model) are obtained from NR fits [17,18]. The QNM frequency $\omega_{\rm QNM}$ and damping time τ are computed by interpolating over a set of 107 values of the final spin and the corresponding QNM values computed in Python using the QNM package [28].

Once NQC coefficients are found using either BOB-derived or NR-informed targets, $h^{\rm insp\text{-}plunge}_{22}$ is constructed by multiplying the factorized waveform $h^{\rm F}_{22}$ by the NQC correction given in Eq. (12). Since the factorized waveform is computed at the sample points of the ODE trajectory, no additional interpolation is required. For $t \geq t_0$, the waveform is subsequently described by the same BOB model used for NQC targets. This construction, leveraging BOB-derived NQC targets and the BOB model for merger-ringdown, inherently ensures $\mathcal{C}^2$ continuity of $h, \dot{h}$, and $\ddot{h}$ at t_0.

In order to perform consistency checks against the `SEOBNRv5HM` model, we also implement the native `SEOBNRv5HM` merger-ringdown model given in Eq. (18). This model is implemented in NRPy as `seobnrv5_nrpy`.

The `SEBOB` waveform code used in this paper can be generated with the following commands (flag `-h` provides instructions to generate specific variants):

```
1. pip install git + [36]
2. python3 -m nrpy.examples.seobnrv5_aligned_spin_inspiral -h
```

The C code can be built and run with the following commands (flag `-h` provides instructions to pass input parameters):

```
1. cd project/seobnrv5_aligned_spin_inspiral
2. make
3. ./seobnrv5_aligned_spin_inspiral -h
```

To access up-to-date versions of developmental features discussed in this paper, we recommend directly cloning the repository—`git clone` [37].

V. RESULTS

We present a systematic evaluation of the SEBOB waveform model. After verifying numerical consistency between our NRPy implementation and the public pySEOBNR code, we compare SEBOB and other state-of-the-art EOB models to all 607 quasicircular, aligned-spin black hole binary, NR waveforms from the SXS catalog (v2025.0.10 release [38]). We then test the key assumptions underlying the BOB merger–ringdown approximation and, finally, benchmark the computational efficiency of our implementation.

Throughout, waveform accuracy is quantified by the noise-free mismatch of the dominant (2,2) strain mode,

$$\mathcal{M}(h_1, h_2) = 1 - \frac{|\langle h_1|h_2\rangle|}{\sqrt{\langle h_1|h_1\rangle\langle h_2|h_2\rangle}}, \tag{40}$$

with inner product

$$\langle h_1|h_2\rangle = \int \tilde{h}_1^*(f)\tilde{h}_2(f)\mathrm{d}f, \tag{41}$$

where $\tilde{h}(f)$ is the Fourier transform of the real part of the time-domain strain $h(t)$. The mismatches are calculated using the open source Python package PyART [39].

A. Numerical consistency

A known implementation difference between NRPy and pySEOBNR is the choice of cubic-spline boundary conditions, which in turn affects NQC coefficients when the peak strain time is not attained by the trajectory. To isolate purely numerical effects we consider two modified pySEOBNR variants,

(1) pyseobnr_pert, which applies random 10^{-15}-level perturbations to the mass ratio and spins.
(2) pyseobnr_nat, which replaces all interpolations with natural splines [40].

Figure 1 shows that altering the interpolation boundary conditions induces slightly larger mismatches than round-off perturbations. By contrast, Fig. 2 demonstrates round-off-level agreement between NRPy and pyseobnr_nat, confirming that our code faithfully reproduces the core SEOBNRv5 dynamics. In Table II, we state the typical frequency of cases when the estimated peak time is not encountered. We profile stop conditions for a random combination of 5000 mass ratios, spins, and start

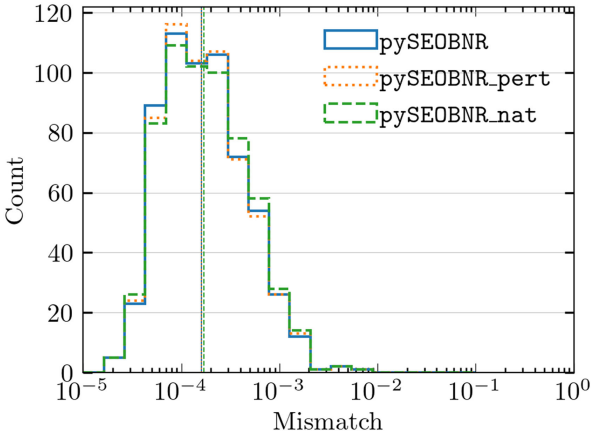

FIG. 1. Mismatch histograms between the public pySEOBNR implementation and the SXS catalog. The orange curve (pyseobnr_pert) shows round-off-level perturbations; the green curve (pyseobnr_nat) adopts natural-spline boundary conditions. Vertical lines mark median mismatches. The green curve has a slightly higher median mismatch than round-off perturbation indicating a small number of cases where the spline boundary conditions affect the accuracy of the calibration.

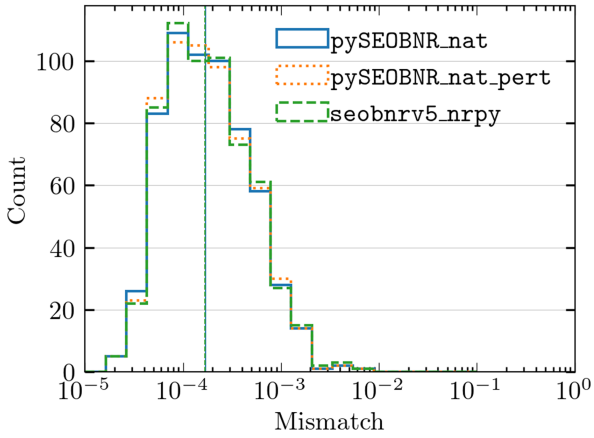

FIG. 2. Mismatch histograms comparing the NRPy implementation of SEOBNRv5 with pyseobnr_nat against the SXS catalog. The orange curve (pyseobnr_nat_pert) is pyseobnr_nat with round-off-level perturbations. The NRPy implementation (green) is consistent with the pyseobnr_nat version due to the mismatches overlapping within the binwidth compared to round-off perturbations.

TABLE II. Frequency of ODE stop conditions for the SEOBNRv5 waveform model. The conditions are calculated over 5000 randomly sampled points with $1 \le q \le 20$, $-0.999 \le \chi_i \le 0.999$, and $5 \times 10^{-3} \le \Omega_{\rm start} \le 1.5 \times 10^{-2}$. The frequency for the 607 quasicircular, spin-aligned, black hole binaries simulated by the SXS catalog is given in brackets. For each stop condition, we specify how often the ODE was terminated before remnant ISCO and/or $t_{22}^{\rm peak}$ crossing. The two non-ISCO crossings when $r < r_{\rm stop} < r_{\rm ISCO}$, happen because $dt_{\rm ODE} \approx 0.1 \equiv dt_{\rm sample}$ at last time leading to the exclusion of $r_{\rm ISCO}$ in the fine sampling.

Condition	Total	ISCO	Peak
Ω peak	3220(369)	78(0)	530(20)
p_{r_*} peak	1080(103)	0(0)	0(0)
$\dot{r} > 0$	6(0)	0(0)	0(0)
$r < r_{\rm stop}$	684(132)	2(0)	0(0)
$r < 3$ and $\Omega > 1$	10(3)	0(0)	0(0)

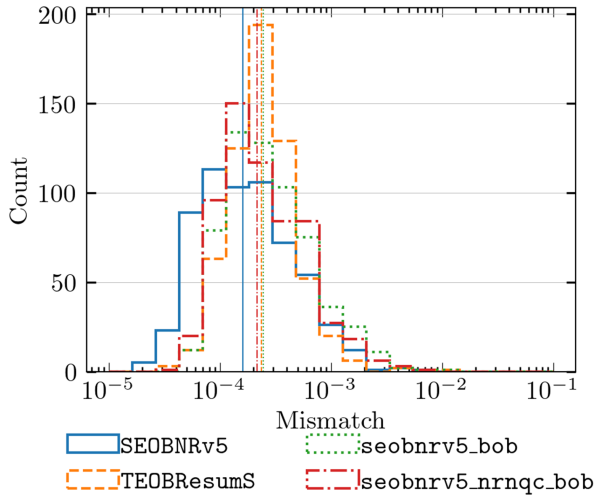

FIG. 3. Mismatch histograms comparing state-of-the-art EOB models, including SEOBNRv5, TEOBResumS-GIOTTO [16], and the two SEBOB variants that differ only in their NQC prescriptions against the SXS catalog. All models achieve median mismatches $\mathcal{M} \approx 2 \times 10^{-4}$.

frequencies to estimate how often this can be encountered during parameter estimation. We note that NQC conditions are evaluated at the endpoint of the trajectory in 10% of the random sample and 3% of SXS simulations.

B. Accuracy of SEBOB

Figure 3 compares mismatches against the SXS catalog for all publicly available EOB models. Both seobnrv5_bob and seobnrv5_nrnqc_bob are competitive with existing models, with a small loss in median accuracy due to the replacement of NR-fitted information in the merger-ringdown regime with BOB. To highlight how BOB-modeled merger-ringdown and NQC corrections manifest in the waveform, we show the waveform strain amplitude (top) and frequency (bottom) of the (2,2) mode for the case where BOB is least accurate against SXS compared to SEOBNRv5 (Fig. 4) and the case where BOB is most accurate against SXS compared to SEOBNRv5 (Fig. 5). Comparing both cases, we find that BOB provides a more accurate description of the frequency evolution compared to the amplitude, even when used as a replacement for NR-fitted NQC corrections. Particularly, in Fig. 5, we note that seobnrv5_bob poorly models the curvature (second derivative) of the amplitude close to the peak. Since BOB uses the assumption of adiabaticity ($|h| \approx |\psi_4|\omega^2$, i.e. $|\psi_4|$ is slowly varying), the variation of the strain amplitude is not described accurately if the variation in the ψ_4 amplitude is non-negligible. In the

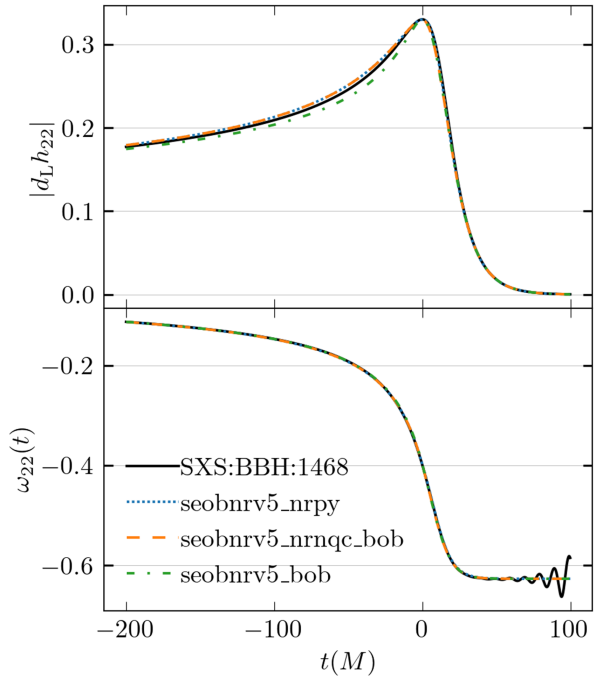

FIG. 4. Waveform strain amplitude (top) and frequency (bottom) of the (2,2) mode for the lowest mismatch case (SXS:BBH:1468) ($q = 3, \chi_1 = -0.599, \chi_2 = -0.399$), and the corresponding seobnrv5_nrpy, seobnrv5_nrnqc_bob, and seobnrv5_bob quantities assuming the same system parameters. We note that the BOB-derived NQC corrections provide an accurate description of the frequency before and after the merger.

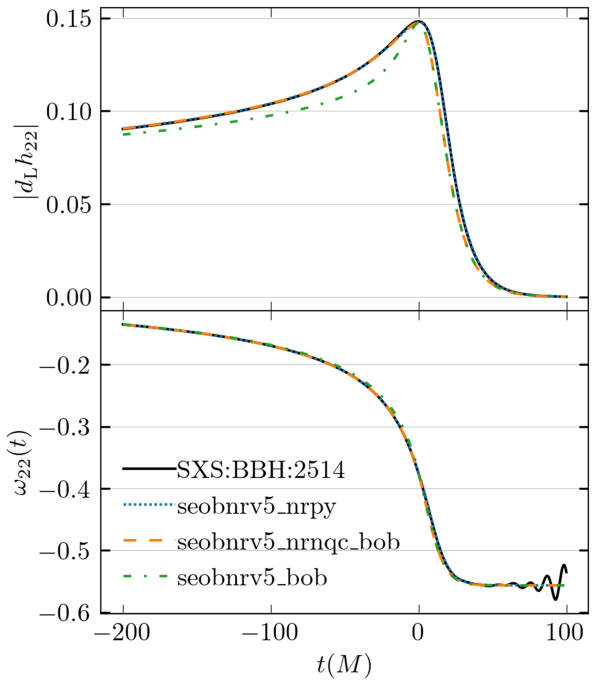

FIG. 5. Waveform strain amplitude (top) and frequency (bottom) of the (2,2) mode for a high-mismatch case (SXS:BBH:2514) ($q = 8, \chi_1 = 0.599, \chi_2 = 0.599$), and the corresponding seobnrv5_nrpy, seobnrv5_nrnqc_bob, and seobnrv5_bob quantities assuming the same system parameters. We note the significant loss in amplitude accuracy prior to merger for this system when BOB is used to compute NQC derivatives.

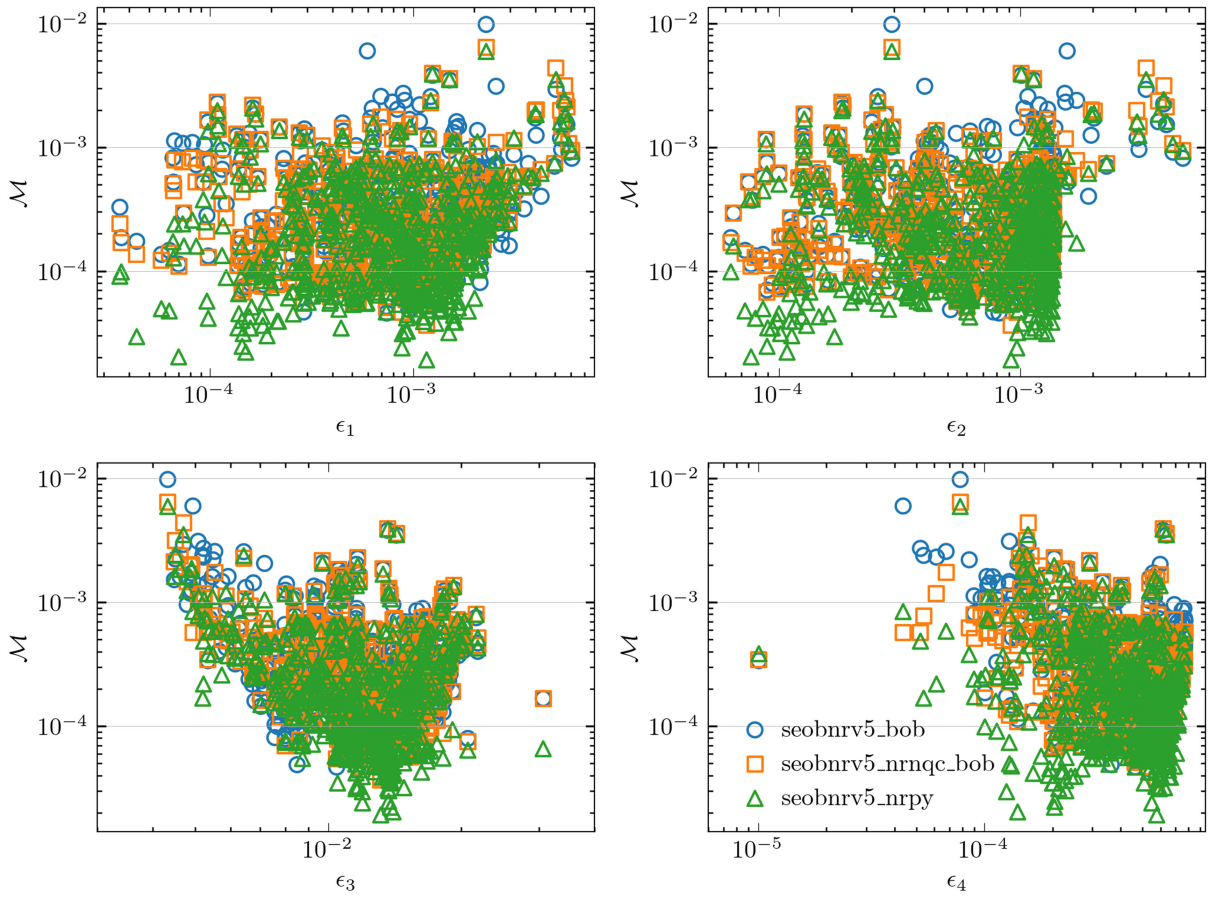

FIG. 6. Scatter plots of mismatch versus the four error metrics ϵ_{1-4} defined in Eqs. (44)–(47). The BOB assumptions are approximately satisfied for the catalog, with only weak correlations to the mismatch.

subsequent discussion, we will assess the validity and impact of the BOB postulates and assumptions.

C. Validity of BOB

The BOB derivation assumes (i) a common phase ϕ for ψ_4, the news $\mathcal{N}$, and the strain h,

$$\phi_{\psi_4} \simeq \phi_{\mathcal{N}} \simeq \phi_h \equiv \phi, \tag{42}$$

and (ii) an adiabatically varying amplitude,

$$|h| \simeq \frac{|\mathcal{N}|}{\omega} \simeq \frac{|\psi_4|}{\omega^2}, \tag{43}$$

where $\omega \equiv \dot{\phi}$. To quantify departures from these postulates we compute four root-mean-squared errors (RMSEs) over the interval $[t_0, t_0 + 20M]$ following peak strain t_0,

$$\epsilon_1 = \mathrm{RMSE}\left(\alpha \frac{\mathrm{d}\omega^2}{\mathrm{d}t}, |\mathcal{N}|^2\right); \quad \alpha = \left(\frac{|\mathcal{N}|^2}{\mathrm{d}\omega^2/\mathrm{d}t}\right)\bigg|_{t_0}, \tag{44}$$

$$\epsilon_2 = \mathrm{RMSE}(A(t), A_p \mathrm{sech}[(t - t_p)/\tau]), \tag{45}$$

$$\epsilon_3 = \mathrm{RMSE}(\omega_{\psi_4}, \omega_h), \tag{46}$$

$$\epsilon_4 = \mathrm{RMSE}(|h|, |\psi_4|/\omega^2). \tag{47}$$

Figure 6 shows that BOB's core relations hold to good accuracy for quasicircular, aligned-spin binaries; the largest mismatches coincide with elevated ϵ_1 and ϵ_2. Improvements will therefore focus on the news-frequency linkage and ψ_4 amplitude model [41].

D. NQC derivatives from BOB

Unlike standard `SEOBNRv5`, which uses NR-fitted derivatives at peak strain, `seobnrv5_bob` computes the second derivative of the amplitude and the first derivative of the frequency analytically from BOB. Their accuracy is assessed in Fig. 7 [42]. The quality of the NR fits and the BOB-derived quantities are measured by relative error,

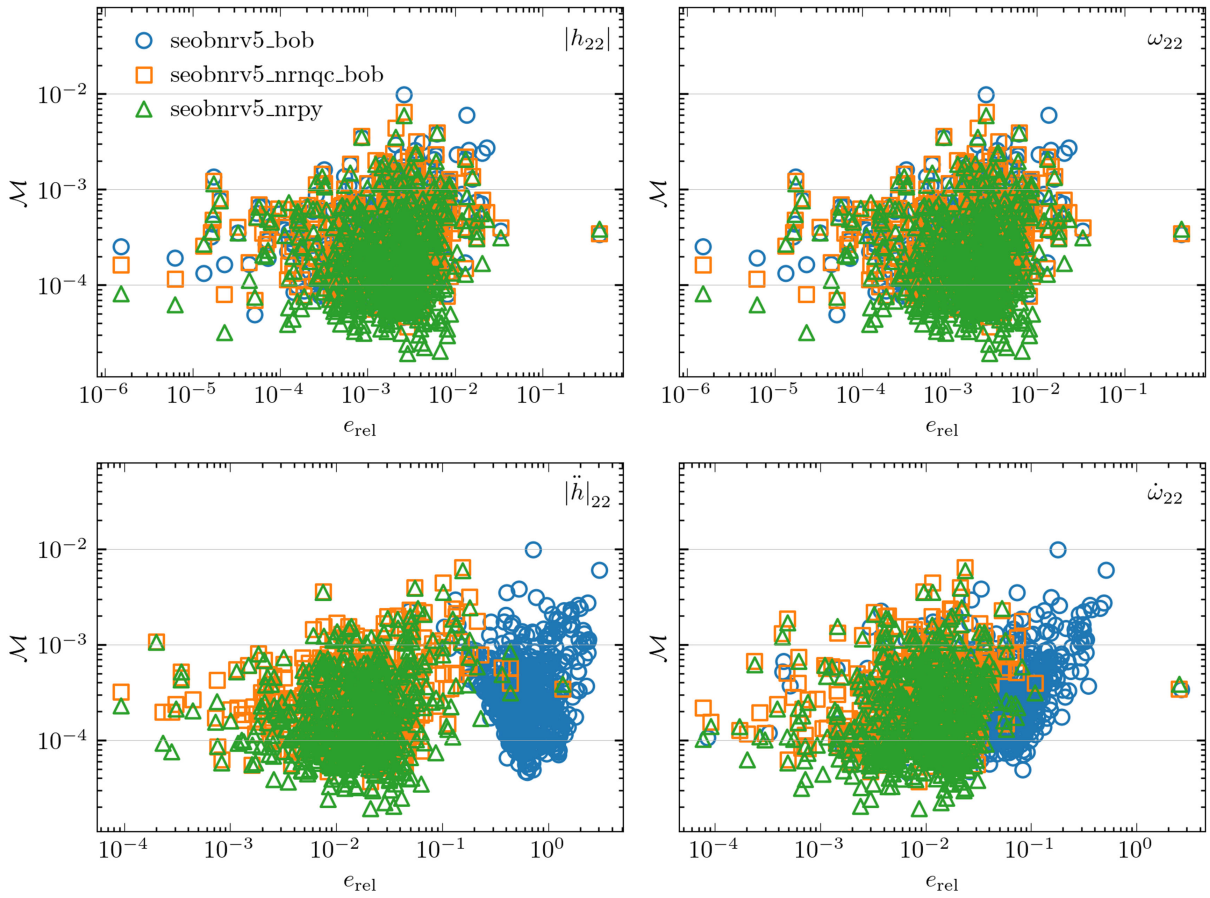

FIG. 7. Relative error $e_{\rm rel}$ of the peak amplitude and frequency fits used in SEOBNRv5 (above) and the fits for the amplitude second derivative and frequency derivative provided as fits in SEOBNRv5 or derived from BOB NQC parameters (below) versus waveform mismatch against the SXS catalog. Each point corresponds to a waveform in the SXS catalog. The relative errors for the NQC parameters are calculated against their corresponding SXS value at the peak strain time. We again note the overall lack of correlation between these errors and the mismatch. The frequency derivative is predicted to within ∼4% on average, while the amplitude's second derivative is less accurate, explaining part of the residual mismatch [42].

$$e_{\rm rel}(y_{\rm NR}, y_{\rm pred}) = \frac{|y_{\rm NR} - y_{\rm pred}|}{|y_{\rm NR}|}, \tag{48}$$

where $y_{\rm NR}$ is the value of the quantity obtained at the time of the peak strain directly from waveforms in the SXS catalog and $y_{\rm pred}$ is the predicted quantity either from the fits given in SEOBNRv5 or derived from BOB.

The frequency derivative is reproduced to ∼4%, close to the ∼1% NR-fit uncertainty, whereas the amplitude's second derivative shows $\mathcal{O}(1)$ relative error. This deficiency stems from the assumption of a slowly varying ψ_4 amplitude. However, even these sizeable parameter errors translate into only mild increases in mismatch; this is evidence that BOB captures the dominant merger-ringdown physics despite its simplified assumptions. We note that these errors can be decreased significantly by formulating BOB to describe the news, rather than ψ_4, and by more accurately calculating the amplitude without assuming adiabaticity, as will be shown in an upcoming paper [41]. Additionally, we note that the single outlying case in the amplitude and frequency fits, as seen in the bottom panels of Fig. 6 and three of the panels in Fig. 7, corresponds to SXS:BBH:1110 ($q = 7$, $\chi_1 = 3 \times 10^{-6}, \chi_2 = -2 \times 10^{-7}$), where the SXS waveform has particularly noisy behavior in the vicinity of merger, as demonstrated in Fig. 8 [42].

E. Efficiency

Because NRPy generates highly optimized C code, we anticipate that our implementation leads to faster waveform generation than the native pySEOBNR implementation.

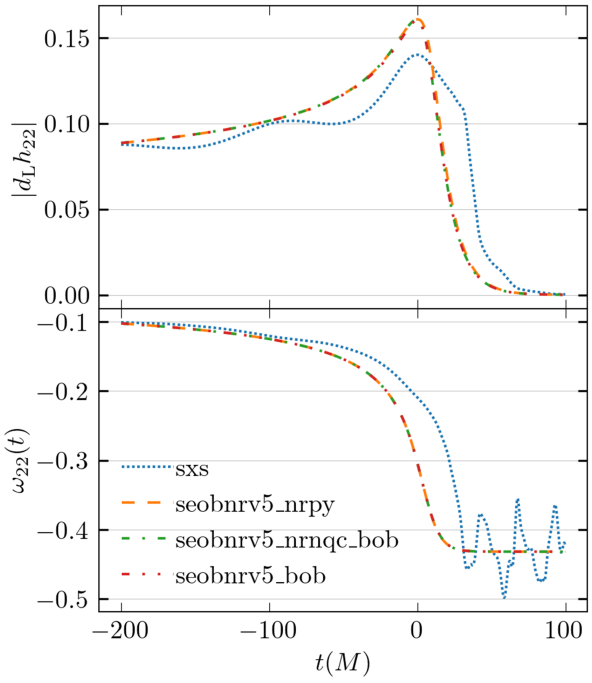

FIG. 8. Waveform strain amplitude (top) and frequency (bottom) of the (2,2) mode for the outlying case (SXS: BBH:1110) ($q = 7, \chi_1 = 3 \times 10^{-6}, \chi_2 = -2 \times 10^{-7}$) [42], and the corresponding `seobnrv5_nrpy`, `seobnrv5_bob`, and `seobnrv5_nrnqc_bob` waveforms assuming the same system parameters. Note that all three models are much more consistent with each other despite the different degrees of NR tuning. This comparison can be particularly helpful in detecting anomalous behavior in large waveform catalogs.

Since pySEOBNR only implements a CSE optimized Hamiltonian, the equivalently optimized flux terms in our implementation should significantly reduce computational time in the ODE right-hand sides. Figure 9 shows that `seobnrv5_bob` is ∼3 times faster than pySEOBNR for $q = [1, 3]$ cases, and is competitive with other state-of-the-art waveform models. The cases shown in Fig. 9 are a representative comparison of wall times, since we found that changing the spin configurations did not affect the relative performance of the waveform models. All timings exclude input/output and postprocessing overhead. `seobnrv4_opt` [43], `NRHybSur3dq8` [13] and `SEOBNRv5_ROM` are timed using their LALSuite's; `TEOBResumS-GIOTTO` uses its stand-alone C code without postadiabatic speedups. One-time Cython compilation costs for pySEOBNR are omitted. In all cases, the wall time is computed by averaging over the wall time of 100 waveform generations. The computations were done using the Thorny Flat cluster at West Virginia University equipped with the Intel® Xeon® Gold 6138 Processor.

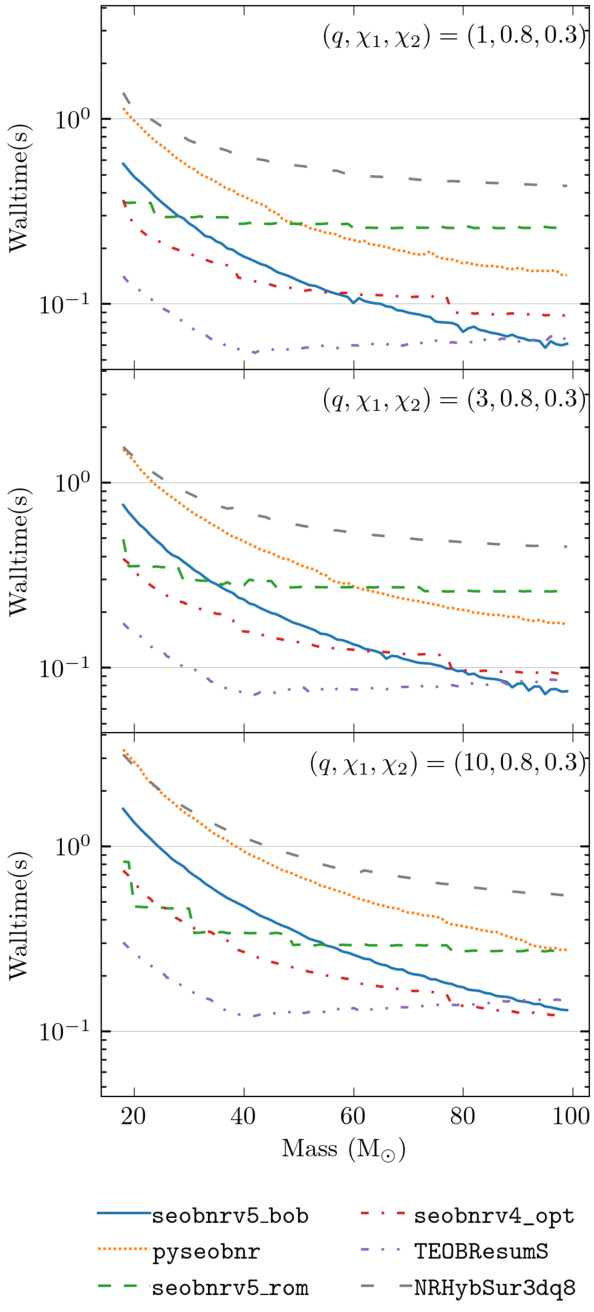

FIG. 9. Wall times for the NRPy implementation of `seobnrv5_bob` compared with leading time-domain generators. Times are measured at a 10 Hz start frequency for $q = \{1, 3, 10\}$ and aligned spins $(\chi_1, \chi_2) = (0.8, 0.3)$.

VI. CONCLUSION

We have introduced `SEBOB`, a hybrid waveform framework that combines an EOB inspiral with a backwards BOB description of the merger–ringdown for aligned-spin binary black holes. Two realizations were constructed atop `SEOBNRv5HM`: (i) `seobnrv5_nrnqc_bob`, which

retains NR-informed NQC corrections and attaches a BOB-based merger–ringdown, and (ii) `seobnrv5_bob`, which uses BOB to supply the NQC targets themselves, reducing reliance on phenomenological fits and enabling $\mathcal{C}^2$ continuity at the attachment time. Both are implemented in the open-source NRPy ecosystem, combining a user-friendly, transparent Python environment and a modularized, efficient C code.

Across a large set of SXS simulations for quasicircular, aligned-spin binaries, both realizations attain median mismatches (with a flat psd) of $\mathcal{M} \simeq 2 \times 10^{-4}$ for the dominant (2,2) mode (Fig. 3), demonstrating that an analytically motivated merger-ringdown can maintain state-of-the-art fidelity. Diagnostics identify a localized imperfection in the curvature of the strain amplitude near its peak when NQC corrections are informed by the current ψ_4-based version of BOB; by contrast, the frequency evolution is captured to within a few percent, with typical peak $|\dot{\omega}|$ errors of ∼4% (Fig. 7). Catalog-wide tests indicate that the core BOB postulates are satisfied to useful accuracy, with only weak correlations between their measured deviations and the overall mismatch (Fig. 6). Even with the localized amplitude shortfall, total mismatches remain at the few-10^{-4} level.

A key practical outcome is computational performance. With symbolic optimization and comprehensive common-subexpression elimination applied to both the Hamiltonian and radiation-reaction fluxes, the NRPy-generated C implementation achieves ∼3× shorter wall times than pySEOBNR for equal and moderate mass ratios while remaining competitive with other time-domain generators (Fig. 9). Since we do not benchmark the time taken to output and process waveforms, a caveat to Fig. 9 is that we do not take into account the time taken to produce a template bank or to perform parameter inference as part of our wall time analysis.

A key limitation of the BOB-informed NQC corrections is a lack of overlap between the EOB and BOB physics in the window where the corrections are necessary. EOB uses PN factorized waveform and fluxes, assumes circularity, and requires late-inspiral corrections to account for highly noncircular orbital motion and the general breakdown of PN close to merger. Upcoming work will explore constructing BOB directly from the news function without the assumption of adiabaticity, which should significantly reduce errors in the amplitude derivatives close to the peak [41]. Within the current paradigm of NQC corrections, this version of BOB can mitigate the need for higher derivatives of NR data to accurately model the amplitude corrections in the late inspiral.

A long-term goal for the `SEBOB` formalism is to eliminate the remaining direct reliance on NR fittings entering the EOB inspiral and dynamics. EOB models rely on extensive calibration of inspiral waveforms to NR waveforms to enhance accuracy. In `SEOBNRv5`, the underlying waveforms already incorporate NR derived through NQC corrections and the phenomenological merger-ringdown (see Table I for a breakdown of precalibration NR parameters). A key advantage of `SEBOB` is that it limits precalibration NR information to four parameters $(|h|_0, \omega_0, M_f, a_f)$, significantly reducing the need for NR to model the merger-ringdown within calibration accuracy. It also reduces the need for extensive calibrations by providing a substantially accurate, physically motivated map from inspiral to merger-ringdown physics. Furthermore, self-consistently modeling noncircularities into the EOB fluxes and inferring remnant properties directly from the EOB inspiral (as suggested in [44]) can potentially reduce the number of NR parameters needed in the inspiral sector. Such improvements would significantly limit the need for NR calibration to a narrow window (between the EOB plunge and the BOB merger) and fewer parameters.

The present study is limited to spin-aligned, quasicircular binaries and to the (2,2) mode, and it continues to use NR fits for remnant properties. Several targeted extensions are natural: (i) improving the amplitude model by formulating BOB in terms of the news rather than ψ_4 and relaxing adiabaticity to sharpen NQC derivatives [41]; (ii) incorporating higher harmonics and, ultimately, generic precessing spins; (iii) tightening the inspiral-to-remnant mapping to further curtail NR-derived inputs; and (iv) developing an open-source pipeline to perform targeted calibrations to NR. By substantially shrinking the role of phenomenological calibration in the most nonlinear phase while retaining competitive accuracy and speed, `SEBOB` offers a robust, interpretable, and extensible path toward waveform modeling in the high-SNR era.

Scientific libraries NumPy [45] and SciPy [46] were used for part of the accuracy analysis.

ACKNOWLEDGMENTS

We would like to acknowledge Anuj Kankani, Matthew Cerep, Lorenzo Pompili, Rossella Gamba, and Aasim Jan for useful comments and discussion. S. M. and S. T. M. were supported in part by National Science Foundation CAREER Grant No. 1945130 and NASA Grants No. 22-LPS22-0022 and No. 24-2024EPSCoR-0010. We acknowledge the computational resources provided by the WVU Research Computing Spruce Knob HPC cluster, which is funded in part by NSF Grant No. EPS-1003907, and the Thorny Flat HPC cluster, which is funded in part by NSF Grant No. OAC-1726534. This work was performed in part at the Aspen Center for Physics, which is supported by National Science Foundation Grant No. PHY-2210452.

DATA AVAILABILITY

The data that support the findings of this article are openly available [37].

[1] R. Abbott *et al.* (KAGRA, Virgo, and LIGO Scientific Collaborations), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, Phys. Rev. X **13**, 041039 (2023).

[2] N. Yunes and X. Siemens, Gravitational-wave tests of general relativity with ground-based detectors and pulsar timing-arrays, Living Rev. Relativity **16**, 9 (2013).

[3] A. Palmese, C. R. Bom, S. Mucesh, and W. G. Hartley, A standard siren measurement of the Hubble constant using gravitational-wave events from the first three LIGO/Virgo observing runs and the DESI legacy survey, Astrophys. J. **943**, 56 (2023).

[4] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Population properties of compact objects from the second LIGO-Virgo gravitational-wave transient catalog, Astrophys. J. Lett. **913**, L7 (2021).

[5] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi *et al.*, LASER INTERFEROMETER SPACE ANTENNA, arXiv:1702.00786.

[6] M. Punturo *et al.*, The Einstein Telescope: A third-generation gravitational wave observatory, Classical Quantum Gravity **27**, 194002 (2010).

[7] D. Reitze *et al.*, Cosmic Explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO, Bull. Am. Astron. Soc. **51**, 035 (2019), https://inspirehep.net/literature/1743201.

[8] V. Kapil, L. Reali, R. Cotesta, and E. Berti, Systematic bias from waveform modeling for binary black hole populations in next-generation gravitational wave detectors, Phys. Rev. D **109**, 104043 (2024).

[9] B. S. Sathyaprakash, B. F. Schutz, and C. Van Den Broeck, Cosmography with the Einstein Telescope, Classical Quantum Gravity **27**, 215006 (2010).

[10] M. Boyle *et al.*, The SXS Collaboration catalog of binary black hole simulations, Classical Quantum Gravity **36**, 195006 (2019).

[11] C. M. Will, Gravity: Newtonian, post-Newtonian, and general relativistic, in *Gravity: Where Do We Stand?*, edited by R. Peron, M. Colpi, V. Gorini, and U. Moschella (Springer, Cham, 2016), pp. 9–72.

[12] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries, Phys. Rev. D **102**, 064002 (2020).

[13] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, Surrogate model of hybridized numerical relativity binary black hole waveforms, Phys. Rev. D **99**, 064045 (2019).

[14] A. Buonanno and T. Damour, Effective one-body approach to general relativistic two-body dynamics, Phys. Rev. D **59**, 084006 (1999).

[15] L. Pompili *et al.*, Laying the foundation of the effective-one-body waveform models SEOBNRv5: Improved accuracy and efficiency for spinning nonprecessing binary black holes, Phys. Rev. D **108**, 124035 (2023).

[16] A. Nagar, G. Riemenschneider, G. Pratten, P. Rettegno, and F. Messina, Multipolar effective one body waveform model for spin-aligned black hole binaries, Phys. Rev. D **102**, 024077 (2020).

[17] X. Jiménez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Pürrer, Hierarchical data-driven approach to fitting numerical relativity data for nonprecessing binary black holes with an application to final spin and radiated energy, Phys. Rev. D **95**, 064024 (2017).

[18] F. Hofmann, E. Barausse, and L. Rezzolla, The final spin from binary black holes in quasi-circular orbits, Astrophys. J. Lett. **825**, L19 (2016).

[19] S. T. McWilliams, Analytical black-hole binary merger waveforms, Phys. Rev. Lett. **122**, 191102 (2019).

[20] E. Newman and R. Penrose, An approach to gravitational radiation by a method of spin coefficients, J. Math. Phys. (N.Y.) **3**, 566 (1962).

[21] E. Berti, V. Cardoso, and A. O. Starinets, Quasinormal modes of black holes and black branes, Classical Quantum Gravity **26**, 163001 (2009).

[22] C. V. Vishveshwara, Scattering of gravitational radiation by a Schwarzschild black-hole, Nature (London) **227**, 936 (1970).

[23] W. H. Press, Long wave trains of gravitational waves from a vibrating black hole, Astrophys. J. **170**, L105 (1971).

[24] I. Ruchlin, Z. B. Etienne, and T. W. Baumgarte, SENR/NRPy+: Numerical relativity in singular curvilinear coordinate systems, Phys. Rev. D **97**, 064036 (2018).

[25] D. P. Mihaylov, S. Ossokine, A. Buonanno, H. Estelles, L. Pompili, M. Pürrer, and A. Ramos-Buades, pySEOBNR: A software package for the next generation of effective-one-body multipolar waveform models, SoftwareX **30**, 102080 (2025).

[26] LIGO Scientific, Virgo, and KAGRA Collaborations, LVK Algorithm Library—LALSuite, Free software (GPL) (2018), 10.7935/GT1W-FZ16.

[27] A. Nagar, P. Rettegno, R. Gamba, S. Albanesi, A. Albertini, and S. Bernuzzi, Analytic systematics in next generation of effective-one-body gravitational waveform models for future observations, Phys. Rev. D **108**, 124018 (2023).

[28] L. C. Stein, QNM: A Python package for calculating Kerr quasinormal modes, separation constants, and spherical-spheroidal mixing coefficients, J. Open Source Software **4**, 1683 (2019).

[29] M. Khalil, A. Buonanno, H. Estelles, D. P. Mihaylov, S. Ossokine, L. Pompili, and A. Ramos-Buades, Theoretical groundwork supporting the precessing-spin two-body dynamics of the effective-one-body waveform models SEOBNRv5, Phys. Rev. D **108**, 124036 (2023).

[30] J. G. Baker, W. D. Boggs, J. Centrella, B. J. Kelly, S. T. McWilliams, and J. R. van Meter, Mergers of non-spinning black-hole binaries: Gravitational radiation characteristics, Phys. Rev. D **78**, 044046 (2008).

[31] B. J. Kelly, J. G. Baker, W. D. Boggs, S. T. McWilliams, and J. Centrella, Mergers of black-hole binaries with aligned spins: Waveform characteristics, Phys. Rev. D **84**, 084009 (2011).

[32] A. Buonanno, Y. Chen, and T. Damour, Transition from inspiral to plunge in precessing binaries of spinning black holes, Phys. Rev. D **74**, 104005 (2006).
[33] Z. B. Etienne, Improved moving-puncture techniques for compact binary simulations, Phys. Rev. D **110**, 064045 (2024).
[34] A. Meurer *et al.*, SymPy: Symbolic computing in Python, PeerJ Comput. Sci. **3**, e103 (2017).
[35] M. C. Galassi *et al.*, *GNU Scientific Library Reference Manual* (2019), ISBN 0954612078.
[36] https://github.com/nrpy/nrpy.git@e16a119.
[37] https://github.com/nrpy/nrpy.git.
[38] M. A. Scheel *et al.*, The SXS Collaboration's third catalog of binary black hole simulations, Classical Quantum Gravity **42**, 195017 (2025).
[39] https://github.com/RoxGamba/PyART.
[40] Natural splines set the second derivative to zero at each endpoint, whereas the public code employs "not-a-knot" conditions.
[41] A. Kankani and S. McWilliams, BOB the (Waveform) builder: Optimizing analytical black-hole binary merger waveforms, arXiv:2510.25012.
[42] The highest error case in the amplitude and frequency fits corresponds to SXS:BBH:1110 ($q = 7, \chi_1 = 3 \times 10^{-6}$, $\chi_2 = -2 \times 10^{-7}$), which was deprecated during final development of the paper.
[43] C. Devine, Z. B. Etienne, and S. T. McWilliams, Optimizing spinning time-domain gravitational waveforms for Advanced LIGO data analysis, Classical Quantum Gravity **33**, 125025 (2016).
[44] A. Buonanno, L. E. Kidder, and L. Lehner, Estimating the final spin of a binary black hole coalescence, Phys. Rev. D **77**, 026004 (2008).
[45] S. van der Walt, S. C. Colbert, and G. Varoquaux, The NumPy array: A structure for efficient numerical computation, Comput. Sci. Eng. **13**, 22 (2011).
[46] P. Virtanen *et al.*, SciPy 1.0–Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).

Having established a robust analytical model for the vacuum dynamics of black hole binaries and the resulting gravitational waveforms, the next chapter will shift focus to the astrophysical environments these binaries inhabit, specifically exploring their interactions with circumbinary accretion disks.

Chapter 3

Impact of black hole binaries on accretion disks

The previous chapter discussed the modeling of the dynamics of black hole binaries due to GR and the subsequent generation of gravitational waves. One consequence of the PN evolution equations is that they apply to black hole binaries of any mass, from the stellar mass binaries that are detected by LIGO/Virgo/KAGRA, to the massive black hole binaries expected to be detected by LISA, to the supermassive black hole binaries detectable by pulsar timing arrays (PTAs). Additionally, the dynamics of the black hole binaries were only used to derive the gravitational wave strain and no other observable quantities.

In reality, black hole binaries of various masses occupy a diverse range of astrophysical environments. It is, therefore, not only possible that the dynamics of the binary are affected by the surrounding environment, but it is also possible to observe these binaries through electromagnetic radiation emitted by this environment. This multi-messenger approach to strongly gravitating systems is important for understanding the formation and evolution of these systems [15].

For the case of massive black hole binaries ($M = 10^5 - 10^7 M_\odot$) that are expected to be detected by LISA, the observable electromagnetic counterpart is expected to come from the accretion disks surrounding the black holes. In this section, I will provide an overview of how the radiation from the accretion disk was expected to follow the evolution of the black hole binary, based on initial theoretical predictions by Goldreich and Tremaine [16]. I will then present results from accurate hydrodynamic simulations that provide a contradictory description. The subsequent sections provide an alternative theoretical description of binary-disk interactions at short timescales that offers

resolution and shows promising predictive power.

For an accretion disk evolving under the gravitational influence of a Newtonian binary system, Goldreich and Tremaine [16] predicted that the binary system would deposit angular momentum in the form of torques at discrete resonant locations called Lindblad resonances. The location of these Lindblad resonances are given by

$$r_{\mathrm{LR}} = \left(\frac{m+1}{N}\right)^{2/3} a. \tag{3.1}$$

where a is the semi-major axis of the binary, and $m, N = 1, 2, 3, ...$ label the resonances and come from the Fourier decomposition of the gravitational potential (note that N here denotes the epicyclic harmonic number, distinct from the principal action used in Chapter 2). The viscosity of the disk exerts an opposing torque that clears the angular momentum deposited by the binary at the resonances. The innermost resonant locations, however, have very strong torques leading to angular momentum being deposited at a rate much higher than that at which viscous damping can clear. As a result, the fluid near these resonances is cleared and a *circumbinary gap* is formed with a width given by the outermost resonance that has a higher gravitational torque than viscous torque.

The balance between these two torques provides a gap maintenance timescale; that is, the timescale over which the viscous torque is able to provide the torque necessary to balance the gravitational torque. This timescale, t_{visc}, depends on the typical gap size, and as a result on the semi-major axis of the binary as [16]

$$t_{\mathrm{visc}} \sim a^2. \tag{3.2}$$

If the binary begins to inspiral due to the emission of gravitational waves, the quadrupole formula discussed in the last section gives an inspiral timescale of [4]

$$t_{\mathrm{GW}} \equiv \frac{a}{|da/dt|} = \frac{5}{64}\frac{c^5 a^4}{G^3 M^3 \nu} \sim a^4. \tag{3.3}$$

Therefore, as the binary inspirals ($a \to 0$), there is a critical semi-major axis, a_c, at which the inspiral timescale becomes much shorter than the gap maintenance timescale. At this point, changes in the binary semi-major axis are not reflected in the circumbinary gap and the dynamics of the disk and binary decouple. This reasoning was provided by Milosavljevic and Phinney [4] to suggest that once the evolution of the binary is dominated by gravitational radiation, the disk, which can radiate light, will only reflect the merger after a viscous timescale and therefore result in an *afterglow*, when

the disk moves inwards due to the lack of any gravitational torques. Therefore, while there is an electromagnetic counterpart to massive black hole mergers, it does not contain any imprints of the gravitational-wave-driven evolution of the binary.

As the capabilities of numerical hydrodynamic simulations evolved, Tang et al. [5] showed that disk-binary systems initialized close to this decoupling scale do not actually decouple and instead the disk moves inwards and interacts with the binary over much shorter timescales. The predicted electromagnetic signals also contain periodic behaviors that relate to the binary dynamics, which can act as a "smoking gun" for the detection of an inspiraling black hole binary system.

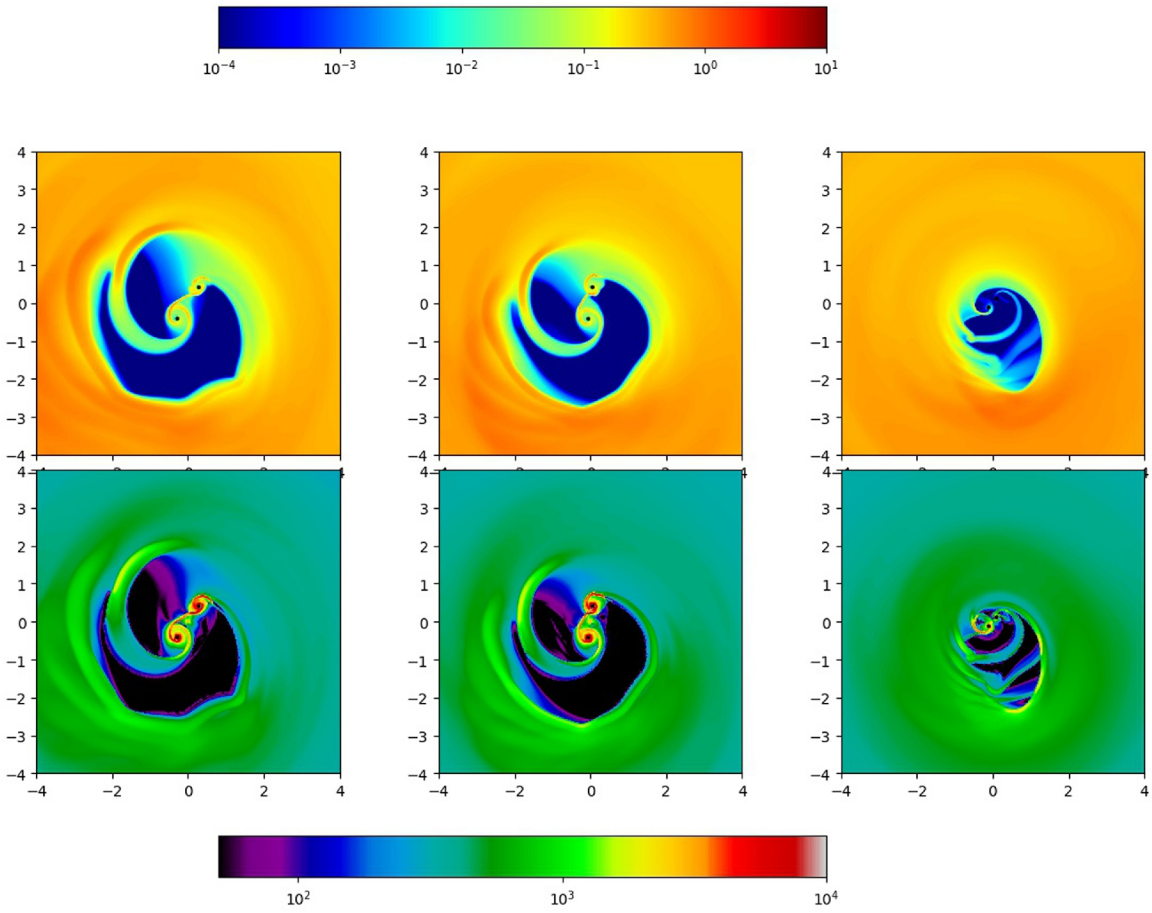

Figure 3.1: Time averaged snapshots of the normalized surface density (top row) and effective temperature (bottom row, in eV) of the simulated accretion disks from [5]. Each column corresponds to different stages in the inspiral of the binary. The dark central region is the circumbinary gap and it can be seen to shrink as the binary inspirals.

The results of [5] are in stark contrast to the predictions of Milosavljevic and Phinney [4]. The proposed reason for this discrepancy was the assumption that while resonant disk-binary interactions are relevant for the long-term evolution of the circumbinary disk, viscosity does not dominate the

disk-binary interactions at the typical distance of the circumbinary gap $\approx 2a$. This motivates an investigation of disk dynamics at much shorter timescale that is dominated by the gravitational force of the binary on individual fluid elements in the disk.

The subsequent section provides a pedagogical derivation of the established theoretical reasoning for the formation and maintenance of the circumbinary gap. The section makes a new proposal for the location of the circumbinary gap based on the treatment of the disk as a collection of test particles evolving purely under the gravitational force of the binary. Additionally, the section provides an analysis of numerical simulations of disk-binary system to verify how the picture derived in [16] may be inaccurate and how the proposed explanation provides a better measure of the location of the circumbinary gap and the timescales of its maintenance.

The disk analysis is divided into two connected studies. The first establishes the analytical framework in the simpler coplanar case, where the binary and disk angular momenta are aligned and the gap structure can be compared directly with earlier resonant-torque predictions. The second tests the generality of the same framework in inclined systems, where the binary potential drives qualitatively new vertical and time-dependent disk behavior. The two studies therefore serve complementary roles: the coplanar case validates the epicyclic-instability picture in the standard setting, while the inclined case demonstrates its predictive power beyond the regime where the traditional resonant picture is most naturally applied.

3.1 First Principles Analysis of Circumbinary Disk Dynamics

All material relevant to this section is contained within the publications[17]. The text in this publication contains a first principles derivation and summary of disk-binary interactions in the test particle and fluid limits.

My primary contributions to the coplanar disk study were fourfold. First, I reconstructed the standard resonant-torque picture of circumbinary gap formation and identified the assumptions required to approximately reproduce the Artymowicz and Lubow prediction. Second, I developed the alternative test-particle framework based on epicyclic stability in the restricted three-body problem. Third, I derived and applied the azimuthal harmonic decomposition used to test whether the disk density satisfies the assumptions of the resonant-torque model. Fourth, I analyzed the simulation data and compared the numerical gap locations with both the resonant-torque and epicyclic-

instability predictions. These contributions produced the main interpretive claim of the study: the observed gap location is better explained by short-timescale orbital instability than by long-timescale resonant torque balance.

The Astrophysical Journal, 973:18 (17pp), 2024 September 20 https://doi.org/10.3847/1538-4357/ad6149

Analytical and Numerical Analysis of Circumbinary Disk Dynamics. I. Coplanar Systems

Siddharth Mahesh[1,2], Sean T. McWilliams[1,2], and Michal Pirog[1,2]
[1] Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA; sm0193@mix.wvu.edu
[2] Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA
Received 2023 December 21; revised 2024 June 26; accepted 2024 June 27; published 2024 September 12

Abstract

We present an analytical and numerical study of a system composed of a stellar binary pair and a massless, locally isothermal viscous accretion disk that is coplanar to the binary orbital plane. Analytically, we study the effect of the binary's gravitational potential over short timescales through the stability of epicyclic orbits, and over long timescales by revisiting the concept of resonant torques. Numerically, we perform two-dimensional Newtonian simulations of the disk-binary system over a range of binary mass ratios. We find that the results of our simulations are consistent with those of previous numerical studies. We additionally show, by comparison of the analytical and numerical results, that the circumbinary gap is maintained on the orbital timescale through the driving of epicyclic instabilities, and does not depend on resonant torquing, contrary to the standard lore. While our results are applicable to any disk-binary system, we highlight the importance of this result in the search for electromagnetic and gravitational-wave signatures from supermassive black hole binaries.

Unified Astronomy Thesaurus concepts: Astrophysical fluid dynamics (101); Accretion (14); Circumstellar disks (235); Orbital theory (1182); Orbital resonances (1181); Computational methods (1965); Astrophysical black holes (98)

1. Introduction

The dynamics of gaseous accretion disks under the gravitational influence of binary point masses is a relevant topic in multiple astrophysical contexts, including protoplanetary disks around binary star systems and accretion disks surrounding supermassive black hole binaries. For an equal mass system (i. e. where the masses of the two stars in the binary system are equal), the binary clears a central low-density cavity out to roughly twice the binary's orbital separation (Milosavljević & Phinney 2005). In the case of supermassive black hole binaries, the traditional picture holds that the binary and the cavity coevolve until the gravitational-wave timescale for the binary drops below the viscous timescale of the inner disk edge, at which point the binary and the disk decouple from each other. This picture is largely motivated by analytical studies (Goldreich & Tremaine 1980; Artymowicz & Lubow 1994), which suggested that the circumbinary gap is opened and maintained by the balancing of dynamical torques from the binary (which deposit angular momentum at discrete locations known as Lindblad resonances) and viscous torques that dissipate the angular momentum through spiral density waves. However, full numerical simulations of circumbinary accretion disks (D'Orazio et al. 2013; Farris et al. 2015; Tang et al. 2018; Tiede et al. 2022; Krauth et al. 2023) have not been consistent with this analytical picture due to the lack of a demonstrable decoupling phase. In addition, in a companion study, we find that the disagreement is more pronounced for disk-binary systems with relative inclinations when we compare an extension of the analytical results presented here with numerical simulations (M. Pirog et al. 2024, in preparation, hereafter Paper II).

Therefore, we revisit the predictions for the circumbinary gap size as presented in Artymowicz & Lubow (1994). We recapitulate the background physics of the system considered in Artymowicz & Lubow (1994) and provide a prescription for reproducing their results. Additionally, in order to reconcile the observed inconsistencies, we suggest a new picture of gap opening that is not contingent on resonant torquing and occurs over timescales comparable to the binary period, namely, the propagation of dynamical instabilities due to the asymmetric nature of the gravitational potential. We show that this new approach successfully predicts the observed trend in the gap sizes with varying mass ratio as studied in this paper, and lends itself naturally to the case of the inclined disk-binary system, which we study in Paper II.

While analytical methods allow for an understanding of the effect of the binary potential on the disk dynamics with minimal inclusion of disk viscosity, numerical studies allow for a full hydrodynamic study of the same system. Previous numerical studies of circumbinary disks were done in Artymowicz & Lubow (1996) using smooth particle hydrodynamics and in Günther & Kley (2002) using a grid-based hydrodynamical model with viscosity. In our approach, we build on the setup introduced in MacFadyen & Milosavljević (2008), where the authors focused on an equal mass binary, and D'Orazio et al. (2013), where the authors considered a varying binary mass ratio and its effect on the accretion process. To conduct our numerical investigations, we use the DISCO code originally described in Duffell (2016), which has been successfully used to solve more advanced problems in circumbinary disk dynamics such as the gap-opening criteria (Duffell & MacFadyen 2013), gas dynamics inside the central cavity (Farris et al. 2014), and disk evolution during the late inspiral stage of a supermassive black hole binary (D'Orazio et al. 2016; Tang et al. 2018).

Our study is divided into two papers on the basis of the two classes of systems being studied (coplanar and inclined), with each paper providing the requisite analytical and numerical

investigations to characterize the system and reconcile the results obtained from both approaches. In this paper, we present the simple, well-studied case of a coplanar binary and circumbinary disk, in order to establish a detailed self-contained reference for the theoretical preliminaries and numerical setup used in this two-part study and future studies. In Paper II, we study the case of inclined disk-binary systems, where we note that the results of the numerical simulations further validate the new analytical approach to understanding the clearing of a central gap.

In this paper, the next section contains the theoretical preliminaries required to motivate the numerical simulations (Section 2.1) and analytical computations (Section 2.2). Section 3 briefly describes the numerical code (we leave readers to consult Duffell 2016 for further details) and our specific setup, as well as the results from our numerical simulations, focusing on the density and angular velocity distributions of the disk. Section 4 presents an analytical investigation of disk dynamics by studying the stability of perturbed orbits in the binary potential and reexamining the resonant torque picture. We introduce two separate definitions for the size of the gap truncated by the binary due to these two descriptions, and present the results of the corresponding gap size computations and a comparison with the numerical results. We summarize and discuss our results in Section 5.

2. Preliminaries

We assume throughout this work that the self-gravity of the disk is negligible, and there is no backreaction on the evolution of the binary due to the disk. The binary itself is composed of two point masses, the more massive one referred to as the *primary* and the less massive one as the *secondary*, and these point masses execute fixed circular Keplerian orbits. The disk, on the other hand, is treated either as a collection of massless test particles for the purpose of analytical calculations, or as a viscous, non-self-gravitating, locally isothermal fluid for the purpose of numerical simulations. This description of the binary is valid in the regime where the binary separation is large relative to the gravitational radii of the two masses, so that the binary does not evolve over the course of the simulation. Even if the binary executes a slow, adiabatic inspiral through the emission of gravitational waves, the orbital separation only changes over a timescale much larger than the orbital period or the viscous timescale, so for our purpose it can be treated as a constant. These two physical treatments approach the dynamics of the disk over different timescales; the analytical treatment addresses the dynamics over timescales comparable to the binary orbital period, whereas the numerical treatment addresses the large-scale structure and long-term evolution of the disk, which requires a full hydrodynamic treatment. The large orbital separations also justify the use of Newtonian physics for the evolution of the overall system.

The binary components are point masses moving in circular, Keplerian orbits in the center-of-mass frame, with the disk moving in the combined Newtonian gravitational potential of both masses. We will see analytically that, for gas orbiting far enough from the binary, the combined binary potential can be expressed as small, periodic perturbations about the potential of a single point mass whose mass equals the total binary mass, and which is at the location of the binary center of mass. The disk is assumed to be geometrically thin and not self-gravitating, justifying a Keplerian rotation law about the center of mass with no additional contribution to the potential due to the density distribution of the gas. Numerically, this rotation law will serve, modulo small corrections, as a leading term in the disk's initial angular velocity distribution. The disk is viscous, with the viscosity following an α-prescription. We work in geometrized units throughout, where the universal gravitational constant, G, and the speed of light, c, satisfy $G = c = 1$; we also set the total mass of the binary system, M, and the orbital separation of the binary, a, to unity. In the following subsections, we provide further details of the disk and binary dynamics.

2.1. Hydrodynamics in the Binary Potential

We work in standard two-dimensional polar coordinates (r, ϕ), and we use the open-source numerical code DISCO (Duffell 2016) to numerically solve the Riemann problem for the four primitive variables: surface mass density $\sigma(\boldsymbol{x}, t)$, surface pressure $p(\boldsymbol{x}, t)$, and radial and azimuthal components of the velocity $\boldsymbol{v}(\boldsymbol{x}, t) = (v^r(\boldsymbol{x}, t), v^\phi(\boldsymbol{x}, t))$. We suppress the functional dependence on $(\boldsymbol{x}, t)$, hereafter and refer to Duffell (2016) for a detailed description of the numerical methods. The Newtonian hydrodynamic field equations can be formulated in the standard "primitive" form typically found in textbooks, or in the "conservative" form typically employed when finding numerical solutions. We employ the latter, and will briefly review the basic equations here. The continuity equation and Euler equation read as (Frank et al. 2002)

$$\partial_t \sigma + \nabla \cdot (\sigma \boldsymbol{v}) = 0, \tag{1}$$

$$\partial_t \boldsymbol{v} + \boldsymbol{v} \cdot \nabla \boldsymbol{v} = \frac{1}{\sigma} \nabla p - \nabla \Phi + \boldsymbol{f}_\nu. \tag{2}$$

Φ is the total gravitational potential due to both test masses,

$$\Phi(\boldsymbol{x}) = \frac{G m_1}{|\boldsymbol{x} - \boldsymbol{x}_1|} + \frac{G m_2}{|\boldsymbol{x} - \boldsymbol{x}_2|}, \tag{3}$$

where m_1 and m_2 are the masses of the binary components, and $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are their positions on fixed circular orbits. $\boldsymbol{f}_\nu$ describes the force due to viscosity ν,

$$\boldsymbol{f}_\nu = \nabla \cdot \nu \nabla \boldsymbol{v} + \nabla \left(\frac{1}{2} \nu \nabla \cdot \boldsymbol{v} \right). \tag{4}$$

The system of coordinates (r, ϕ) lies in the plane of the disk and is centered on the binary's center of mass. We complete this system of equations using the α prescription for viscosity (Shakura & Sunyaev 1973),

$$\nu(r) = \alpha s^2 r^{3/2} \tag{5}$$

and a locally isothermal equation of state,

$$\begin{aligned} p &= s^2 \sigma, \\ s(r) &= \chi / \sqrt{r}, \end{aligned} \tag{6}$$

where s is the speed of sound and $\chi = h/r$ is the constant scale height-to-radius ratio. While this study primarily sets the free parameter of $\alpha = 0.01$, we ran simulations with $\alpha \in \{0.001, 0.003, 0.01, 0.03, 0.1\}$, in order to assess the generality of our outcomes. We set the free parameter $\chi = 0.1$ in all of our simulations.

2.2. Test Particles in the Binary Potential

In our units where $G = M = a = 1$, the only remaining free parameter is the mass ratio of the point masses, which we choose to express as the secondary mass ratio $\mu \equiv m_2/M$, where m_2 is the mass of the secondary. In the center-of-mass frame, the locus of the primary, $\boldsymbol{x}_1$, and the secondary, $\boldsymbol{x}_2$, are given in two-dimensional Cartesian coordinates by

$$\begin{aligned} \boldsymbol{x}_1(t) &= \mu(\cos t, \sin t), \\ \boldsymbol{x}_2(t) &= (\mu - 1)(\cos t, \sin t). \end{aligned} \tag{7}$$

We note that the binary orbital frequency $\Omega_b = \sqrt{\frac{GM}{a^3}}$ has a magnitude of unity due to our choice of normalization, so that the orbital frequency of the binary is simply t in these units. The corresponding Newtonian gravitational potential, Φ, experienced by a test particle located at $\boldsymbol{x}$ is given by

$$\Phi = -\frac{1-\mu}{|\boldsymbol{x} - \boldsymbol{x}_1|} - \frac{\mu}{|\boldsymbol{x} - \boldsymbol{x}_2|}. \tag{8}$$

With some algebra, we can write the potential in two-dimensional polar coordinates (r, ϕ). In this form, one can exploit the generating function of Legendre polynomials as in Poisson & Will (2014) to obtain the multipolar decomposition

$$\Phi = -\sum_{l=0}^{\infty} \mathcal{Q}_l r^{(-l-1)} P_l(\cos(\phi - t)), \tag{9}$$

where $\mathcal{Q}_l \equiv (-\mu)^l(1-\mu) + \mu(1-\mu)^l$ is the dimensionless multipole moment and P_l is the Legendre polynomial of order l. We separate the monopole ($l = 0$) term, which we will refer to as $\Phi_{\rm M}$ and is given by

$$\Phi_{\rm M}(r) = -\frac{1}{r}. \tag{10}$$

$\Phi_{\rm M}$ is static and spherically symmetric, and can be considered the background potential. We then bundle all other multipole terms into a single perturbing potential $\Phi_{\rm P}$, to indicate that the effect of all higher multipoles will enter the subsequent perturbation theory treatment at the same order. We also note that all higher multipoles manifest as harmonics of the variable $(\phi - t)$, and it is therefore convenient to reexpress $\Phi_{\rm P}$ as a Fourier series of the form

$$\begin{aligned} \Phi_{\rm P} &= -\sum_{l=2}^{\infty} \mathcal{Q}_l r^{(-l-1)} \mathcal{P}_l(\cos(\phi - t)) \\ &= \sum_{m=2}^{\infty} \Phi_m \cos(m(\phi - t)). \end{aligned} \tag{11}$$

We begin our summation from $l = 2$, since the dipole moment $\mathcal{Q}_1 = 0$. The expression for the Fourier coefficients, Φ_m, can be found in Equation (16) of Miranda & Lai (2015) for the general case of an inclined, eccentric binary source. For the purposes of this paper, we set the inclination and eccentricity to zero. Subsequent papers will tackle the modifications to the physics introduced by these additional parameters both in the analytics and the numerics.

The Hamiltonian, $\bar{\mathcal{H}}$, for a test particle evolving under these potentials takes the form

$$\bar{\mathcal{H}} = \frac{1}{2}\left(p_r^2 + \frac{p_\phi^2}{r^2}\right) + \Phi_{\rm M} + \Phi_{\rm P}, \tag{12}$$

where p_r and p_ϕ are the canonical momenta corresponding to the position variables R and φ, respectively. We then perform a canonical transformation to a frame of reference still centered on the center of mass, but now corotating with the binary. The transformed coordinates are related to the original coordinates as

$$r \to r; \quad \phi \to \varphi = \phi - t. \tag{13}$$

To obtain the corresponding transformation for the momenta, we invoke a generating function of the second kind

$$\begin{aligned} F_2 &= r p_r + (\phi - t) p_\phi, \\ (r, \varphi) &= \left(\frac{\partial F_2}{\partial p_r}, \frac{\partial F_2}{\partial p_\phi}\right). \end{aligned} \tag{14}$$

We can see that the transformed momenta are identical since

$$p_r = \frac{\partial F_2}{\partial r} = p_r, \tag{15}$$

$$p_\varphi = \frac{\partial F_2}{\partial \phi} = p_\phi. \tag{16}$$

Finally, the new Hamiltonian is given by

$$\mathcal{H} = \bar{\mathcal{H}} + \frac{\partial F_2}{\partial t}. \tag{17}$$

Thus, the final Hamiltonian, after plugging in the value of the potentials in this frame, is given by

$$\mathcal{H} = \frac{1}{2}\left(p_r^2 + \frac{p_\varphi^2}{r^2}\right) - p_\varphi - \frac{1}{r} - \sum_{m=2}^{\infty} \Phi_m \cos(m\varphi). \tag{18}$$

Before we solve the corresponding Hamilton's equations perturbatively, we state the full equations of motion of this Hamiltonian system as

$$\dot{r} = p_r, \tag{19}$$

$$\dot{\varphi} = \frac{p_\varphi}{r^2} - 1, \tag{20}$$

$$\dot{p}_r = \frac{p_\varphi^2}{r^3} - \frac{1}{r^2} - \Phi'_m \cos(m\phi), \tag{21}$$

$$\dot{p}_\varphi = m\Phi_m \sin(m\varphi). \tag{22}$$

For brevity, we use the notation $\dot{x}$ to denote the time derivative of x, Φ'_m to denote the derivative of the potential mode Φ_m with respect to the radial position r. Unless stated otherwise, there is an implied summation over the Fourier mode label m wherever Φ_m appears We assume a Keplerian rotation profile instead of including the axisymmetric components of the higher multipoles because they represent negligible $\mathcal{O}(r^{-l})$ corrections to the Keplerian rotation profile and as a result, to the locations of the resonances and other results. It should be noted that while the corrections become non-negligible at radii closer to the binary separation, the perturbation theory treatment breaks

down in those regions and the dynamics are no longer linear. The perturbation theory treatment follows from the derivation of Lindblad resonances in Binney & Tremaine (2008).

We consider perturbations about a circular orbit at $r = r_0$, the background solution for which is given by

$$\begin{aligned} r &= r_0, \\ p_r &= 0, \\ \varphi &= \varphi_0 = \omega t, \\ p_\varphi &= l_0 = \sqrt{r_0}, \\ \omega &= \sqrt{\frac{1}{r_0^3}} - 1. \end{aligned} \tag{23}$$

The variable ω represents the angular frequency of the circular orbit at r_0 in the corotating frame. The reason for choosing circular orbits is due to the fact that previous numerical studies to find stable test particle orbits under the binary potential have returned either circular or near-circular orbits with a maximum eccentricity, e, of $\mathcal{O}(10^{-1})$ (see Table 1 in Rudak & Paczynski 1981 and Table 4 of Pichardo et al. 2005). Such small eccentricities can be attributed to the apparent eccentricity introduced by the perturbative effects of the potential Φ_P.

We now expand on this circular orbit by introducing new variables, given in terms of the canonical coordinates as $(r, p_r, \varphi, p_\varphi) \rightarrow (r_0, 0, \omega t, l_0) + (r_1, p_1, \varphi_1, l_1)$. The subscript "1" indicates that these variables are first-order perturbations to the background results, the same order at which the perturbing potential enters the equations of motion. After plugging the new variables into Equation (19) and canceling the background terms, we combine the radial position and momentum equations into one second-order differential equation. As we will see in Section 4, the only place where φ is needed is in building the stability matrix. However, φ only enters inside the cosines of Φ_P, which are already first-order perturbations. Thus, we can simplify $\varphi = \varphi_0$ and ignore the evolution of φ_1. The equations of motion then reduce to

$$\dot{l}_1 = m\Phi_m \sin(m\omega t), \tag{24}$$

$$\ddot{r}_1 + (1+\omega)^2 r_1 = \left(-\Phi'_m - \frac{2(1+\omega)}{\omega}\frac{\Phi_m}{r_0}\right)\cos(m\omega t). \tag{25}$$

We solve for the angular momentum, l_1, by integrating the first equation to give

$$l_1 = -\frac{\Phi_m}{\omega}\cos(m\omega t). \tag{26}$$

The equation of motion for r_1 resembles a driven harmonic oscillator and has the standard solution

$$r_1 = -\frac{\cos(m\omega t)}{|(1+\omega)^2 - m^2\omega^2|}\left(\Phi'_m + \frac{2(1+\omega)}{\omega}\frac{\Phi_m}{r_0}\right). \tag{27}$$

We have thus obtained an analytical solution for the radial and angular momentum evolution in the epicyclic approximation. Since r_1 oscillates at the same period as the circular orbit, it is called an epicycle. An important result is the amplitude of the epicycles

$$A(r_0) = \frac{1}{|(1+\omega)^2 - m^2\omega^2|}\left|\Phi'_m + \frac{2(1+\omega)}{\omega}\frac{\Phi_m}{r_0}\right| \tag{28}$$

given as a function of the background radius r_0. We note immediately that we have *Lindblad resonances* whenever the denominator $(1+\omega)^2 - m^2\omega^2$ vanishes and the amplitude diverges as a result. While these resonances are located at discrete radii, the breakdown of the epicyclic approximation happens over a wider region surrounding these resonances. We shall study the nature of this breakdown and its implications for orbital stability in further sections.

3. Numerical Calculations

3.1. Numerical Setup and Data Analysis

The numerical code DISCO (Duffell 2016) is used here as a moving mesh, two-dimensional Harten-Lax-van Leer contact solver (Toro 2013) for Newtonian hydrodynamic field equations. Our numerical domain is a flat ring-shaped surface that is centered on the center of mass of the binary and spans in the radial direction from a to $100a$. The central circle-shaped region of radius a where the binary is located is excluded from the computational domain. The numerical grid is composed of 480 concentric radial rings of varying widths. The width of the nth ring is calculated as the difference between two neighboring nodes $\Delta r_n = r_{n+1} - r_n$, where

$$r_n = 1 + A\sinh(BC), \tag{29}$$

where

$$A = \frac{5}{\sinh(1)}, \quad B = \frac{n-1}{480}, \quad C = \sinh^{-1}(99/A). \tag{30}$$

The size of the innermost ring is $\Delta r_1 \approx 0.034$, and the outermost ring is $\Delta r_{480} \approx 0.790$. Each radial ring is composed of a different number of azimuthal zones, which are refined or de-refined as needed to keep the aspect ratio $r\Delta\phi/\Delta r$ close to unity. According to Dong et al. (2011) and Duffell & MacFadyen (2012), this choice allows the density waves excited by the binary potential to be well resolved over their wavelength $\sim 0.2\pi r$. The time step Δt is set to be half of the shortest propagation time across any cell in the grid, i.e., the width of the cell divided by the local sound speed. The mesh rotates according to $\Omega_{\rm grid} = r^{-3/2}$ to minimize diffusive advection errors.

For simplicity, we express all timescales in terms of the binary revolution period P. We add that, in our units, $P = 2\pi$. To present the data as a simple function of the distance from the rotation axis, we use two different kinds of averaging. For any quantity $\xi(r, \phi, t)$ we introduce the following notation:

$$\bar{\xi}(r, t) = \frac{1}{2\pi}\int_0^{2\pi} \xi(r, \phi, t)\, d\phi, \tag{31}$$

$$\hat{\xi}(r, \phi) = \frac{1}{\Delta}\int_t^{t+\Delta} \xi(r, \phi, t)\, dt, \tag{32}$$

and finally

$$\langle\xi\rangle(r) = \frac{1}{2\pi\Delta}\int_t^{t+\Delta}\int_0^{2\pi} \xi(r, \phi, t)\, d\phi\, dt. \tag{33}$$

In our simulations we use $\Delta \equiv 50P$. We dump grid data every 6° of the binary revolution, so the time averages over the interval Δ are calculated from 3000 instantaneous snapshots of the grid values.

3.2. Initial Data and Boundary Conditions

We start the simulations with the initial density

$$\sigma(r, t=0) = \sigma_0 \left(\frac{r_s}{r}\right)^3 \exp\left[-\left(\frac{r_s}{r}\right)^2\right], \tag{34}$$

where σ_0 is an arbitrary normalization constant, and r_s is a constant that sets the radial scale of σ. The function arguments $(r, t=0)$ are implied throughout the rest of this section. We set $r_s \equiv 10$, so that the maximum initial value of σ occurs at $r = 10\sqrt{2/3} \approx 8.16$. The initial pressure is calculated from the density using the equation of state, Equation (6). The initial velocity distribution is given by MacFadyen & Milosavljević (2008) as

$$\boldsymbol{v} = (v^r, v^\phi) = \left\{\frac{2}{r\sigma v^\phi}\frac{\partial}{\partial r}\left(r^3\sigma\nu\frac{\partial}{\partial r}\frac{v^\phi}{r}\right),\right.$$
$$\left.\left[\Omega_K^2\left(1 + \frac{3}{4}\frac{a^2}{r^2}\frac{q}{(1+q)^2}\right)^2 + \frac{1}{r\sigma}\frac{dp}{dr}\right]^{1/2}\right\}, \tag{35}$$

where $q \equiv m_1/m_2$ is the *mass ratio*. We calculate v^ϕ first and use it to calculate v^r, as described in MacFadyen & Milosavljević (2008).

Boundary conditions are imposed on the primitive variables at both the inner and the outer edges of the grid. The outer boundary conditions for all primitive variables are fixed at the initial data values for the two outermost rings. The same procedure is applied to the components of the fluid velocity at the inner boundary. The exceptions are the density and pressure boundaries, which instead are given by

$$\sigma_i = \bar{\sigma}_{i+1}(r_{i+1}/r_i)^{-1/2}, \tag{36}$$

$$p_i = \bar{p}_{i+1}(r_{i+1}/r_i)^{-1/4}, \tag{37}$$

where $i \in \{1, 2\}$ corresponds to the number of each of the two innermost radial rings. We note that these boundary conditions do not prevent the flow of matter into or out of the numerical domain. However, the central excised region does not contain any significant density throughout the entirety of the simulations. We have investigated the accretion rate through the grid boundaries and the total mass of the system as a function of time, and confirmed that the flow of matter across the boundaries does not impact our results.

3.3. Mass Density and Torque Distributions

We have simulated the different binary mass ratios q and viscosity parameters α given in Table 1 until all of the systems settled into a quasi-steady state. That is, until the averaged values (in the sense of Equation (33)) of the hydrodynamic variables are no longer evolving secularly. Unless stated otherwise, all figures present data for $\alpha = 0.01$. Figure 1 presents the mass density in the central region of two of these systems. In the left panel, we present the equal mass ($q = 1$) case, and on the right, the $q = 1/10$ case.

Figure 2 presents the azimuthally averaged surface mass density $\bar{\sigma}$ for $q \in \{0, 1/100, 1/10, 1\}$ at $t = 4000P$. All cases have smaller density maxima than the initial peak density due to mass loss through the outer boundary. These cases also show varying degrees of periodic structure related to the binary's periodic gravitational potential, which is obviously not present in the single mass ($q = 0$) case. Figure 3 presents the averaged mass density $\langle\sigma\rangle$, calculated using $\Delta = 50P$, as a function of radius for $q \in \{0, 1/100, 1/10, 1/4, 3/7, 2/3, 1\}$, i.e., for all of the cases we simulated and summarized in Table 1. We emphasize that the central part of the system, which is most affected by the periodic binary potential, stabilizes between $3000P$ and $4000P$ for all cases. Our conclusions about the governing dynamics and our description of the final gap size are based on simulations as long as $12000P$, to ensure that the observed quasi-steady state is maintained. The location of the density maximum moves inward as the mass ratio decreases from 1 to 1/100, but moves outward again for the single mass case. However, Figure 4, which presents four different stages of the systems' evolution at $t = 1000P$, $t = 2000P$, $t = 3000P$, and $t = 4000P$, shows that the secular trend of an inward migration of the gap continues all the way from the equal mass case through the single mass case.

Table 1
Summary of All Simulated Configurations

q	$\mu = m_2/M$	α	t_*	r_{dT}	r_{max}	$r_{10\%}$
1:1	0.5	0.1	400	1.818	3.333	1.655
2:3	0.4	0.1	400	1.820	3.295	1.641
3:7	0.3	0.1	400	1.800	3.103	1.606
1:4	0.2	0.1	400	1.713	2.952	1.473
1:10	0.(09)	0.1	400	1.561	2.766	1.322
1:100	0.(0099)	0.1	400	1.139	4.059	1.133
1:1	0.5	0.03	1250	1.975	3.893	2.012
2:3	0.4	0.03	1250	1.970	3.893	1.999
3:7	0.3	0.03	1250	1.948	3.934	1.946
1:4	0.2	0.03	1250	1.898	3.609	1.830
1:10	0.(09)	0.03	1250	1.725	3.256	1.571
1:100	0.(0099)	0.03	1250	1.251	4.101	1.232
1:1	0.5	0.01	4000	2.079	4.757	2.307
2:3	0.4	0.01	4000	2.074	4.711	2.296
3:7	0.3	0.01	4000	2.052	4.532	2.242
1:4	0.2	0.01	4000	1.993	4.228	2.146
1:10	0.(09)	0.01	4000	1.862	3.411	1.828
1:100	0.(0099)	0.01	4000	1.402	3.103	1.285
0	0	0.01	⋯	⋯	4.357	1.207
1:1	0.5	0.003	9000	2.148	5.033	2.603
2:3	0.4	0.003	9000	2.151	4.987	2.557
3:7	0.3	0.003	9000	2.127	4.757	2.470
1:4	0.2	0.003	9000	2.074	4.532	2.388
1:10	0.(09)	0.003	9000	1.957	3.811	2.075
1:100	0.(0099)	0.003	9000	1.518	2.620	1.388
1:1	0.5	0.001	24,000	2.190	5.176	2.800
2:3	0.4	0.001	24,000	2.201	5.128	2.798
3:7	0.3	0.001	24,000	2.206	4.940	2.726
1:4	0.2	0.001	24,000	2.129	4.757	2.636
1:10	0.(09)	0.001	24,000	2.040	4.143	2.297
1:100	0.(0099)	0.001	24,000	1.600	2.915	1.508

Note. Columns label the mass ratio q, secondary mass ratio μ, viscosity α, time at which we report results (i.e., after the system reaches a quasi-steady state) t_*, radius where the torque densities balance r_{dT}, radius of the maximum density r_{max}, and radius where the density inside the cavity falls to 10% of the final density maximum $r_{10\%}$.

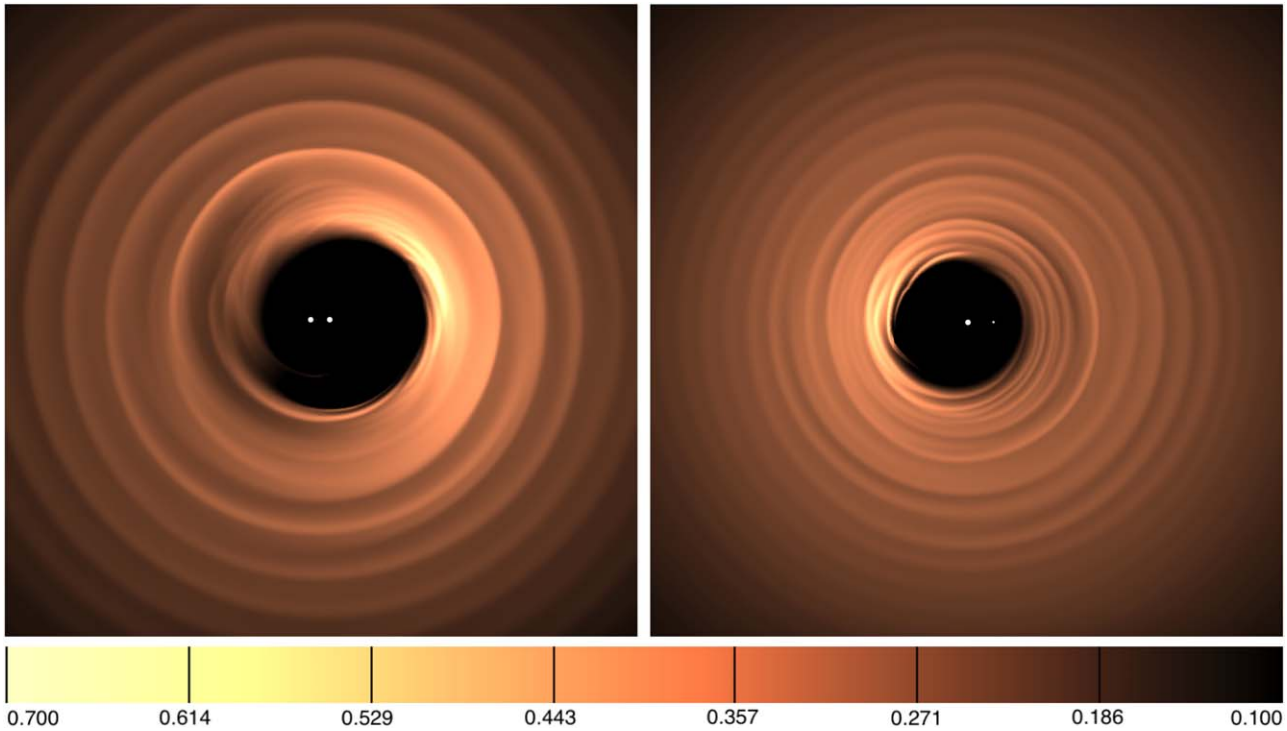

Figure 1. Surface mass density σ/σ_0 at $t = 4000P$ for $q = 1$ (left) and $q = 1/10$ case (right). The pictures show the region $-15r/a \leqslant (x, y) \leqslant 15r/a$.

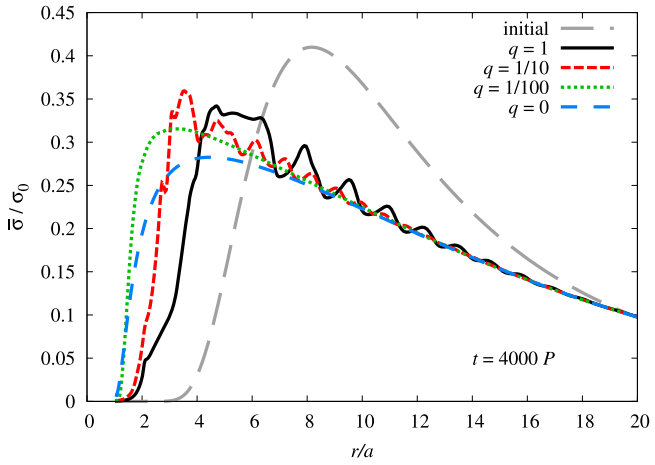

Figure 2. Azimuthally averaged surface mass density, at $t = 4000P$ for $q \in \{0, 1/100, 1/10, 1\}$ together with the initial density for comparison.

Following MacFadyen & Milosavljević (2008) and D'Orazio et al. (2013), we calculate the dynamical torque density

$$\frac{dT_{\rm d}}{dr} = -2\pi r \left\langle \sigma \frac{d\Phi}{d\phi} \right\rangle, \tag{38}$$

the integrated dynamical torque

$$T_{\rm d} = \int_a^r \frac{dT_{\rm d}}{dr} dr, \tag{39}$$

the viscous torque

$$T_\nu = 2\pi r^3 \nu \left\langle \sigma \frac{\partial}{\partial r} \frac{v^\phi}{r} \right\rangle, \tag{40}$$

and the viscous torque density $\frac{dT_\nu}{dr}$.

Figure 5 shows the final density profile overall in comparison to the initial data for all of the simulated cases (left panel), as well as a zoom-in on the innermost part of the system (right panel) where all of the potential definitions for the location of a gap occur. Figure 5 specifically illustrates $r_{\rm dT}$—the radius where the viscous torque density equals the dynamical torque density. This concept is further highlighted in Figure 6, which shows both types of torque (viscous and dynamical) and their densities as a function of radius. Vertical dashed lines are placed at the radii where the two different types of torque (densities) intersect.

In Table 1, in addition to the mass ratio q and the secondary mass ratio μ, we have listed the following system characteristics for all of our simulated cases: the radial location of the maximum density, $r_{\rm max}$, and the maximum density value $\rho(r_{\rm max})$; the radius where the torque densities balance, $r_{\rm dT}$, and corresponding torque density value, $\rho(r_{\rm dT})$; and $r_{10\%}$, the radius where the density inside the cavity falls to 10% of the final density maximum. This last quantity is introduced as an additional intuitive way to define a gap. The specific choice of 10% is essentially arbitrary. We note that $r_{10\%}$ is found by interpolation, and is not therefore restricted to occur at a cell boundary. To show their general dependence on the mass ratio, all of these radii are plotted in the left panel of Figure 7. The right panel of Figure 7 shows the stability of the $r_{\rm dT}$ with time; its value is established very early, by the time t reaches $1000P$—$2000P$, while the other referenced radii (especially $r_{\rm max}$) are still actively evolving toward their quasi-steady states.

3.4. Azimuthal Mode Analysis

An important aim of this study is to better understand the nature of angular momentum transport in circumbinary disks as a way to interpret astrophysical disk observations and what they can tell us about the underlying dynamics. A frequent primary assumption in previous analytic studies of these disks

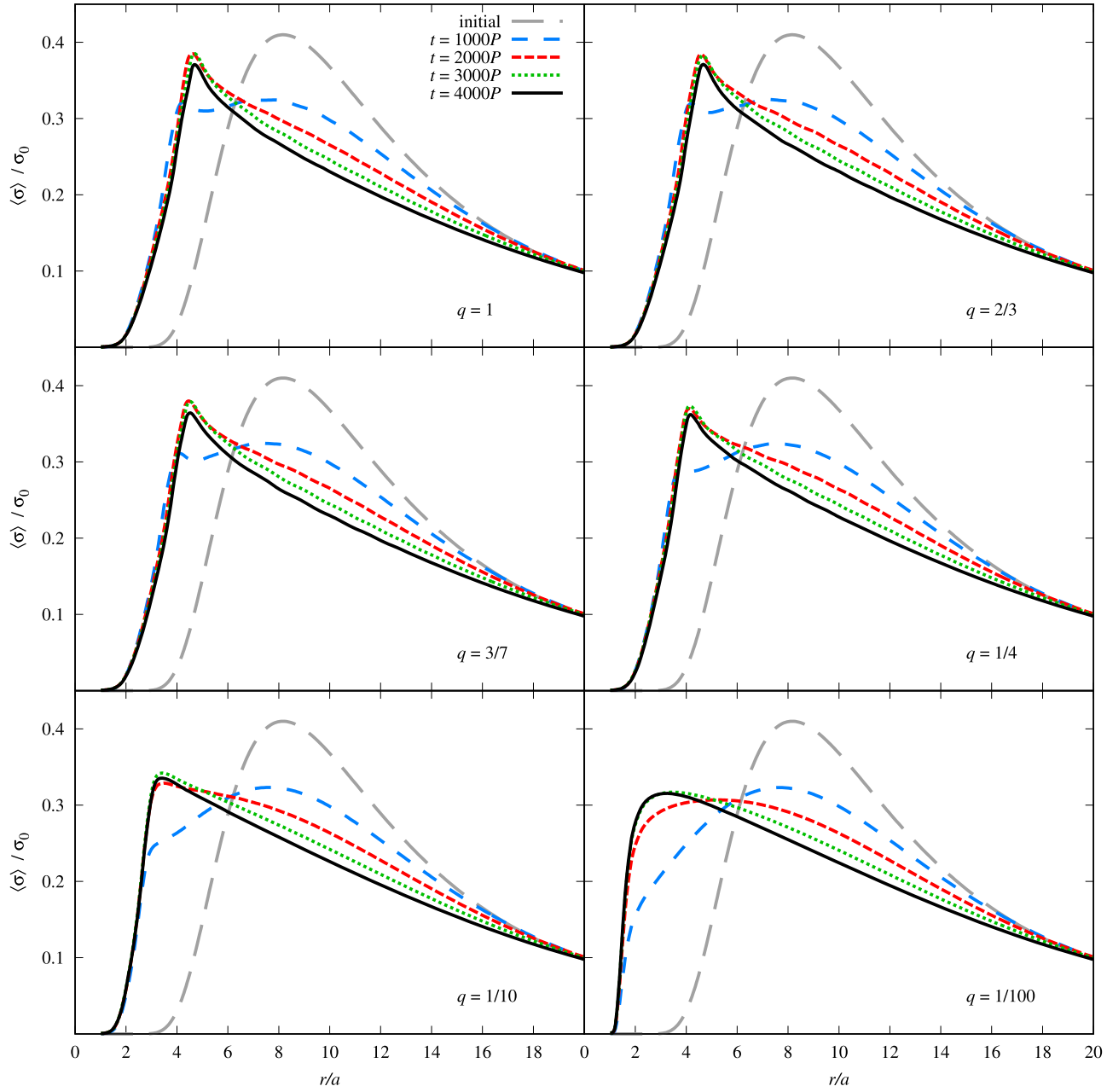

Figure 3. Averaged surface mass density initially and at four different times $t \in \{1000P, 2000P, 3000P, 4000P\}$, with each panel considering one of the mass ratios summarized in Table 1.

has been the perturbative treatment of azimuthal symmetry breaking due to the binary potential. In order to provide a measure of accuracy for such a perturbative treatment, and to better assess the importance of the *resonant* transfer of angular momentum, we compute the corresponding azimuthal breakdown of the density that indicates the presence or absence of a perturbative effect. In these calculations, the density is decomposed into azimuthal harmonics in an identical fashion to the gravitational potential in Equation (11). This decomposition is given by

$$D_m = \frac{1}{2\pi^2}\int_0^{2\pi} d\phi \int_0^{2\pi} d(\Omega_b t)\,\rho e^{im(\phi-\Omega_b t)}, \tag{41}$$

where Ω_b is the binary's orbital angular frequency.

In Figure 8, we present the decomposition of the density distribution into azimuthal harmonic modes for all of our simulated cases. In the perturbative picture presented in previous studies (Goldreich & Tremaine 1980; Artymowicz & Lubow 1994), these modes are driven by the corresponding harmonic mode of the gravitational potential. Since each harmonic mode represents a contribution from a higher multipole moment, we expect the strength of the driving force to weaken with higher mode number m, and for odd harmonic modes to weaken as we approach the equal mass due to the absence of odd-m potential modes in that case, especially the case of equal mass where we expect the $m = 1$ mode to vanish. However, as illustrated in Figure 8, we do not see an unambiguous decay with increasing mode number beyond $m = 1$ at any length scale across the next three m modes, with the possible exception of the $q = 1/10$ and $1/100$ cases at large radii. We also note the dominance of the $m = 1$ mode everywhere except the $q = 1/100$ case at small radii,

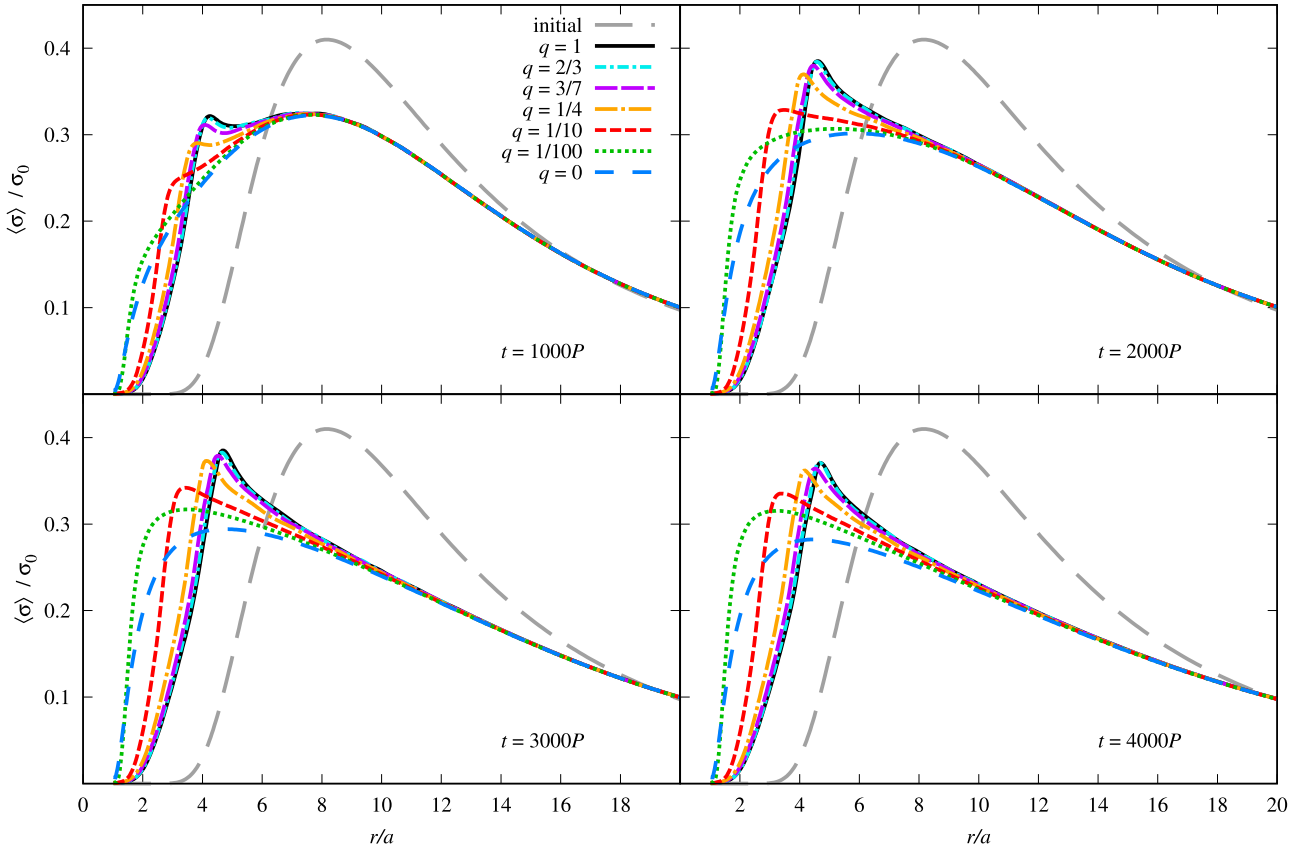

Figure 4. Averaged surface mass density for all of the mass ratios summarized in Table 1, with each panel considering one of the evolution times $t \in \{1000P, 2000P, 3000P, 4000P\}$. The initial density is also shown in each panel for reference.

particularly its dominance for the equal mass case, where the contribution from any odd-m mode of the gravitational potential vanishes. In the case of a linear coupling of the density modes, which is attributed to a resonant gap clearing, we would ideally see the $m = 1$ mode diminish significantly toward equal mass and dominance of the $m = 2$ mode. Additionally, the amplitude of each mode would drop as higher power laws beyond the inner region. The lack of these features suggests a departure from a linear coupling of the potential and density modes in the dynamics. This absence strongly suggests the dynamical irrelevance of resonant torquing in this scenario. In our orbital stability picture of gap opening, we sum over multiple modes of the gravitational potential and do not assert that any particular harmonic mode is responsible for the opening of the gap as is the case in resonant torquing. We will discuss the consequences of this and make suggestions for a better analytic treatment of circumbinary disk dynamics in the next section.

4. Analytical Calculations

4.1. Orbital Stability

In order to reconcile the results from the hydrodynamics simulations with analytic theory, and thereby obtain a coherent understanding of circumbinary disk behavior, it is important to look at the stability of the approximate solutions derived in Section 2.2. We note that in order to translate from the hydrodynamic picture to the test particle picture, the disk is considered to be a collection of fluid elements which each occupy an infinitesimal area centered about a hypothetical test particle. Then, the orbit of the test particle acts as an approximation for the motion of this fluid element in the absence of viscosity. In that case, the exponential blow-up of small perturbations about the test particle's orbit, especially over timescales comparable to or shorter than the binary orbital period, provides an important insight into the effect of the binary potential on the stability of the fluid elements themselves. In other words, instability in the test particle picture suggests that the corresponding fluid elements in the hydrodynamic picture will either be expelled to larger radii or accreted by the binary.

The physical quantity relevant to orbital stability is the Lyapunov exponent. For a Hamiltonian system with phase space variables $\eta = (r, \phi, p_r, p_\phi)$, the equation of motion is given in the symplectic form as (see, e.g., Equation (8.39) of Goldstein et al. 2002)

$$\dot{\eta}_i = \bar{I}_{ij} \frac{\partial \mathcal{H}}{\partial \eta_j}, \tag{42}$$

where

$$\mathbb{I} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{43}$$

maps the derivatives of the Hamiltonian to the appropriate time derivative with the appropriate sign. Let us assume there exists a particular solution, $\eta^{(0)}$, to the above equations of motion. Since a particular solution can be entirely specified by a choice

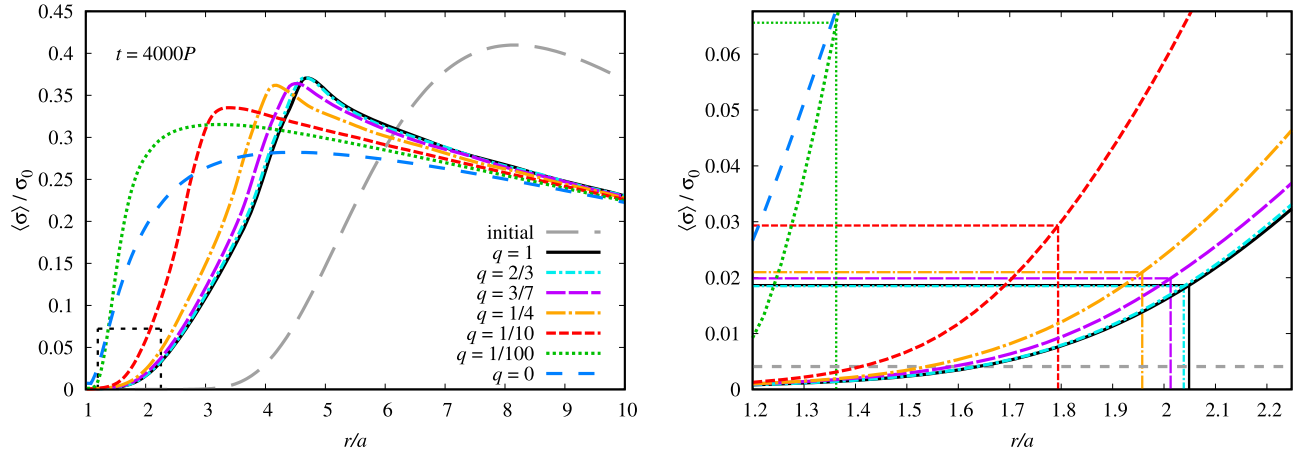

Figure 5. Averaged surface mass density at $t = 4000P$ for all cases in Table 1, along with the initial density. The right panel is a zoom-in of the region inside the dashed box in the left panel. The right panel also shows $r_{\rm dT}$ and $\rho(r_{\rm dT})$ as we have defined them in the text.

of initial conditions, the stability of that solution can be defined as the extent to which a new solution, $\eta^{(1)}$, with approximately similar initial conditions, departs from the original solution. To quantify this, let us assume that $\eta^{(1)}{=}\eta^{(0)} + \delta\eta$. Here, $\delta\eta$ represents the difference between the two solutions with $\delta\eta$ $(t = 0)$ being infinitesimally small. We can then linearize the above equations of motion to get

$$\delta\dot{\eta}_i = \bar{I}_{ij}\frac{\partial^2\mathcal{H}}{\partial\eta_j\partial\eta_k}\bigg|_{\eta^{(0)}}\delta\eta_k. \tag{44}$$

Thus, the departure of the new solution due to a small initial perturbation $\delta\eta(0)$ is given by the matrix equation

$$\delta\dot{\boldsymbol{\eta}} = \mathbb{K}\delta\boldsymbol{\eta}, \tag{45}$$

$$K_{ij} = \bar{I}_{im}\frac{\partial^2\mathcal{H}}{\partial\eta_m\partial\eta_j}\bigg|_{\eta^{(0)}}. \tag{46}$$

The above system of differential equations has a general solution given by a superposition of the eigenvectors of $\mathbb{K}$ which evolve exponentially at a rate given by the corresponding eigenvalues, λ, which are known as the Lyapunov exponents. Since $\mathbb{K}$ will in general have complex values, the real part, $\mathfrak{Re}(\lambda)$, gives the timescale over which the perturbations decay (if $\mathfrak{Re}(\lambda) < 0$), or blow up (if $\mathfrak{Re}(\lambda) > 0$); the imaginary part, $\mathfrak{Im}(\lambda)$, gives the period over which the perturbations oscillate. We note that the eigenvectors of the stability matrix represent the direction in phase space of the initial perturbations associated with each of these exponents. A particular orbit may be metastable, which indicates that one set of perturbations tends to decay while an orthogonal set may blow up. It is useful to study the stability of particular "families" of orbits, as we do with the epicyclic orbits in this study.

4.1.1. Stability Matrix for Epicyclic Orbits

Using the tools from Section 4.1, we now study the stability of circumbinary orbits under the epicyclic approximation. Since we will solve the equations perturbatively, we approximate the stability matrix to the same order as the linearized solution for consistency. The Hamiltonian presented in Section 2.2 is given by

$$\mathcal{H} = \frac{p_r^2}{2} + \frac{p_\varphi^2}{2r^2} - p_\varphi - \frac{1}{r} + \epsilon\,\Phi_m(r)\cos(m\varphi). \tag{47}$$

Since the potential Φ_m contains terms of $\mathcal{O}(q)$ and higher, it must be treated as a perturbation to the Kepler potential. To this end, we introduce the parameter ϵ that keeps track of the perturbation terms. Once we have separated the perturbation terms, we can set $\epsilon = 1$ without loss of generality. For the above Hamiltonian, the stability matrix, $\mathbb{K}$, takes the form

$$\mathbb{K}= \begin{pmatrix} 0 & 0 & 1 & 0 \\ -\dfrac{2l}{r^3} & 0 & 0 & \dfrac{1}{r^2} \\ -\dfrac{3l^2}{r^4} + \dfrac{2}{r^3} - \epsilon\,\Phi''_m\cos(m\varphi) & \epsilon\,\Phi'_m\cos(m\varphi) & 0 & \dfrac{2l}{r^3} \\ m\epsilon\,\Phi'_m\sin(m\varphi) & m^2\epsilon\,\Phi_m\cos(m\varphi) & 0 & 0 \end{pmatrix}. \tag{48}$$

We evaluate Equation (48) for the approximate orbital solution

$$r = r_0 + \epsilon\, r_1, \tag{49}$$

$$p = \epsilon\frac{r_1}{\omega}, \tag{50}$$

$$l = \sqrt{r_0} + \epsilon\, l_1. \tag{51}$$

The first-order solutions to the above expressions can be found in Equations (26) and (27). We note again that we do not need a first-order solution for φ since it enters the matrix at $\mathcal{O}(\epsilon)$. Thus, we can use

$$\varphi = \omega t,$$
$$\omega = \sqrt{\frac{1}{r_0^3}} - 1. \tag{52}$$

The stability matrix can then be split as

$$\mathbb{K} = \mathbb{K}_0 + \epsilon\delta K, \tag{53}$$

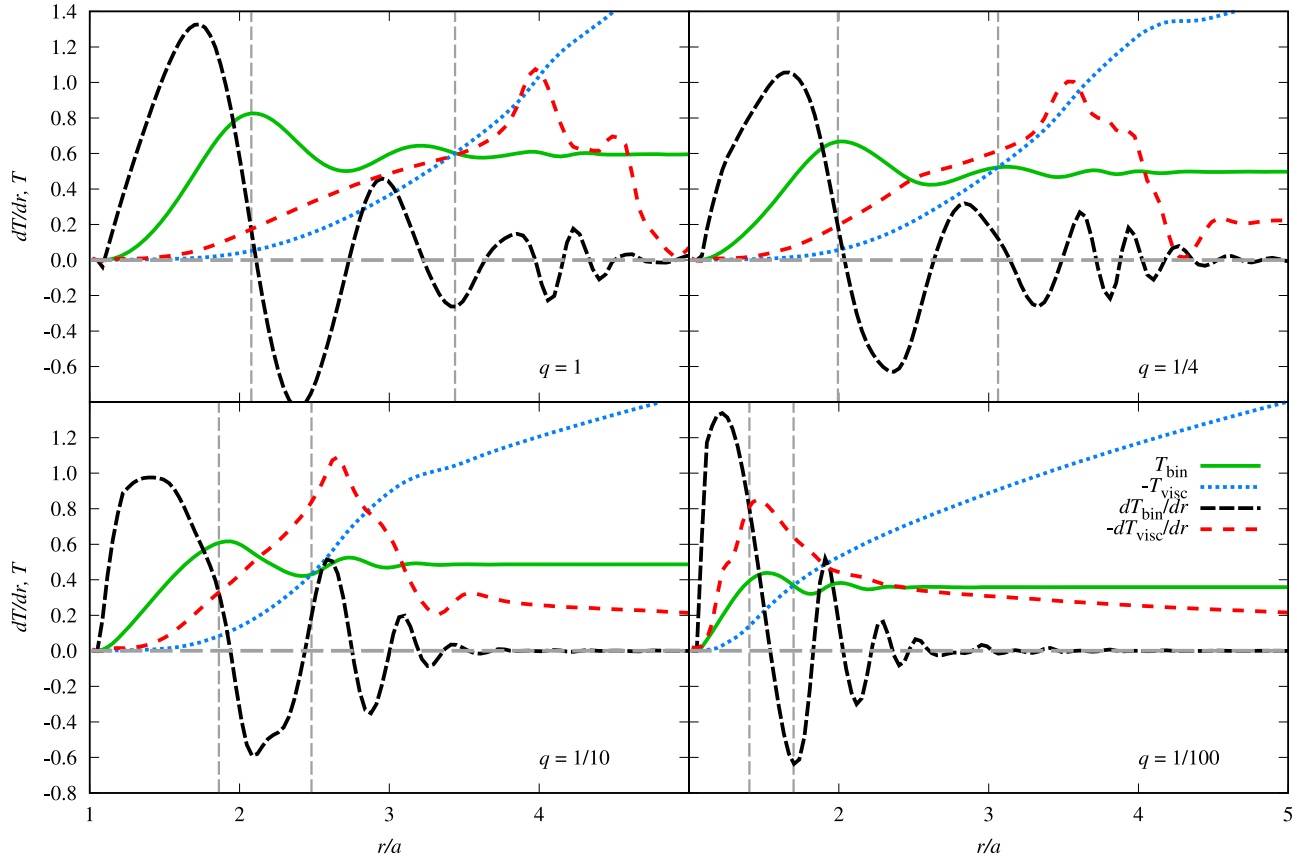

Figure 6. Four different types of torques for four different mass ratios. Solid lines correspond to torques, broken lines correspond to torque densities. Red lines correspond to dynamical torques and green lines correspond to viscous torques. Vertical broken lines help to read the intersection radii.

where $\mathbb{K}_0$ comes from the Keplerian background motion and $\delta\mathbb{K}$ contains all $\mathcal{O}(\epsilon)$ terms. $\mathbb{K}_0$ is given by

$$\mathbb{K}_0 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ -\frac{2l_0}{r_0^3} & 0 & 0 & \frac{1}{r_0^2} \\ -\frac{1}{r_0^3} & 0 & 0 & \frac{2l_0}{r_0^3} \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{54}$$

and the perturbative matrix $\delta\mathbb{K}$ is

$$\delta\mathbb{K} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -2\left(\frac{l_1}{r_0^3} + \frac{3l_0 r_1}{r_0^4}\right) & 0 & 0 & -2\frac{r_1}{r_0^3} \\ 6\left(\frac{r_1}{r_0} - \frac{l_0 l_1}{r_0^4}\right) - \Phi_m'' \cos(m\phi) & m\Phi_m' \sin(m\phi) & 0 & 2\left(\frac{l_1}{r_0^3} - \frac{3l_0 r_1}{r_0^4}\right) \\ m\Phi_m' \sin(m\phi) & m^2\Phi_m \cos(m\phi) & 0 & 0 \end{pmatrix}. \tag{55}$$

4.1.2. Lyapunov Exponents for Epicyclic Orbits

The computation of closed-form eigenvalues for the stability matrix given in Equation (53) at linear order is complicated by the fact that regular matrix perturbation theory does not hold for degenerate matrices. Since the Keplerian stability matrix, Equation (54), has a doubly degenerate, trivial eigenvalue of zero, it is not invertible, and therefore we cannot proceed to solve the eigenvalue problem analytically; we therefore use the Python package SCIPY (Virtanen et al. 2020) to numerically compute the eigenvalues.

We compute the eigenvalues for a family of epicyclic orbits around the background radius r_0 in the range $r_0 \in [1.1, 3.0]$, as this covers the relevant resonances and the typical gap scale, which we will define below. For the binary mass ratio, we chose the range $\mu \in [0.01, 0.5]$, as this is the range probed by the numerical studies.

We separate the real part of the Lyapunov exponents, and search for the largest positive value at each radius. Since the real part is the inverse timescale over which instabilities propagate, the largest value corresponds to the shortest timescale, which we will refer to as the Lyapunov timescale. To define a gap size r_L based on the Lyapunov timescale τ, we find the radius that satisfies

$$\tau(r_L) \equiv [\mathfrak{Re}(\lambda(r_L))_{\max}]^{-1} = P, \tag{56}$$

where, $\lambda(r_L)$ denotes the Lyapunov exponent at r_L, "max" indicates that we are choosing the Lyapunov exponent with the largest real part, and $P = 1$ in our units. In Figure 9, we present the Lyapunov timescales for a range of systems. We note that there are regions where the timescale τ drops drastically. We confirm that these are the locations of the Lindblad Resonances where $\tau \to 0$ or, more importantly, $\lambda \to \infty$. This effect only demonstrates a *complete* breakdown of the epicyclic approximation. However, as can be seen from the rest of the graph, the epicyclic approximation is unstable for radii larger than the resonances. Binney & Tremaine (2008) provide a method for

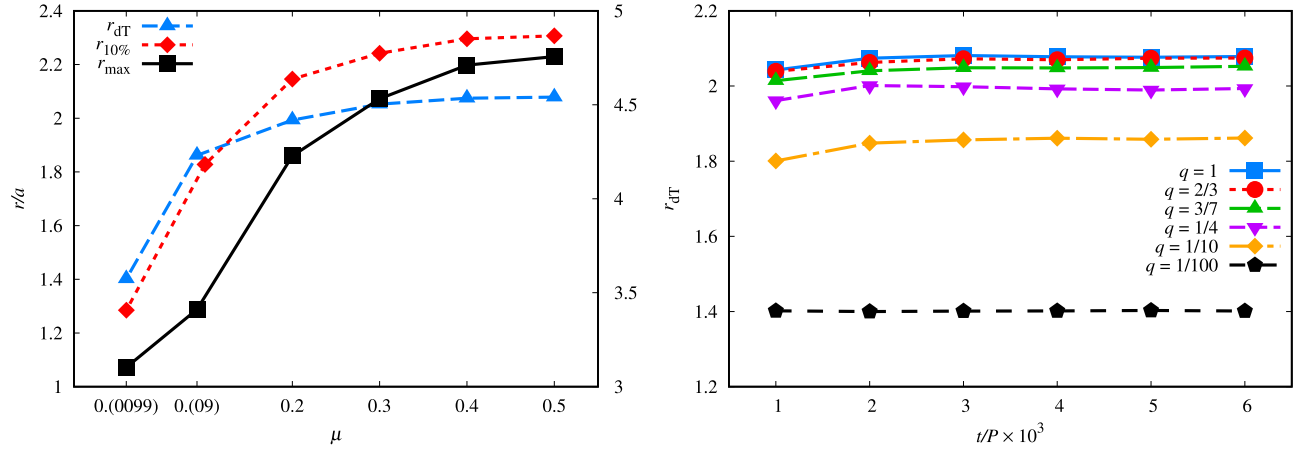

Figure 7. Three crucial radii that characterize each configuration as a function of binary mass function μ. The squares denote r_{dT}—the radii where the dynamical torque densities intersect viscous torque densities, the diamonds denote $r_{10\%}$—the radii where the disk's density reaches 10% of the maximal value. For those two, the scale on the left vertical axis is applied. Finally, triangles denote r_{max}—the location of the density maximums. In this case the right vertical axis is valid. Let us underline that the values of the densities $\sigma(r_{dT})$ are different for each configuration. The details are in Table 1.

regularizing, i.e., removing the infinite behavior at the resonances. We did not perform this regularization as it does not affect the outcome of the gap sizes, which occur at larger radii than the resonances. The rationale behind the choice $\tau = P$ in Equation (56) is that, while the binary still induces instabilities at larger radii, the Lyapunov timescale increases with radius far faster than the orbital period. We also note that the analysis of orbital stability has been done by perturbing on circular orbit solutions about the total mass while studies going back to MacFadyen & Milosavljević (2008) indicate that the circumbinary gap is eccentric. This is because the apparent eccentricity induced by the perturbing potential of the binary is sufficient to explain the eccentricity in the circumbinary gap. Additionally, any perturbative treatment of an eccentric orbit will still yield the instabilities in the circular ($\mathcal{O}(\epsilon^0)$) order that can explain the gap formation. However, a treatment of generic orbits would benefit future studies where the eccentricity and semimajor axis are dynamical due to the effect of accretion from the disk (see Siwek et al. 2023).

We finally emphasize that this definition is independent of the nature of the fluid characterization. The study of test particles indicates that instabilities propagate over timescales shorter than an orbit, and therefore far shorter than the characteristic viscous timescale as believed in Goldreich & Tremaine (1980). Therefore, if this is a viable explanation for the formation and maintenance of a circumbinary gap, then it should be independent of the viscous prescription as noted in D'Orazio et al. (2013).

4.2. Resonant Torquing Picture

Prior analytical studies of circumbinary accretion disks have arrived at a criterion for the opening of a circumbinary gap from hydrodynamic considerations (see Artymowicz & Lubow 1994; Miranda & Lai 2015; Goldreich & Tremaine 1980). Instead of studying the stability of orbits due to the gravitational potential of the binary, those works have explored the behavior of the primitive fluid variables that evolve under the influence of the binary. In that case, the equations of fluid mechanics are solved approximately using a Wentzel–Kramers–Brillouin approximation to study the behavior of the primitive variables over timescales that are larger than the binary timescale but still comparable to or shorter than the viscous timescale.

While the two pictures of test particles and hydrodynamic flows can be related by associating particle orbits with steady streamlines of the gas, they differ in the physical mechanism that is invoked to explain the opening and maintenance of a circumbinary gap. In our analysis, the opening and maintenance of the gap is contingent on the divergence of an infinitesimal element over short timescales due to the tidal nature of the binary potential; in the fluid picture, the gap is contingent on the ability of the Lindblad resonances associated with the binary potential to transfer angular momentum outward through the disk faster than viscous damping forces can dissipate it. The mechanisms primarily differ over the role of resonances in the formation of the gap and the timescales over which the binary potential maintains the gap.

Since our proposed mechanism is different from that of the hydrodynamics one, we must compare the resultant gap size in both cases. Therefore, in this subsection, we will recap some of the quantities derived in Goldreich & Tremaine (1980), proceed to define the criterion for a resonance to be gap opening, and compute gap sizes from this definition.

The gap-opening timescale is defined as the time it takes for a torque at a Lindblad resonance to deposit the angular momentum needed to open a gap in the form of an infinitesimal ring of width Δr at a radius r. To quantify this, we first to define the amount of angular momentum, ΔL, required to open such a gap in a disk with surface density σ as

$$\Delta L = \sigma\Omega(r\Delta r)^2. \tag{57}$$

If T_{LR} is the torque delivered by a Lindblad resonance, then the gap-opening timescale is given by

$$\Delta t_{open} = \frac{\Delta L}{T_{LR}}. \tag{58}$$

Goldreich & Tremaine (1980) also give the expression for the torque at the mth Lindblad resonance as

$$T_m = -m\pi^2\left[\sigma\left(\frac{dD}{d\ln r}\right)^{-1}|\Psi_m|^2\right], \tag{59}$$

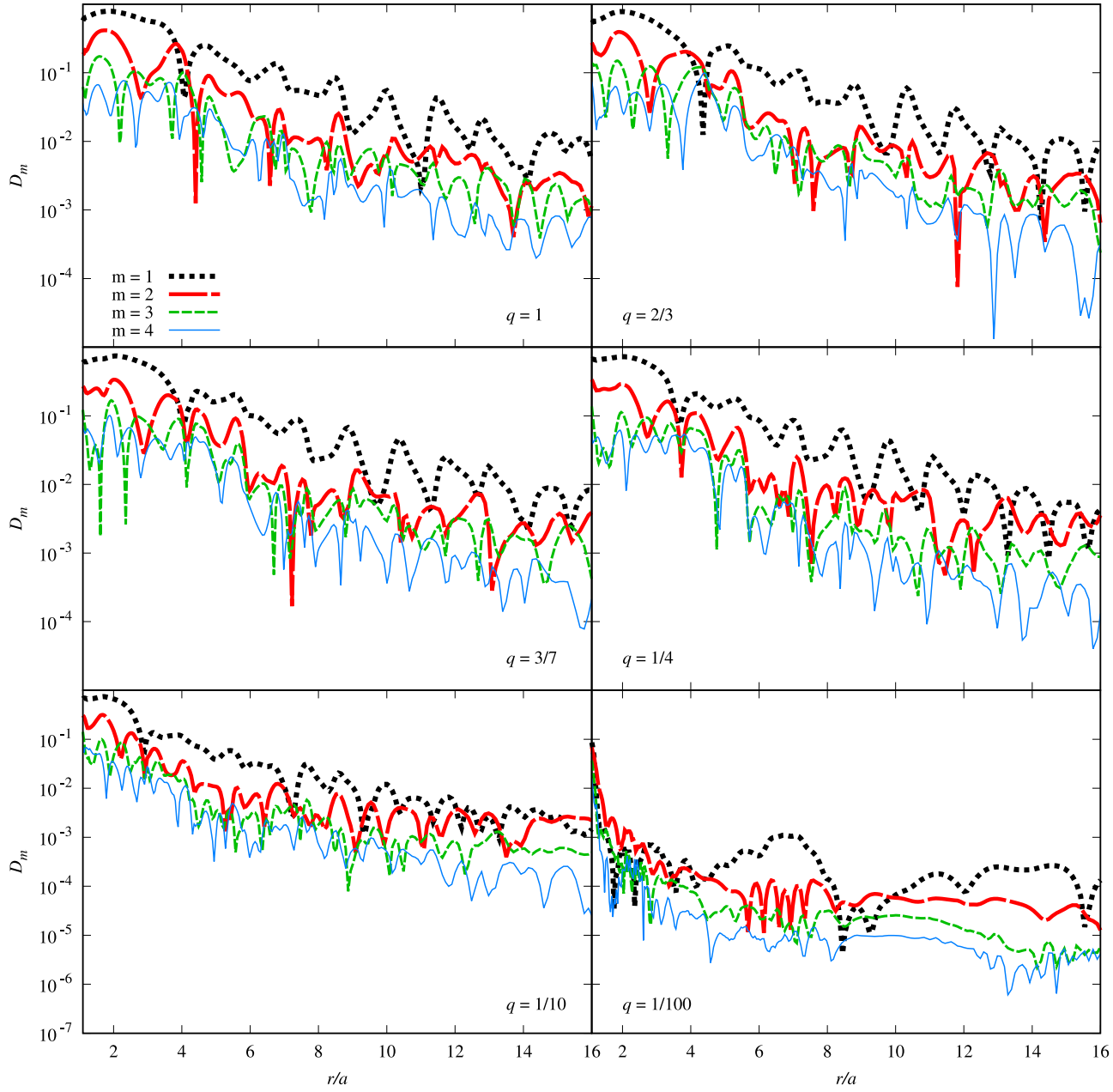

Figure 8. Azimuthal modes decomposition of the density at $t = 4000P$ for all of our simulated cases.

where Ψ_m is defined as

$$\Psi_m = \frac{d\Phi_m}{d\ln r} - 2m\Phi_m, \tag{60}$$

and D is defined in terms of the epicyclic frequency, κ, and the orbital frequency, ω, as $D \equiv \kappa^2 - m^2(\omega - 1)^2$. For the Keplerian background ($\kappa = \omega = r^{-3/2}$) considered below, $\frac{dD}{d\ln r}$ evaluated at the Lindblad resonance is then given by

$$\left.\frac{dD}{d\ln r}\right|_{\rm LR} = -\frac{3m^2}{m+1}. \tag{61}$$

To complete this picture of gap maintenance due to Lindblad resonances, we also need the gap-closing timescale, which is defined as the time it will take for viscous damping forces to close a ring-shaped gap of similarly infinitesimal radial size Δr at a radius r,

$$\Delta t_{\rm close} = \frac{(\Delta r)^2}{\nu}, \tag{62}$$

where ν is the coefficient of viscosity for the accretion disk. For the α-disk prescription, again assuming a Keplerian background, one has from Equation (5)

$$\nu = \frac{\alpha\chi^2}{r\Omega}, \tag{63}$$

where α and χ are defined in Section 2.1.

We therefore arrive at the first gap-opening criterion for this picture, which states that a resonance is considered gap opening if the gap-closing timescale exceeds the gap-opening timescale. The gap-opening resonance is the outermost Lindblad resonance where the criterion $\Delta t_{\rm close} > \Delta t_{\rm open}$ is met, or

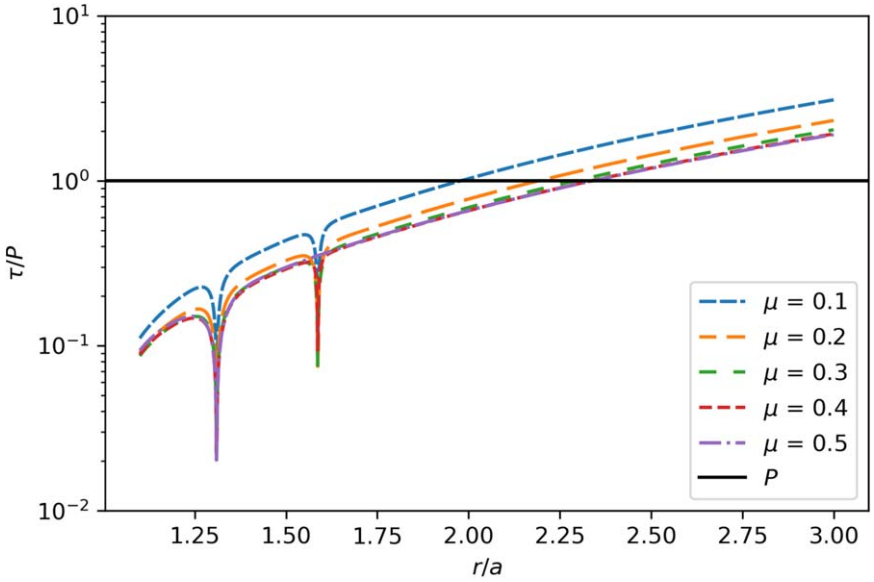

Figure 9. The dependence of the quantity τ defined in Equation (56) on the radius. The black line represents the binary period and τ is normalized by the binary period for convenience. The gap size in the orbital stability picture is defined as the location where the quantity τ intersects with the black line.

equivalently, where the parameter $\zeta_{\Delta t}$ given by

$$\zeta_{\Delta t} = \frac{\Delta t_{\rm close}}{\Delta t_{\rm open}} = \frac{1}{\alpha\chi^2}\frac{\bar{T}_m}{r} \tag{64}$$

exceeds unity, where the term $\bar{T}_m$ refers to the torque as defined in Equation (59), normalized by the surface density.

A second gap-opening criterion, which is used in Artymowicz & Lubow (1994) and Miranda & Lai (2015), considers the balance between the viscous and resonant torques instead of their associated timescales. It requires that the resonant torque, which clears the gap, be stronger than the viscous torque as defined in Equation (40) that brings fluid in to fill the gap. These two criteria differ by a factor of 3π. We define in this case a corresponding parameter, $\zeta_{\rm T}$, such that

$$\zeta_{\rm T} = \frac{T_m}{T_\nu} = \frac{1}{3\pi\alpha\chi^2}\frac{\bar{T}_m}{r} = \frac{\zeta_{\Delta t}}{3\pi}. \tag{65}$$

We note that, unlike the numerical computation of the torque in Equation (40), we do not need azimuthal averaging, since the background densities used to derive these expressions are already spherically symmetric. While the parameter $\zeta_{\rm T}$ provides a stronger check of gap opening due to the factor 3π, we note that both criteria ultimately yield the same results for the gap-opening resonance. In order to obtain a measure of gap size, we must refine the range of influence of the gap-opening resonance, since the torque is not, in reality, deposited at the discrete resonance radius, but rather it is spread over some neighborhood surrounding the resonance. In Artymowicz & Lubow (1994), this spread is addressed by considering the extent to which intersecting fluid streamlines bundle around the resonances. However, an explicit method of calculating their Table 1 is not provided. Therefore, we take this effect into consideration by turning, once again, to the theory of epicycles. We define the gap size to be the outermost radius where the amplitude of epicycles (given by Equation (28)) allows for the orbit to intersect the resonance location. That is, we can define the gap size at $r_{\rm T}$ to obey the condition

$$A(r_{\rm T}) = |r_{\rm T} - r_\zeta|, \tag{66}$$

where r_ζ is the location of the gap-opening resonance. Since this equation is nonlinear, we invert it to find $r_{\rm T}$ using a simple bisection solver since the form of the amplitude of epicycles is strictly decreasing from the location of the resonance. In order to better resolve the solution to the above equation, we choose to perform the bisection search on the logarithmic version of the same condition, that is

$$f(r_{\rm T}) \equiv \log\left(\frac{A(r_{\rm T})}{|r_{\rm T} - r_\zeta|}\right) = 0. \tag{67}$$

We present the results of the above gap size computation along with results from Artymowicz & Lubow (1994) in Figure 10 to show that they are consistent. Thus, we have recapitulated the resonant gap-opening picture as well as a measure of the gap size predicted by this picture. We can now provide a comparison of these gap sizes with those predicted by the orbital stability picture.

4.3. Comparison of Analytical and Numerical Results

Having arrived at multiple definitions of the gap size, we note that there is a stark contrast in the physics relevant to maintaining a circumbinary gap between the two analytical pictures. In the test particle case, the gap is maintained by instabilities propagated by the binary potential over timescales comparable to the binary period, while in the resonant torque case, the gap is maintained by viscous transport of angular momentum deposited at resonances over the viscous timescale. Additionally, the test particle picture would indicate that the gap size should not depend on the viscosity parameters, whereas the resonant torque picture would require a strong correlation between gap sizes and viscosity parameters.

It is also evident that the two pictures predict different trends in the variation of gap size over different mass ratios. In the test particle case, gap sizes increase monotonically with increasing

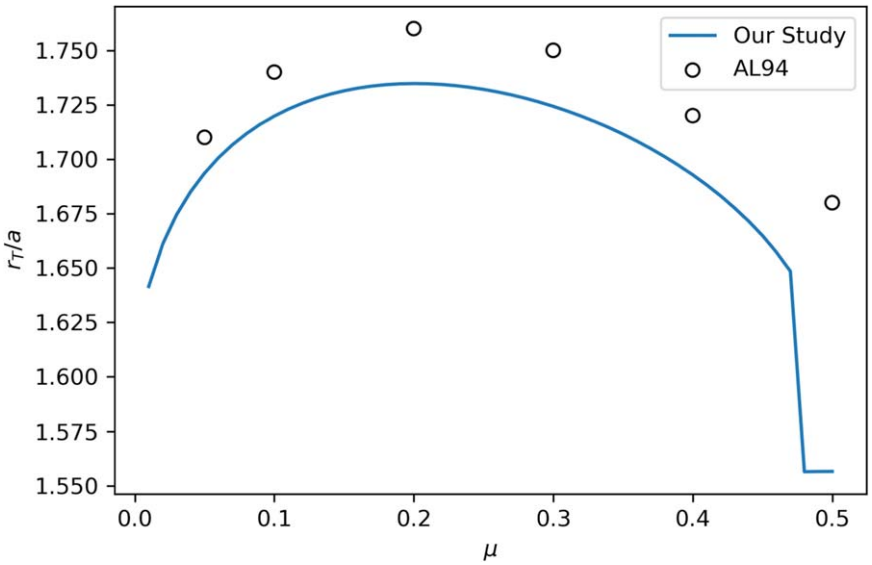

Figure 10. Comparison of the dependence of r_T from Equation (67) and the gap sizes from Table 1 of Artymowicz & Lubow (1994) over mass ratio μ. We find that there is an overall agreement of the trend against mass ratios except for the offset at $\mu \approx 0.5$. The torque from the $m = 1$ resonance abruptly drops at $\mu \approx 0.5$ leading to an immediate dominance of the $m = 2$ resonance.

Table 2
Summary of All Gap Size Definitions

Label	Definition	Defined
$r_{10\%}$	Radius where surface density is 10% of the final maximum	Table 1
r_{dT}	Radius where dynamical and viscous torque densities intersect	Table 1
r_L	Radius where the Lyapunov timescale is equal to the binary orbital period	Equation (56)
r_T	Radius where the epicyclic amplitude reaches the gap-opening resonance	Equation (67)

mass ratio. In contrast, in the resonant torque scenario, the gap size initially increases, reaches a maximum, and then decreases with increasing mass ratio, with a more abrupt decrease very close to the equal mass case where the transition from the $m = 1$ to the $m = 2$ resonance occurs. Thus, it is crucial to look to the numerical simulations of the disk-binary system to arrive at a conclusion as to which physics is actually responsible for creating and maintaining gaps in circumbinary disks.

We emphasize that we wish to compare the trends predicted in these two pictures, rather than the exact values for the gap size at any given mass ratio since a realistic circumbinary gap is just a diffuse region and does not have a hard boundary. We therefore normalize the results for the several radii that we have defined throughout this work, to clearly identify the trends they make when we overlay them graphically. To that end, we define the quantities $\bar{r}_X$ such that,

$$\bar{r}_X = \frac{r_L}{r_X}\Big|_{\mu=0.5} r_X, \tag{68}$$

where the subscript X denotes the subscript of the relevant radius from the numerical or analytical computations, e.g., $r_{10\%}$ as defined in Section 3. The vertical line denotes that the values are taken at the mass ratio $\mu = 0.5$, so that the gap size as defined by the instability timescale and the numerical/analytical measure of gap size are normalized to agree for equal masses. Finally, we recap the multiple definitions of the gap size in Table 2.

We plot the results of this comparison in Figure 11, which is the central result of this work. In each plot, the black solid line refers to the gap size determined by the Lyapunov timescale while the red dashed line corresponds to the gap size as determined by the matching of the Lindblad and viscous torques. The black and red points refer to the numerical results for $r_{10\%}$ and r_{dT}. In all cases, the lines are rescaled to match the result for $r_{10\%}$ at equal mass. We note the clear agreement in the trend of the numerical results with the gap size as predicted by the Lyapunov timescale, indicating that the physics responsible for the gap truncation and maintenance is the instabilities that are rapidly propagated at sub-binary timescales.

The disagreement between the resonant torque picture and the numerical results in Figure 11 reinforces the disagreement we saw in the azimuthal variation of the density distributions in Section 3.4. In the analytical picture, we expect that each successive multipolar contribution from the binary potential should affect the disk dynamics ever more weakly, due to the ever larger inverse power-law drop-offs of those higher-order contributions. Additionally, as we approach the equal mass case, there is a symmetry that ensures that odd multipole moments have zero contribution to the potential, and therefore there should be no density perturbations excited at those harmonics in the resonant torque scenario.

However, the numerical results show a stark contrast from this description. The magnitudes of azimuthal modes for all harmonics are within 2 orders of magnitude of each other, with odd harmonics never being suppressed to any degree as we approach the equal mass case. We note especially in the equal mass case that the $m = 1$ azimuthal perturbations dominate by almost an order of magnitude over the $m = 2$ mode that should

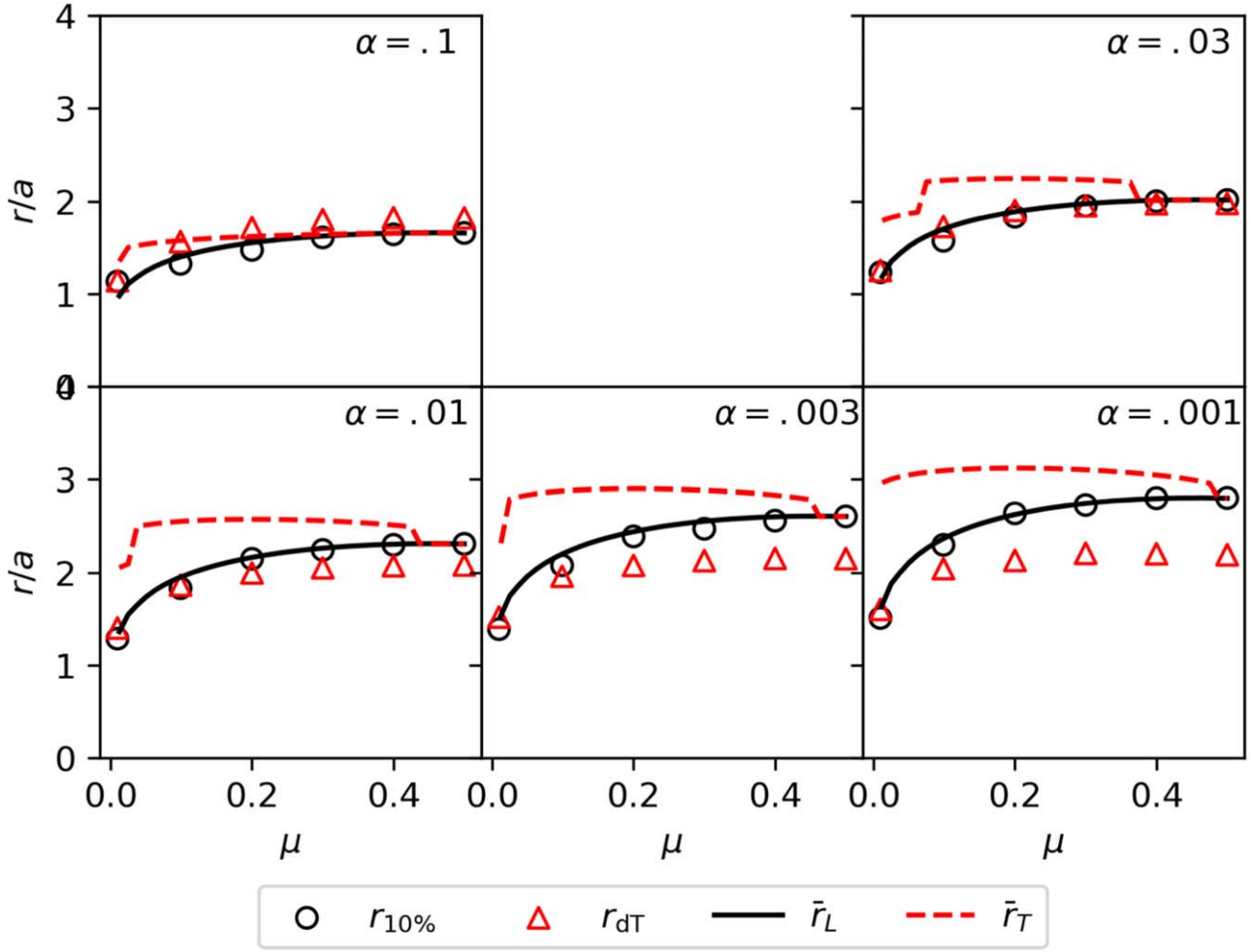

Figure 11. Comparison of numerical results of $r_{10\%}$ and $r_{\rm dT}$ in Table 1, and $r_{\rm L}$ and $r_{\rm T}$ computed in Equations (56) and (67). We notice that the trends for the numerical radii agree better with the gap size as defined by the instability timescale as opposed to the resonant torque balance criteria for the simulated range of α.

be the leading order multipole moment in the analytic theory. Additionally, we note that the strength of the higher azimuthal modes does not decay, which indicates strongly that the different azimuthal modes are being propagated nonlinearly.

These results, and the ability of test particle orbit instabilities to explain the numerically observed behavior of the circumbinary gap, lead us to the conclusion that angular momentum transport in circumbinary disks is not well explained by a competition of dynamical and viscous torques. Specifically, the location of the circumbinary gap does not correlate with the viscosity of the disk. Additionally, we do not observe a drop-off in the circumbinary gap size as we approach equal mass ratios. Azimuthal variations in the density profile are also not modulated in any way by the multipole moments that excite them in the perturbative picture, be it the radial drop-off in individual cases or the weakening of odd multipole moments as we approach the symmetrical equal mass case.

5. Conclusions

We have provided a comprehensive study of the effect of a binary on a surrounding circumbinary disk through both analytical theory as well as through two-dimensional, Newtonian, viscous hydrodynamic simulations. In this work, the first of two parts, we limit ourselves to the case of a coplanar disk and binary system. In doing so, we have been able to consider multiple configurations of the only free parameter——the mass ratio of the binary. In both analytical and numerical computations, we have modeled the binary as a pair of point masses with a corresponding Newtonian gravitational potential. We have modeled the disk as a non-self-gravitating, locally isothermal fluid with an alpha-type viscosity. Lastly, in both the analytical and numerical studies, we have limited the domain of interest to the exterior of the binary's orbit by excising the inner low-density region from the numerical domain, and by choosing a harmonic decomposition of the potential in the regime of space outside the semimajor axis of the binary, and the behavior of test particle orbits under the static background of the combined potential for our analytical computations.

We have numerically investigated the behavior of the two-dimensional density distribution of the viscous and locally isothermal fluid under the influence of the binary gravitational potential. The computations initialized and evolved four primitive variables: the surface mass density, the surface pressure, and the radial and azimuthal components of the velocity. The knowledge of these variables provides sufficient information for us to calculate other astrophysically relevant characteristics of the disk in post-process.

Since we are focused on the density profiles, the most intuitive and natural way to present our output is to plot the

azimuthal- and time-averaged density as a function of radius. Together with certain examples of the full two-dimensional distributions, we can draw broad conclusions about the behavior of matter in the disk. To fully characterize the disk, we also studied other significant characteristics, including the locations of the density maxima and their values, the radii where the viscous and dynamical torque densities are equal, and their value at that point, and the radii where the density in the central cavity drops to 10% of its instantaneous peak value.

In the analytical computations, we reviewed the theory of approximate orbits in the binary potential, including what the Lindblad resonances are and how they have been used, in combination with linearized fluid mechanics, to explain the formation of a circumbinary gap. We have additionally introduced a measure of the Lyapunov timescale for an analytically tractable set of approximate orbits and, in doing so, we have defined a new measure of the gap size. This new measure reflects the rapidity with which instabilities in the linear regime propagate. We find that the new gap sizes agree better with the trends observed in the full numerical simulations. This supports the conclusion that the circumbinary gap is maintained by the effect of the binary potential on the disk over timescales much shorter than previously suggested. Additionally, we conclude that the viscosity plays a minimal role in the formation of the gap, as the gap sizes are determined solely by the characteristics of the binary. Once again, this suggests a different picture than the case of resonant torque dampening since that effect would depend on the magnitude of the viscous torque density.

We add that this conclusion reflects the α-disk viscosity profile and the values of α that we have studied. Studies have shown that the morphology of the inner regions of the circumbinary disks is still affected by changing the nature of the viscosity (D'Orazio et al. 2013; Dittmann & Ryan 2022; Dittmann et al. 2023). Particularly, D'Orazio et al. (2013) introduce an effective viscosity force to the test particle dynamics when studying the stability of test particle orbits. In such a case, the stability of the Lagrange points L4 and L5 (stationary solutions to Equation (19)) indicate a strong dependence of the timescale on the viscous force at mass ratios $q < 0.04$. Extending the orbital stability regime developed here to perturbed orbits around the individual point masses could provide an understanding of the kinematics at radial scales where the point masses are present.

Dittmann & Ryan (2022) simulate large ranges of viscosities, and demonstrate substantial variability of the gap size as they define it. In addition, Dittmann et al. (2023) study the existence of a decoupled regime as indicated in Milosavljević & Phinney (2005), where the gap size is demonstrated to not track the evolution of the binary at effectively $\alpha\chi^2 > 3 \times 10^{-4}$. We note that since we fix $\chi = 0.1$, this latter range corresponds to $\alpha > 0.03$ in our case, where we have shown that the locations of the Lyapunov-timescale- and torque-balance-based estimates for the gap size begin to overlap. We also expect viscosity to be more dynamically important when it is very large. In addition, we note that the definition for the gap sizes in Dittmann & Ryan (2022) and Dittmann et al. (2023) is the location where the density is 20% of σ_0. This reference density will generally be much larger than ours since the final density maximum is less than the initial maximum, which in turn is less than σ_0 (the initial maximum density is only $\sim 0.4\sigma_0$ in our case), and obviously $10\% < 20\%$; the gaps as defined in Dittmann & Ryan (2022) and Dittmann et al. (2023) will therefore occur at much larger distances from the binary, where viscosity might become more important. Also, the criterion for the decoupling phase used in Dittmann et al. (2023) is different from that of Milosavljević & Phinney (2005), and Dittmann et al. (2023) identify the decoupling phase as when the binary separation and the disk radial velocity are comparable as opposed to the comparison of viscous and gravitational-wave radiation timescales.

In our follow-up paper (Paper II), we extend the parameter space of configurations by introducing a constant inclination angle between the binary and disk planes. We will apply the same tools used here to characterize the binary's potential and the disk's fluid properties. We will also continue to restrict the domain of interest to the exterior of the binary.

The primary focus of these two studies is to provide a coherent physical description of the properties of the circumbinary disk and its central gap by combining analytical and numerical approaches. To that effect, the first paper accomplishes the task of providing a self-contained overview of both the analytical and numerical approaches that we consider in this two-part study as well as future applications. It then provides an overview of the inconsistencies in the current analytical understanding of circumbinary disks. Finally, we suggest a new approach, the stability of epicyclic orbits, to reconcile these inconsistencies and to more accurately describe the physics governing the behavior of circumbinary disks.

Acknowledgments

The authors thank Zoltan Haiman and Alex Dittman for their comments and discussions. S.M. and S.T.M. were supported in part by a National Science Foundation CAREER grant No. 1945130 and a National Aeronautics and Space Administration grant No. 22-LPS22- 0022. The authors acknowledge the computational resources provided by the WVU Research Computing Spruce Knob HPC cluster, which is funded in part by NSF EPS-1003907, and the Thorny Flat HPC cluster, which is funded in part by NSF OAC-1726534. This research was supported in part by the National Science Foundation under grant No. NSF PHY-1748958.

ORCID iDs

Siddharth Mahesh https://orcid.org/0000-0002-8340-614X
Sean T. McWilliams https://orcid.org/0000-0003-2397-8290

References

Artymowicz, P., & Lubow, S. H. 1994, ApJ, 421, 651
Artymowicz, P., & Lubow, S. H. 1996, ApJL, 467, L77
Binney, J., & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton, NJ: Princeton Univ. Press)
Dittmann, A. J., & Ryan, G. 2022, MNRAS, 513, 6158
Dittmann, A. J., Ryan, G., & Miller, M. C. 2023, ApJL, 949, L30
Dong, R., Rafikov, R. R., & Stone, J. M. 2011, ApJ, 741, 57
D'Orazio, D. J., Haiman, Z., Duffell, P., MacFadyen, A., & Farris, B. 2016, MNRAS, 459, 2379
D'Orazio, D. J., Haiman, Z., & MacFadyen, A. 2013, MNRAS, 436, 2997
Duffell, P. C. 2016, ApJS, 226, 2
Duffell, P. C., & MacFadyen, A. I. 2012, ApJ, 755, 7
Duffell, P. C., & MacFadyen, A. I. 2013, ApJ, 769, 41
Farris, B. D., Duffell, P., MacFadyen, A. I., & Haiman, Z. 2014, ApJ, 783, 134
Farris, B. D., Duffell, P., MacFadyen, A. I., & Haiman, Z. 2015, MNRAS, 446, L36

Frank, J., King, A., & Raine, D. 2002, Gas Dynamics (3rd ed.; Cambridge: Cambridge Univ. Press), 8
Goldreich, P., & Tremaine, S. 1980, ApJ, 241, 425
Goldstein, H., Poole, C., & Safko, J. 2002, Classical Mechanics (Reading, MA: Addison-Wesley)
Günther, R., & Kley, W. 2002, A&A, 387, 550
Krauth, L. M., Davelaar, J., Haiman, Z., et al. 2023, MNRAS, 526, 5441
MacFadyen, A. I., & Milosavljević, M. 2008, ApJ, 672, 83
Milosavljević, M., & Phinney, E. S. 2005, ApJL, 622, L93
Miranda, R., & Lai, D. 2015, MNRAS, 452, 2396
Pichardo, B., Sparke, L. S., & Aguilar, L. A. 2005, MNRAS, 359, 521
Poisson, E., & Will, C. M. 2014, Gravity (Cambridge: Cambridge Univ. Press)
Rudak, B., & Paczynski, B. 1981, AcA, 31, 13
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 500, 33
Siwek, M., Weinberger, R., & Hernquist, L. 2023, MNRAS, 522, 2707
Tang, Y., Haiman, Z., & MacFadyen, A. 2018, MNRAS, 476, 2249
Tiede, C., Zrake, J., MacFadyen, A., & Haiman, Z. 2022, ApJ, 932, 24
Toro, E. F. 2013, Riemann Solvers and Numerical Methods for Fluid Dynamics (Berlin: Springer),
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261

3.2 Predictions for Misaligned Disk-Binary Systems

The previous section provided a new and successful theoretical framework for understanding the formation of the circumbinary gap. The resulting depiction anchors the gap formation timescale to the orbital timescale of the binary orbital period, explaining why the disk moves inward with the inspiraling binary. One key prediction of this model is that the gap size should be sensitive to relativistic effects that become important as the binary separation decreases. One such important effect is the precession of the binary orbital plane due to spin-orbit interactions in general relativity.

At large separations, the leading order effect of spin-orbit precession is the rotation of the orbital angular momentum of the binary, and the individual black hole spins around the total angular momentum vector. The relevant timescale for the precession is given by the spin precession frequency, Ω_{prec} [18],

$$\Omega_{prec} = \frac{7GJ}{2c^2a^3} \Rightarrow \Delta t_{prec} = \frac{4\pi c^2 a^3}{7GJ}, \tag{3.4}$$

where J is the total angular momentum of the binary. At large separations, the total angular momentum is dominated by the orbital angular momentum of the binary, L. Using Kepler's laws to relate $L = \nu M\sqrt{GMa}$, we get

$$\frac{\Delta t_{prec}}{R_g c^{-1}} = \frac{4\pi}{7\nu}\left(\frac{a}{R_g}\right)^{5/2} \tag{3.5}$$

where $R_g = \frac{GM}{c^2}$ is the gravitational radius. At large separations, $a \gg R_g$, the precession timescale is much longer than the orbital period but not longer than the characteristic viscous timescale. As a result, it was not initially expected that the precession of the binary would have any significant effect on the circumbinary disk. On the other hand, with the new understanding that the maintenance of the gap is tied to the orbital period of the binary, it is possible that the precession of the binary could have a significant effect on the circumbinary disk.

In the disk-binary system, the effect of the spin-orbit precession is to cause a relative inclination between the binary orbital plane and the disk. Since the precession timescale is a large multiple of the orbital time period, it is possible to assess the impact of the misalignment numerically in the limit where the inclination between the disk and binary planes is assumed to be constant in time. More importantly, the change in the location of the circumbinary gap over inclinations provides a strong scientific validation of the test particle model of the circumbinary gap, extending the astrophysical importance of this model beyond a reconciliation between analytical and numerical results.

The rest of this section provides the results of this study and validates the predictions of the gap size for inclined disk-binary systems in both cases. The work relevant to the study is contained in the publication [19]. The text recapitulates the theoretical preliminaries introduced in the last section and generalizes them to the case where the binary plane is inclined with respect to the disk plane. The resulting predictions are then compared to numerical simulations of inclined disk-binary systems.

My contribution to the inclined disk study involved extending the resonant and epicyclic-stability analyses to misaligned disk-binary systems, identifying the predicted unstable sector in inclination, comparing the analytical predictions with numerical simulations, and interpreting the resulting time-dependent disk behavior. This work tests whether the instability-based gap picture developed in the coplanar case remains predictive when the disk and binary angular momenta are not aligned.

The Astrophysical Journal, 980:130 (23pp), 2025 February 10 https://doi.org/10.3847/1538-4357/ada55e

Analytical and Numerical Methods for Circumbinary Disk Dynamics. II. Inclined Disks

Michal Pirog[1,2,3], Siddharth Mahesh[2,3], and Sean T. McWilliams[2,3]
[1] The Harriet L. Wilkes Honors College, Florida Atlantic University, Boca Raton, FL 33431, USA; mpirog@fau.edu
[2] Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA
[3] Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA
Received 2024 October 24; revised 2024 December 14; accepted 2025 January 1; published 2025 February 7

Abstract

To gain insight into the dynamical influence of a supermassive black hole binary on a circumbinary accretion disk, we investigate the binary and viscous torque densities throughout such a disk, with emphasis on the final density distribution, particularly the size and stability of the central gap between the binary and the inner edge of the disk. We limit ourselves to the simplified case of a massless viscous thin accretion disk under the influence of the gravitational potential from a binary system whose orbital plane is inclined relative to the disk. We employ two-dimensional Newtonian hydrodynamics simulations to examine the influence of two model parameters: the mass ratio of the binary and the inclination angle between the binary and the disk. We investigate their impact on the density and torque distribution. In our analytical approach, we consider the stability of epicycles induced by the perturbative effect of the asymmetric inclined binary gravitational potential on Keplerian circular orbits. Through our simulations, we observe that certain configurations never attain a quasi-steady state, where the density profile averaged over many orbits stabilizes. This instability occurs when the inclination is close to 45°. Furthermore, we identify configurations where there is never a persistent balance between the dynamical and viscous torque densities, as well as cases where the location of this balance oscillates or exhibits other time-dependent behavior over viscous timescales. These findings have implications for understanding both the expected gravitational-wave signal and electromagnetic counterparts from supermassive black hole binaries.

Unified Astronomy Thesaurus concepts: Astrophysical fluid dynamics (101); Accretion (14); Circumstellar disks (235); Orbital theory (1182); Orbital resonances (1181); Computational methods (1965); Astrophysical black holes (98)

1. Introduction

The evolution of gaseous accretion disks surrounding supermassive black hole binaries is a crucial and active area of research in astrophysics. Observationally, understanding the dynamics of these circumbinary disks is vital for interpreting any electromagnetic signatures that might be used to identify supermassive black hole binary candidates (I. Bartos et al. 2017; N. C. Stone et al. 2017; B. McKernan et al. 2018; H. Tagawa et al. 2020). Theoretically, the mechanisms by which a gap between the binary and the inner edge of the disk is created and maintained are interesting open problems. Early analytical work on resonant torques and disk dynamics addressing the problem of gap formation in circumbinary disks was presented in P. Goldreich & S. Tremaine (1980) and P. Artymowicz & S. H. Lubow (1996) for coplanar and eccentric cases, respectively. R. Miranda & D. Lai (2015) investigated the problem of eccentric binaries inclined with respect to the disk's plane, focusing on the same problem; namely, the balance between binary and viscous torques and its impact on the location of the central gap. Numerically, simulating the progression of the unequilibrated disk toward a quasi-steady state (or, in some cases, the lack thereof) and the maintenance of the circumbinary gap over viscous timescales are significant challenges.

Numerical simulations of these systems have been conducted using both smooth particle hydrodynamics (see e.g., M. R. Bate et al. 1995; P. Artymowicz & S. H. Lubow 1996; A. Escala et al. 2005; K. Hayasaki et al. 2007; J. Cuadra et al. 2009; C. Roedig et al. 2012; F. I. Pelupessy & S. Portegies Zwart 2013; E. Ragusa et al. 2016), and grid-based hydrodynamics methods (e. g. R. Günther & W. Kley 2002; A. I. MacFadyen & M. Milosavljević 2008; T. Hanawa et al. 2010; M. de Val-Borro et al. 2011; D. J. D'Orazio et al. 2013; S. Lines et al. 2015; R. Miranda et al. 2017). Our approach to this problem originates from the work of A. I. MacFadyen & M. Milosavljević (2008), who investigated a two-dimensional coplanar configuration of an equal-mass binary and a thin, viscous, isothermal disk. Since then, several generalizations have been made. D. J. D'Orazio et al. (2013) studied different mass ratios of the binary within Newtonian hydrodynamics. Generalizations from pure Newtonian hydrodynamics to Newtonian magnetohydrodynamics (MHD), then to post-Newtonian MHD, and finally to General-Relativistic MHD have been made in e. g. J.-M. Shi et al. (2012), S. C. Noble et al. (2012), and B. D. Farris et al. (2012), respectively. Further studies of the topic can be found, for example, in J.-M. Shi & J. H. Krolik (2015) and D. B. Bowen et al. (2019). Additionally, recent studies exploring the vicinity of each black hole and extending the description to the Kerr metric are discussed in L. Combi et al. (2021) and L. Combi et al. (2022).

Circumbinary disks that are inclined with respect to the binary plane have also been investigated by E. I. Chiang & R. A. Murray-Clay (2004) in the context of planetary systems. The authors provided evidence that the inclination of the disk may explain the observed electromagnetic flux from the pre-main-sequence star KH 15D. Similarly, T. Hioki et al. (2011) considered the misalignment between circumstellar and

circumbinary disks as an explanation for the observed behavior of FS Tauri. In both of these studies, the models' parameters were chosen to match those specific astronomical systems. Analytical studies of disks inclined with respect to the binary plane were carried out by F. Foucart & D. Lai (2014), who derived an analytical description for the long-term evolution of the disk in an axisymmetric perturbed gravitational potential. Inclined three-dimensional configurations were investigated in A. Dittmann et al. (2024), which focused on the effect of an active galactic nucleus disk on an embedded binary. The authors analyzed numerous configurations with various inclinations, including retrograde cases. They found that the duration of the inspiral stage, due to gravitational interactions with the disk, could be up to 4 times shorter than prograde configurations. However, the implications of this result for the isolated binary problem remain uncertain.

In this work, we attempt a more general study of the dynamics of circumbinary disks, covering a wide range of our model parameters. However, our approach differs from the aforementioned studies in one important aspect: we do not consider varying eccentricities and instead assume circular orbits. While eccentricity can play an important role, here, we specifically concentrate on exploring how the disk dynamics varies with the binary mass ratio and inclination angle. This work is a continuation of our earlier work, S. Mahesh et al. (2024; hereafter Paper I), where we studied the same configuration but with zero inclination.

We emphasize that, despite the physical problem being three-dimensional, the hydrodynamical field equations we employ primarily involve two-dimensional functions. The only three-dimensional aspect of this work pertains to the inclination between the binary's orbital plane and the disk, where the total gravitational potential within the disk is due to the combined influence of the two binary components that are executing an inclined circular orbit. In our numerical analysis, we employ two-dimensional Newtonian hydrodynamics simulations to examine the influence of two model parameters: the *mass ratio* of the binary and the *inclination angle* between the binary and the disk. We investigate their impact on the density and torque distribution. It is important to acknowledge the limitations of our two-dimensional simulations. By construction, these simulations do not account for several three-dimensional effects such as disk warping, twisting, disruption, and differential precession, which are likely significant in inclined astrophysical systems (S. Facchini et al. 2010). These three-dimensional effects can lead to complex disk morphologies and dynamics that are not captured in our current model.

In disk–binary interactions, it has long been argued that the binary will clear a gap in the disk through resonant torquing (P. Artymowicz & S. H. Lubow 1996). In that scenario, the gap size would be determined by imposing that the timescale of angular momentum deposition by the Lindblad Torques (gap opening timescale) far exceeds the timescale of viscous dissipation of angular momentum through the disk (gap closing timescale; P. Goldreich & S. Tremaine 1980). Using these ideas, it has been suggested that an electromagnetic afterglow from a circumbinary disk could brighten and be observed long after a pair of massive black holes have merged (M. Milosavljević & E. S. Phinney 2005). However, ideas like this rely on the assumption that a gap is maintained due to the resonant behavior of the disk–binary interaction. However, in Paper I, a study of perturbed orbits in the binary potential revealed incredibly high epicyclic advances due to the harmonic modes of the binary potential at a broader range of distances around the corresponding Lindblad resonance. Furthermore, the perturbed orbits at these distances exhibit instabilities with an e-fold timescale much less than the binary orbital period. Thus, the driving of short-timescale instabilities in the inner disk was suggested as being responsible for opening and maintaining circumbinary gaps for these systems.

The effect of misalignment of the binary and disk planes is to weaken the effect of the binary potential (R. Miranda & D. Lai 2015). While both gap opening mechanisms discussed in Paper I rely on the strength of the gravitational potential, the effect of the weakening manifests differently in each picture. The predictions from the orbital stability picture tend to portray a gradual decrease in the size of the circumbinary gap. On the other hand, the weakening of the resonant torques leads to sharp transitions of the dominant Lindblad resonances, sometimes predicting an increase in the gap size close to counterrotating inclinations. We demonstrate that, while the numerical gap sizes show some transitionary trends, they do not coincide with the inclinations predicted by the resonant torque picture. However, when considering the unstable nature of the quasi-stationary state for these inclinations, the overall trend is still better predicted by the orbital stability picture.

Astrophysically, misalignment of binary and disk planes can be induced by the general-relativity-induced precession of the spin and orbital angular momenta of the binary over an associated timescale (E. Poisson & C. M. Will 2014). Should the short-timescale picture hold over a range of binary–disk inclinations, the precession timescale would exceed the gap opening timescale. This would allow for an adiabatic treatment of the disk–binary environment where the gap size fluctuates over the range of inclinations accessed by the binary. Additionally, the effect of the inner disk instabilities could contribute to an electromagnetic counterpart to the gravitational radiation from the binary.

In order to compare our analytical predictions with numerical simulations, we utilized the DISCO code, which has previously demonstrated its effectiveness as a solver in similar studies (see, e.g., B. D. Farris et al. 2014; D. J. DOrazio et al. 2016; Y. Tang et al. 2017; P. C. Duffell et al. 2020). A detailed description of this numerical tool can be found in P. C. Duffell (2016). Furthermore, an extensive study comparing various numerical codes designed for binary–disk interactions, including DISCO, is provided in P. C. Duffell et al. (2024).

The rest of the paper is structured as follows. Section 2 provides a detailed description of the model's construction, features, and parameters, accompanied by the necessary mathematical formulas to express it within the framework of Newtonian gravity and hydrodynamics. We then delve into the numerical approach employed, followed by a presentation of our initial data and boundary conditions. In Section 3, we focus on the results of our numerical simulations as summarized in Table 1 and introduce the concept of the quasi-steady state. For the sake of convenience, we adopt the terms moderately inclined, highly inclined, and counterrotating systems to denote inclinations $\iota \in \langle 0°, 30° \rangle$, $\iota \in \langle 55°, 90° \rangle$, and $\iota \in (90°, 180° \rangle$, respectively. The range of $\iota \in (30°, 55°)$ is referred to as the unstable region. We present the details of our analytical model in Section 4 and compare it with our numerical results in Section 5. Finally, Section 6 provides a summary, draws key conclusions, and outlines potential future research directions.

Table 1
Summary of Simulated Configurations

Run's Code	q	ι	t_*	$r_{\rm dT}$	$\tilde{r}_{\rm dT}$	$\varsigma(r_{\rm dT})$ $(r_{\rm dT})$	$r_{\rm max}$	$\tilde{r}_{\rm max}$	$\varsigma\,(r_{\rm max})$	r_{10} %
1-000	1:1	0°	12,000	2.0786	2.0817	0.0045	4.7568	4.7643	0.0342	2.3194
1-015	1:1	15°	12,000	2.0599	2.0749	0.0223	4.7568	4.7719	0.0234	2.2943
1-020	1:1	20°	36,000	2.0632	2.0817	0.0045	4.7568	4.7643	0.0342	2.2922
1-025	1:1	25°	12,000	2.1203	2.0879	0.0572	4.7568	4.7492	0.0186	2.2723
1-030	1:1	30°	30,000	2.1827	2.0665	0.0987	4.7112	4.7492	0.0186	2.2464
1-035	1:1	35°	24,000	2.2467	2.0531	0.1338	4.7112	4.7491	0.0342	2.2163
1-040	1:1	40°	24,000	1.9178	2.0858	0.2130	4.7568	4.7415	0.0371	2.1927
1-045*	1:1	45°	48,000	1.9681	1.8663	0.0566	4.7568	4.8404	0.0535	2.1935
1-050	1:1	50°	24,000	1.7466	1.7467	0.0001	2.4390	2.4390	0.0000	1.7150
1-055	1:1	55°	12,000	1.7356	1.7357	0.0001	2.4030	2.4030	0.0000	1.6999
1-060	1:1	60°	12,000	1.7240	1.7240	0.0001	2.3672	2.3672	0.0000	1.6647
1-075	1:1	75°	12,000	1.6905	1.6906	0.0002	2.2959	2.2959	0.0000	1.5833
1-090	1:1	90°	12,000	1.6428	1.6427	0.0001	2.2959	2.2959	0.0000	1.4778
1-105	1:1	105°	12,000	1.6029	1.6029	0.0002	2.2959	2.2959	0.0000	1.3683
1-120	1:1	120°	12,000	1.5760	1.5745	0.0013	2.5472	2.5533	0.0148	1.2993
1-135	1:1	135°	12,000	1.5320	1.5282	0.0035	3.6491	3.6557	0.0164	1.2394
1-150	1:1	150°	12,000	…	…	…	4.1855	4.2069	0.0234	1.2294
1-165	1:1	165°	12,000	1.1595	1.1595	0.0000	4.7568	4.7794	0.0248	1.2457
1-180	1:1	180°	12,000	1.1617	1.1617	0.0000	4.8479	4.8786	0.0377	1.2576
2-000	2:3	0°	12,000	2.0745	2.0736	0.0020	4.7112	4.7340	0.0250	2.2997
2-015	2:3	15°	12,000	2.0913	2.0738	0.0233	4.7112	4.7188	0.0186	2.2788
2-020	2:3	20°	36,000	2.0220	2.0736	0.0020	4.7112	4.7340	0.0250	2.2681
2-025	2:3	25°	12,000	1.9962	2.0550	0.0626	4.6662	4.6887	0.0246	2.2488
2-030	2:3	30°	24,000	2.0295	2.0503	0.0830	4.6662	4.6737	0.0184	2.2313
2-035*	2:3	35°	24,000	2.1543	2.1149	0.1185	4.5771	4.6439	0.0469	2.2108
2-040*	2:3	40°	24,000	1.8632	2.0589	0.2060	4.6662	4.6142	0.0593	2.1792
2-045	2:3	45°	48,000	1.8808	1.8414	0.0242	4.8020	4.8021	0.0288	2.1860
2-050**	2:3	50°	12,000	1.7366	1.8065	0.1280	2.4390	2.9067	0.5343	1.6875
2-055	2:3	55°	12,000	1.7355	1.7352	0.0005	2.4030	2.4090	0.0147	1.6797
2-060	2:3	60°	12,000	1.7214	1.7214	0.0002	2.4030	2.4030	0.0000	1.6574
2-075	2:3	75°	12,000	1.6811	1.6807	0.0004	2.2959	2.2840	0.0184	1.5692
2-090	2:3	90°	12,000	1.6443	1.6442	0.0001	2.5109	2.5109	0.0000	1.5029
2-105	2:3	105°	12,000	1.6062	1.6062	0.0000	2.3315	2.3315	0.0000	1.3884
2-120	2:3	120°	12,000	1.5785	1.5794	0.0008	2.9147	2.9147	0.0000	1.2869
2-135	2:3	135°	12,000	1.5501	1.5467	0.0049	3.5691	3.5757	0.0163	1.2687
2-150	2:3	150°	12,000	…	…	…	4.0588	4.0658	0.0171	1.2595
2-165	2:3	165°	12,000	…	…	…	4.4883	4.4957	0.0180	1.2526
2-180*	2:3	180°	12,000	…	…	…	4.8020	4.8556	0.0346	1.2581
3-000	3:7	0°	12,000	2.0524	2.0521	0.0007	4.5325	4.5325	0.0000	2.2462
3-015	3:7	15°	12,000	2.0754	2.0458	0.0268	4.4883	4.5104	0.0242	2.2343
3-020	3:7	20°	12,000	2.0595	2.0521	0.0007	4.5325	4.5325	0.0000	2.1915
3-025	3:7	25°	12,000	2.0832	2.0355	0.0639	4.4883	4.4957	0.0180	2.1980
3-030	3:7	30°	24,000	1.9146	2.0224	0.1005	4.4444	4.4518	0.0330	2.1648
3-035*	3:7	35°	12,000	2.1695	2.0126	0.1275	4.3142	4.3936	0.0508	2.1254
3-040	3:7	40°	12,000	2.0339	2.0111	0.1406	4.2712	4.2855	0.0222	2.1135
3-045*	3:7	45°	48,000	2.0463	1.9729	0.0427	4.4007	4.5255	0.0857	2.1135
3-050	3:7	50°	12,000	1.7155	1.7288	0.0119	4.3142	4.3502	0.0326	1.6648
3-055	3:7	55°	12,000	1.6951	1.6964	0.0014	4.3142	4.3142	0.0000	1.6560
3-060	3:7	60°	12,000	1.6848	1.6855	0.0005	4.2712	4.2712	0.0000	1.6529
3-075	3:7	75°	12,000	1.6795	1.6795	0.0001	3.8516	3.8516	0.0000	1.6084
3-090	3:7	90°	12,000	1.6358	1.6357	0.0001	3.9339	3.9339	0.0000	1.5234
3-105*	3:7	105°	12,000	1.5969	1.5970	0.0001	3.4505	3.3852	0.0320	1.4155
3-120	3:7	120°	12,000	1.5811	1.5811	0.0007	3.0651	3.0462	0.0207	1.3067
3-135	3:7	135°	12,000	1.5949	1.5929	0.0038	3.3333	3.3333	0.0000	1.3024
3-150	3:7	150°	12,000	…	…	…	3.7702	3.7566	0.0210	1.2841
3-165	3:7	165°	12,000	…	…	…	4.3574	4.3790	0.0237	1.2589
3-180	3:7	180°	12,000	…	…	…	4.8020	4.8403	0.0345	1.2606
4-000	1:4	0°	12,000	1.9934	1.9925	0.0019	4.2283	4.2283	0.0000	2.1530
4-015	1:4	15°	12,000	1.9618	1.9861	0.0260	4.1855	4.1714	0.0218	2.1133
4-020*	1:4	20°	12,000	2.0120	1.9925	0.0019	4.1008	4.2283	0.0000	2.0770
4-025	1:4	25°	12,000	2.0549	1.9872	0.0580	4.0588	4.0588	0.0000	2.0562
4-030	1:4	30°	24,000	2.0353	1.9567	0.0872	3.9755	3.9962	0.0347	1.9972

Table 1
(Continued)

Run's Code	q	ι	t_*	$r_{\rm dT}$	$\tilde{r}_{\rm dT}$	$\varsigma(r_{\rm dT})$ $(r_{\rm dT})$	$r_{\rm max}$	$\tilde{r}_{\rm max}$	$\varsigma(r_{\rm max})$	r_{10} %
4-035	1:4	35°	12,000	2.0981	1.9396	0.1241	3.9339	3.9616	0.0215	1.9682
4-040*	1:4	40°	12,000	1.8819	1.8603	0.0846	3.8516	3.9002	0.1199	1.9311
4-045**	1:4	45°	48,000	2.0057	2.0679	0.0573	4.0169	3.9075	0.1533	1.9717
4-050	1:4	50°	12,000	1.7301	1.7240	0.0058	4.4444	4.4737	0.0227	1.6726
4-055	1:4	55°	12,000	1.6733	1.6730	0.0003	4.0169	4.0169	0.0000	1.6312
4-060	1:4	60°	12,000	1.6642	1.6641	0.0004	4.0588	4.0588	0.0000	1.6181
4-075	1:4	75°	12,000	1.6265	1.6269	0.0003	4.1008	4.1008	0.0000	1.5703
4-090	1:4	90°	12,000	1.5962	1.5963	0.0002	4.1432	4.1220	0.0232	1.5011
4-105	1:4	105°	12,000	1.5598	1.5596	0.0001	3.9339	3.9133	0.0226	1.4043
4-120	1:4	120°	12,000	1.5590	1.5595	0.0007	3.1793	3.1665	0.0197	1.3137
4-135	1:4	135°	12,000	1.5950	1.5963	0.0055	3.2175	3.2175	0.0000	1.3163
4-150	1:4	150°	12,000	…	…	…	3.5691	3.5558	0.0205	1.2956
4-165	1:4	165°	12,000	…	…	…	4.2283	4.2712	0.0272	1.2658
4-180	1:4	180°	12,000	…	…	…	4.7568	4.8022	0.0407	1.2659
5-000	1:10	0°	12,000	1.8616	1.8616	0.0020	3.4112	3.4440	0.0160	1.8305
5-015	1:10	15°	12,000	1.8531	1.8329	0.0229	3.2559	3.2817	0.0200	1.7858
5-020*	1:10	20°	24,000	1.8822	1.8616	0.0020	3.3333	3.4440	0.0160	1.7678
5-025	1:10	25°	12,000	1.7315	1.8123	0.0766	3.3333	3.3398	0.0158	1.7292
5-030	1:10	30°	24,000	1.8619	1.7939	0.0752	3.3722	3.3982	0.0320	1.6829
5-035*	1:10	35°	24,000	1.7022	1.7042	0.0412	3.3333	3.3918	0.0479	1.6103
5-040	1:10	40°	12,000	1.8393	1.8413	0.0035	3.6090	3.5890	0.0219	1.7081
5-045	1:10	45°	36,000	1.6428	1.6471	0.0098	3.6892	3.6691	0.0220	1.4868
5-050	1:10	50°	12,000	1.6062	1.6064	0.0002	3.8516	3.8516	0.0000	1.4589
5-055	1:10	55°	12,000	1.5969	1.5969	0.0005	3.9755	3.9755	0.0000	1.4531
5-060	1:10	60°	12,000	1.5865	1.5861	0.0004	4.0169	4.0169	0.0000	1.4479
5-075	1:10	75°	12,000	1.5543	1.5535	0.0006	4.1008	4.1220	0.0232	1.4136
5-090	1:10	90°	12,000	1.5225	1.5233	0.0010	4.1008	4.1079	0.0173	1.3652
5-105	1:10	105°	12,000	1.4853	1.4863	0.0008	3.6491	3.6424	0.0163	1.2929
5-120	1:10	120°	12,000	1.4304	1.4332	0.0036	3.4112	3.3852	0.0202	1.2693
5-135	1:10	135°	12,000	1.3766	17.5756	39.6563	3.4112	3.4112	0.0000	1.2793
5-150	1:10	150°	12,000	…	…	…	3.6491	3.6557	0.0164	1.2721
5-165	1:10	165°	12,000	…	…	…	4.1432	4.1291	0.0219	1.2547
5-180	1:10	180°	12,000	…	…	…	4.6662	4.6963	0.0369	1.2646
6-000	1:100	0°	12,000	1.4017	1.4014	0.0004	3.1030	3.1030	0.0000	1.2789
6-015	1:100	15°	12,000	1.3709	1.3709	0.0006	2.9147	2.8961	0.0204	1.2450
6-020	1:100	20°	12,000	1.3379	1.4014	0.0004	3.1410	3.1030	0.0000	1.2364
6-025	1:100	25°	12,000	1.2967	1.2908	0.0061	3.3722	3.3398	0.0158	1.2302
6-030	1:100	30°	12,000	…	…	…	3.4899	3.5031	0.0203	1.2220
6-035	1:100	35°	12,000	…	…	…	3.6090	3.6090	0.0000	1.2172
6-040	1:100	40°	12,000	…	…	…	3.6892	3.6959	0.0165	1.2208
6-045	1:100	45°	12,000	…	…	…	3.7702	3.7566	0.0210	1.2166
6-050	1:100	50°	12,000	…	…	…	3.8106	3.8106	0.0000	1.2226
6-055	1:100	55°	12,000	…	…	…	3.8516	3.8448	0.0167	1.2167
6-060	1:100	60°	12,000	…	…	…	3.8516	3.8584	0.0168	1.2150
6-075	1:100	75°	12,000	…	…	…	3.8926	3.9064	0.0213	1.2136
6-090	1:100	90°	12,000	…	…	…	3.9339	3.9408	0.0170	1.2141
6-105*	1:100	105°	12,000	…	…	…	4.2283	4.3071	0.0503	1.2081
6-120	1:100	120°	12,000	…	…	…	4.3142	4.3286	0.0223	1.2069
6-135	1:100	135°	12,000	…	…	…	4.3142	4.3286	0.0223	1.2066
6-150	1:100	150°	12,000	…	…	…	4.3142	4.3430	0.0353	1.2063
6-165	1:100	165°	12,000	…	…	…	4.3574	4.3718	0.0223	1.2063
6-180	1:100	180°	12,000	…	…	…	4.4007	4.4153	0.0226	1.2183

Note. All cases have a viscosity coefficient $\alpha = 0.01$. The first column contains the run code in the format X-YYY, where X's refer to the mass ratio, and YYY encodes the inclination angle. The other columns give the mass ratio q, inclination angle ι, total simulated time t_*, the radius where the viscous and dynamical torque densities balance $r_{\rm dT}$ at $t = 6000\delta$, the mean of $r_{\rm dT}$ over the subsequent 5000 binary revolutions $\tilde{r}_{\rm dT}$, the standard deviation of $r_{\rm dT}$ over that same time interval $\varsigma(r_{\rm dT})$, the radius of the density maximum $r_{\rm max}$ at $t = 6000\delta$, its mean over the subsequent 5000 binary revolutions $\tilde{r}_{\rm max}$, its standard deviation over the subsequent 5000 binary revolutions $\varsigma(r_{\rm max})$, and the radius where the density reaches 10% of the final density maximum $r_{10\%}$ at $t = 6000\delta$. Runs marked with a single asterisk (*) denote those that never reach a quasi-steady state, while those marked with double asterisks (**) denote runs that at first appear to reach a quasi-steady state over many orbits, but eventually depart from the quasi-steady state within a viscous timescale. All other cases persist in their quasi-steady state over a viscous timescale.

2. Theoretical Foundations

2.1. Analytical Preliminaries

As with our previous study, we work in units where the gravitational constant G, the combined mass of the binary M, and the semimajor axis of the binary a are all set to unity. With this in mind, we consider the gravitational potential experienced by test particles within the plane of a disk due to a binary system of point masses, such that the binary is inclined with respect to the disk by an angle ι. We again note that we only study the planar motion of the test particles.

Following R. Miranda & D. Lai (2015), we decompose the gravitational potential in the plane of the disk into azimuthal and temporal harmonics:

$$\Phi(r, \phi) = -\frac{1}{r} + \sum_{m,n} \Phi_{m,n}(r, \iota)\cos m\phi - nt, \tag{1}$$

where r is the radial distance, ϕ is the azimuthal angle within the plane of the disk, m and n are the azimuthal and temporal harmonic indices, and t is the time in units of the binary orbital period.

Compared to the coplanar case studied in Paper I, we note that an additional harmonic index has been introduced. This is to take into account *eccentric* modes, i.e., harmonic modes where the azimuthal frequency is offset from the temporal frequency of the mode due to noncircular behavior. Even when the binary in question is executing a circular orbit, the projection of an inclined orbit onto the plane of the disk is not circular, and therefore allows for eccentric harmonic modes to propagate and affect the disk dynamics.

Using the Wigner matrix formulation from quantum mechanics, we can decompose the gravitational potential within the plane of the disk into spherical harmonic as suggested in R. Miranda & D. Lai (2015) to get the following expression for the Fourier modes $\Phi_{m,n}$:

$$\Phi_{m,n}(r, q, \iota) = -2 \sum_{l=l_{\min}} \mathcal{Q}_l W_{l,m} W_{l,n} d^l_{n,m}(\iota) r^{-l-1}, \tag{2}$$

where l labels the index of the spherical harmonic, $\mathcal{Q}_l = (-\mu)^l(1-\mu) + \mu(1-\mu)^l$ is the multipole moment with μ being the ratio of the secondary mass to the total mass, $W_{l,m}$ is given in Equation (9) of R. Miranda & D. Lai (2015) and relates to the equatorial spherical harmonics, and $d^l_{n,m}$ is the Wigner "little" d matrix (N. Zettili 2009). As with the coplanar case, $l_{\min} = \max(2, m, n)$.

Importantly, we analyze the effect of the Fourier potential as a perturbation of circular orbits in the background Keplerian potential. The effect of the perturbation is to force epicyclic motion about the circular orbits. The amplitude of the epicycles for any Fourier mode is, in this case, given by

$$A_{m,n}(r_0) = \frac{2\frac{\Phi_{m,n}}{r_0}\left(1 - \frac{n}{m\omega}\right) + \partial_r \Phi_{m,n}}{|(m\omega - n)^2 - \omega^2|}, \tag{3}$$

where $\omega = r_0^{-3/2}$ is the Keplerian angular frequency, and r_0 is the radius of the background circular orbit; see Section 2.2 of Paper I for a detailed derivation of the epicyclic approximation.

In this study, we focus once again on the stability of the epicyclic motion as well as the strength of Lindblad torques in light of the disk-orbit inclination. In comparison to the coplanar case, we notice two additional components—the time harmonic in the cosine of the Fourier decomposition and the dependence of the strength of the potential (and, correspondingly, the amplitude of epicycles) on the inclination of the binary's orbit. The time harmonic introduces new resonances to the picture, which are conventionally referred to as *eccentric resonances*, since the $m \neq n$ modes are only present when the projection of the binary orbital plane onto the disk plane is noncircular. These resonances can be found farther away from the circular Lindblad resonances (i.e., when $m = n$), and therefore can occur at orbits farther from the typical gap scale where the instability timescale is much larger. On the other hand, the dependence of the epicyclic amplitude on inclination affects the magnitude of the Lyapunov exponents, and this alters the instability timescale. The strength of the potential is weakened for inclined systems, and this brings the gap closer to the binary, which can ultimately be used to explain the observed trend in the gap sizes from our numerical study.

We note here that, while the gravitational potential is written in terms of the radius and azimuthal angle, we neglect the role that polar angle or vertical displacement plays in shaping the dynamics of the test particles. This is to ensure that the analytical description matches the numerical setup of the disk hydrodynamics. Ideally, the inclined binary system will force particles to move toward the point masses, displacing the test particles from the disk plane. In the resonant torque picture, a new set of vertical resonances (S. H. Lubow & G. I. Ogilvie 1998) will need to be included in addition to the eccentric resonances discussed earlier. In the orbital stability picture, the dynamics of the test particles will have an additional degree of freedom resulting in an additional Lyapunov exponent that corresponds to instabilities driving the test particles out of the disk plane. The consequence of off-plane dynamics to the formation and maintenance of the circumbinary gap will be left for future studies.

2.2. Hydrodynamics Equations

In our setup, the only source of the Newtonian gravitational potential $\Phi(\boldsymbol{x})$ is two point masses, m_1 and m_2, moving on fixed circular orbits. The potential at an arbitrary point in three-dimensional space, represented by the vector $\boldsymbol{x}$, can be expressed as

$$\Phi = \frac{m_1}{\boldsymbol{x} - \boldsymbol{x}_1} + \frac{m_2}{\boldsymbol{x} - \boldsymbol{x}_2}, \tag{4}$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are the positions of the point masses.

The dynamics of the massless disk is governed by the standard equations of Newtonian hydrodynamics for mass (areal) density $\sigma(\boldsymbol{x})$ and fluid velocity $\boldsymbol{v}(\boldsymbol{x})$,

$$\partial_t \sigma + \nabla \cdot (\sigma \boldsymbol{v}) = 0, \tag{5}$$

$$\partial_t \boldsymbol{v} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\frac{1}{\sigma}\nabla p - \nabla\Phi + \boldsymbol{f}_\nu, \tag{6}$$

where $p(\boldsymbol{x})$ represents the pressure, and the viscous force $\boldsymbol{f}_\nu(\boldsymbol{x})$ is given by

$$\boldsymbol{f}_\nu = \nabla \cdot (\nu \nabla \mathbf{v}) + \nabla\left(\frac{1}{2}\nu \nabla \cdot \mathbf{v}\right). \tag{7}$$

To complete the system of equations and make it solvable, an equation of state and a prescription for the viscosity are also required (J. Frank et al. 2002).

2.3. Numerical Setup

The numerical calculations are performed using the open-source numerical code DISCO (P. C. Duffell 2016). This code is designed for grid-based, three-dimensional, moving mesh simulations and is specifically tailored to solve Newtonian MHD problems in axial symmetry. The code's architecture allows for easy reduction of dimensions to handle two-dimensional problems. The moving mesh capability enables the user to optimize accuracy and computational efficiency by selecting the angular velocity of the grid, thereby minimizing diffusive advection errors.

In our study, we consider a three-dimensional configuration consisting of a black hole binary and a thin, viscous, and massless disk inclined with respect to the binary. The first component of our system is the black hole binary, which is modeled using point masses within the Newtonian framework. We define a "primed" polar coordinate system, denoted as (r', ϕ'), centered on the center of mass of the binary and coplanar with the binary's orbit. In this coordinate system, the gravitational potential, Equation (4), takes the form

$$\Phi(r') = \frac{Gm_1}{r' - r_1'} + \frac{Gm_2}{r' - r_2'}, \tag{8}$$

where r_1' and r_2' are determined by the first model parameter, the mass ratio $q = m_1/m_2$ (also denoted $m_1{:}m_2$ in the text); specifically, they are given by $r_1' = q/(1+q)$ and $r_2' = 1/(1+q)$.

The second component of our system is the two-dimensional viscous disk. We define a cylindrical coordinate system (r, ϕ, z) with the origin at the center of the binary, where the surface $z = 0$ is tangential to the orbital plane of the binary at an angle ι. In this configuration, the disk exists solely on the surface $(r, \phi, z = 0)$, and all hydrodynamics quantities are evaluated on this surface. To describe the viscous force, we adopt the α-*type viscosity* following the original approach of N. I. Shakura & R. A. Sunyaev (1973). Here, the viscosity is given by $\nu(r) = \alpha s^2/r^{3/2}$, where α is a constant, $s(r) = \chi/r^{1/2}$ is the speed of sound, and $\chi = h/r$ represents the constant ratio of the pressure scale height to the radius. For our simulations, we choose $\chi = 0.1$.

The focus on two-dimensional simulations in this work was primarily motivated by the computational cost of three-dimensional simulations. While we recognize that three-dimensional simulations could capture additional phenomena like disk warping, which could impact the evolution of disk parameters in the midplane, our preliminary investigation in two dimensions provided evidence in favor of our proposed gap opening mechanisms. The discovery of an unstable sector and the behavior of torques at particular configurations has given insight into areas that could be further explored with three-dimensional simulations in the future.

The system of Equations (5) and (6) consists of four primitive variables: the surface mass density σ, the pressure p, and the two components of the fluid's velocity vector $\boldsymbol{v} = (v^r, r\Omega)$, all of which depend on the cylindrical coordinates r and ϕ. Additionally, the pressure is related to the density through the locally isothermal equation of state, given by $p = s^2\sigma$. Figure 1 illustrates the meaning of the *inclination* angle ι as the angle between the disk and binary planes.

We utilize the two-dimensional version of the open-source Riemann solver, DISCO (P. C. Duffell 2016), specifically the HLLC version (E. F. Toro 2013), to solve the field equations for the four primitive variables: density $\sigma(t, r, \phi)$, pressure $p(t, r, \phi)$, radial linear velocity $v^r(t, r, \phi)$, and angular velocity $\Omega(t, r, \phi)$. This choice is suitable due to the thinness of the disk. As the point masses are described in the primed coordinates (r', ϕ'), it is necessary to transform their positions to the unprimed coordinates (r, ϕ, z) using the following coordinate transformations:

$$\begin{aligned} r_i &= r_i'(\cos^2\phi_i' + \sin^2\phi_i'\cos^2\iota)^{1/2}, \\ \phi_i &= \arctan(\tan\phi_i'\cos\iota), \\ z_i &= -r_i'\sin\phi_i'\sin\iota. \end{aligned} \tag{9}$$

Here, $i \in \{1, 2\}$ corresponds to the first and second mass, respectively. It is important to note that, although our model is limited to a two-dimensional problem, we include the three-dimensional information of the point masses' position (r, ϕ, z) when calculating the gravitational potential and evaluating the hydrodynamical equations in the two-dimensional setup at the surface $(r, \phi, z = 0)$.

We adopt the binary separation a as the unit of length. Our computational domain is defined as the region $S = S_{100a} \backslash S_a$, where $S_{100a} = r \times \phi$: $r \leqslant 100a \wedge \phi \in \langle 0, 2\pi)$ is a filled circular region from which we exclude the subdomain $S_a = r \times \phi$: $r < a \wedge \phi \in \langle 0, 2\pi)$. The purpose of excluding the inner region S_a is to optimize computational efficiency, since we are focusing solely on the circumbinary disk. However, it is worth mentioning that future studies will aim to incorporate this excluded region in order to expand the scope of our investigation. For reference, Figure 2 provides a visual comparison between the case labeled as $i00q1$ from Table 1 (equal mass, not inclined) and the corresponding simulation where the input parameters are the same, but the region $r < a$ is included in the computational domain.

The numerical grid used in our simulations incorporates moving mesh capabilities, allowing it to rotate with an angular velocity of $\Omega_{\rm grid} = r^{-3/2}$. The grid is composed of nearly square-shaped cells, with an aspect ratio close to unity. We introduce the term "ring" to refer to a group of grid cells that are equidistant from the center. Our grid consists of a total of $n = 480$ radial rings, with each ring containing a varying number of cells denoted by n_j, where the index "j" corresponds to the radial ring. The width of the rings changes with radius, ranging from approximately 0.034 for the innermost ring to approximately 0.790 for the outermost ring. Explicitly, the width of the nth ring is given by the difference between two neighboring nodes $\Delta r_n = r_{n+1} - r_n$, where

$$r_n = 1 + A\sinh(BC), \tag{10}$$

with

$$A = \frac{5}{\sinh(1)}, \quad B = \frac{n-1}{480}, \quad C = \sinh^{-1}\left(\frac{99}{A}\right). \tag{11}$$

This particular choice ensures that variations in the density are adequately resolved across multiple cells. The time intervals employed in our simulations vary between time steps and are calculated as half of the minimal propagation time that may occur across the entire grid. The timescale is expressed in terms of the binary revolution period δ and is related to the viscous

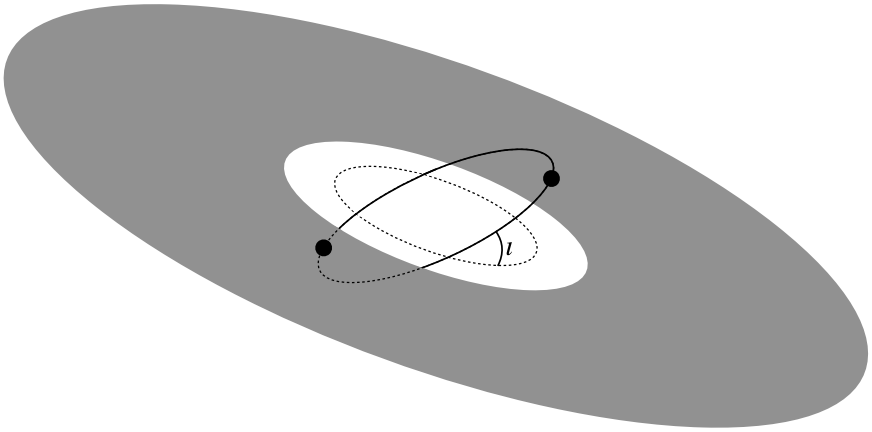

Figure 1. The configuration we are considering, where the orbital plane of the binary is inclined by an angle ι relative to the plane of the disk.

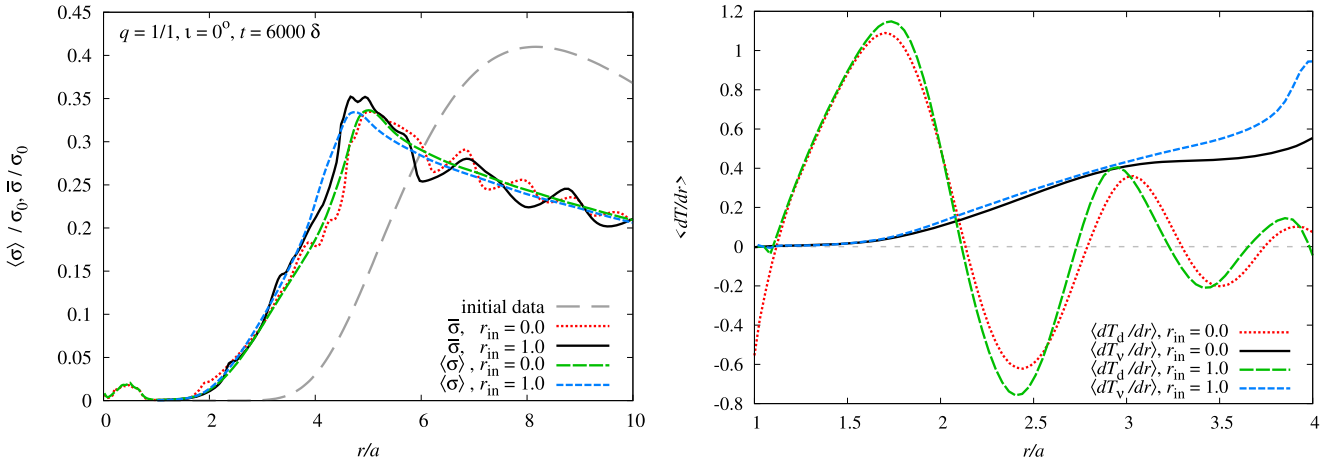

Figure 2. The plots show the comparison of our standard approach where we intentionally, due to computational cost, excluded the inner region $r < a$ from our numerical domain (red curves), and a test run without any limits of this kind (green curves). The left plot shows the instantaneous (solid lines) and averaged mass density distributions (dotted lines) for both approaches. The right plot shows that the location where the dynamical (solid lines) and viscous torque densities (dotted lines) intersect does not depend on the choice for the inner boundary. Both plots present a stage of the evolution at $t = 6000\delta$.

timescale by the well-known approximate formula

$$t_\nu \approx \frac{r^2}{3\nu} = \frac{1600\delta}{2^{3/2}} r^{3/2}. \tag{12}$$

The majority of our results are expressed in terms of azimuthally and time-averaged quantities. To obtain an exclusively r-dependent averaged function $\langle\xi\rangle(r)$ from a general variable $\xi(t, r, \phi)$, we perform the following procedure:

$$\langle\xi\rangle(r) = \frac{1}{2\pi\Delta}\int_t^{t+\Delta}\int_0^{2\pi} \xi(t, r, \phi)\, d\phi dt \tag{13}$$

where Δ is the time interval over which the quantity is averaged. From a numerical perspective, let us introduce the notation ξ_{ijk}, which corresponds to the variable ξ evaluated at the ith time sample, the jth radial ring, and the kth azimuthal cell. The averaging procedure can then be written as

$$\langle\xi\rangle_j = \frac{1}{\kappa n_\delta n_j}\sum_{i=m}^{m+\kappa n_\delta}\sum_{k=1}^{n_j}\xi_{ijk} \tag{14}$$

where we introduce additional parameters for mathematical correctness and transparency: κ represents the number of time samples per binary revolution, n_δ denotes the number of binary revolutions over which we average, and m represents the number of binary revolutions at which we start the averaging procedure. In our case, we choose $\Delta = 50\delta$, so $n_\delta = 50$ and $\kappa = 60$. This results in 3000 time samples for 50 binary revolutions, which corresponds to one sample for each 6° of the binary's circular motion.

2.4. Initial Data and Boundary Conditions

Following the approach of M. Milosavljević & E. S. Phinney (2005) and D. J. D'Orazio et al. (2013), we initialize our primitive variables for a locally isothermal equation of state with a surface density that is peaked around $r = 8a$ and falls off significantly at $r < 3a$. The orbital frequency and radial velocity are initialized taking into account the pressure gradient and viscous drift, respectively, so that the initial surface density, pressure, orbital frequency, and radial velocity are given by

$$\sigma(t = 0, r) = \sigma_0\left(\frac{r_s}{r}\right)^3 \exp\left(-\frac{r_s}{r}\right)^2, \tag{15}$$

$$p(t = 0, r) = s^2\sigma, \tag{16}$$

$$\Omega(t = 0, r) = \sqrt{\Omega_K^2\left[1 + \frac{3}{4}\frac{a^2}{r^2}\frac{q}{(q+1)^2}\right]^2 + \frac{\partial_r p}{r\sigma}}, \tag{17}$$

$$v_r(t=0,\, r) = \frac{2\partial_r(r^3\nu\sigma\partial_r\Omega)}{r^2\Omega\sigma}. \tag{18}$$

Here, σ_0 and r_s are constants that determine the value and location of the initial density peak; s is the sound speed, while $\Omega_K = r^{-3/2}$ represents the Keplerian angular velocity. It is important to note that this initial data incorporate the mass ratio parameter q but do not include the inclination ι.

The boundary conditions are imposed on the two innermost and outermost radial rings. To apply these conditions, we only need azimuthally averaged quantities:

$$\bar{\xi}(t,\, r) = \frac{1}{2\pi}\int_0^{2\pi} \xi(t,\, r,\, \phi)\, d\phi. \tag{19}$$

In a discretized form, this becomes

$$\bar{\xi}_{ij} = \frac{1}{n_j}\sum_{k=1}^{n_j} \xi_{ijk}. \tag{20}$$

The outer boundary conditions are simply the initial data for all primitive variables. For $j \in \{n-1,\, n\}$, we have

$$\begin{aligned} \sigma_{ij} &= \sigma_{0j} = \sigma_j(t=0), \\ p_{ij} &= p_{0j} = p_j(t=0), \\ v^r_{ij} &= v^r_{0j} = v^r_j(t=0), \\ \Omega_{ij} &= \Omega_{0j} = \Omega_j(t=0). \end{aligned} \tag{21}$$

On the other hand, the inner boundary conditions for $j \in \{1, 2\}$ are as follows:

$$\begin{aligned} \sigma_{ij}(t) &= \bar{\sigma}_{i-1,j+1}(r_{j+1}/r_j)^{-1/2}, \\ p_{ij}(t) &= \bar{p}_{i-1,j+1}(r_{j+1}/r_j)^{-1/4}, \\ v^r_{ij} &= v^r_{0j} = v^r_j(t=0), \\ \Omega_{ij} &= \Omega_{0j} = \Omega_j(t=0). \end{aligned} \tag{22}$$

The index "i" corresponds to the number of time steps, where "$i=0$" represents the initial data, and "i-1" represents the state immediately preceding time step "i." Similarly, the index "j" corresponds to the number of radial rings, so "j+1" corresponds to the next neighboring radial ring. It is important to note that the "i" index here denotes time steps dictated by the solving procedure, rather than the number of time samples used for data postprocessing. This second order numerical approach requires boundary conditions to be imposed onto two radial rings due to the fact that the Riemann solver uses the cells' *faces*, not nodes or central points. More details about this problem can be found in references (P. C. Duffell & A. I. MacFadyen 2011; P. C. Duffell 2016).

3. Numerical Results

3.1. Quasi-steady State

We considered 114 configurations characterized by the viscosity coefficient $\alpha = 0.01$, mass ratios $q \in \{1:1,\, 2:3,\, 3:7,\, 1:4,\, 1:10\}$, and inclinations $\iota \in \{0°,\, 15°,\, 20°,\, 25°,\, 30°,\, 35°,\, 40°,\, 45°,\, 50°,\, 55°,\, 60°,\, 75°,\, 90°,\, 105°,\, 120°,\, 135°150°,\, 165°,\, 180°\}$. They are presented in Table 1. In addition, we simulated 30 additional configurations with different viscosity coefficients, namely, $\alpha = 0.03$ and $\alpha = 0.003$, ranging over a smaller sampling of mass ratios $q \in \{1:1,\, 1:4,\, 1:10\}$ and inclinations $\iota \in \{0°,\, 45°,\, 90°,\, 135°,\, 180°\}$. These additional configurations are presented in Table 2, and give us 144 configurations in total.

There is no rigorous definition in the literature for the *quasi-steady state*. Our definition is based on the behavior of the averaged density $\langle\sigma\rangle$, which is computed using Equation (14). For brevity, we drop the term "averaged" and refer to it simply as the "density" later in the text. We also drop the index "j," keeping in mind that the numerical data are discretized for each radial ring. For each configuration, we analyze the state of the system at time intervals spaced by 1000δ (250δ for $\alpha = 0.03$ cases). We introduce here the following notation: $t_m = m \times 1000\delta$ ($m \times 250\delta$ for $\alpha = 0.03$ cases), the time at which we sample the data, with $m \in \mathrm{IN}$; $\langle\sigma\rangle_m^{\max}$, the maximum value of the density at $t = t_m$; and $r_m^{\max}$, the value of r at which the density reaches its maximum at $t = t_m$.

Unless otherwise noted, we chose $m = 6$ to report our results, corresponding to a time $t = 6000\delta$. Additionally, we denote the means of essential quantities over the time samples $m \in \{6, 7, 8, 9, 10, 11\}$ with a tilde symbol. The corresponding standard deviations over that same interval are denoted by $\varsigma()$. Furthermore, we use t_* to represent the total number of binary revolutions over which we investigated each system. The summary of these results is presented in Table 1.

According to our definition of quasi-steady state, the difference between $r_6^{\max}$ and the average taken from the locations of the maxima $\tilde{r}^{\max} = \frac{1}{6}\sum_{i=6}^{11} r_i^{\max}$ should not exceed the distance between two neighboring grid nodes, denoted as ϵ, i.e.,

$$|\tilde{r}^{\max} - r_6^{\max}| \leqslant \epsilon. \tag{23}$$

The definition of quasi-steady state can vary significantly in the literature. In A. I. MacFadyen & M. Milosavljević (2008), density profiles were observed to stabilize over 4000 orbits, which we applied to Paper I. However, we observed that the time required for the system to reach a quasi-steady state differed significantly over inclinations. The above, mathematically rigorous definition has been successful and can work with different values of the viscosity parameter, which is responsible for the timescale of the disk evolution.

Given that all the maxima are situated between $r = 2.0$ and $r = 5.0$, the maximum distance between two neighboring nodes, $\epsilon = 0.047$, accounts for less than 3% of the r value. We observed that certain cases, marked with an asterisk or double asterisk in Table 1, do not conform to our definition of approaching a quasi-steady state.

All cases with $\iota \leqslant 30°$ and $\iota \geqslant 55°$, as indicated in Table 1, satisfy the aforementioned condition from Equation (23). Additionally, these cases exhibit density maxima that either are stable or undergo movement/oscillation within one or two zones over 1000 binary revolutions. On the other hand, cases marked with an asterisk in the range of $35° \leqslant \iota \leqslant 50°$ demonstrate a relatively stable peak location, although it oscillates over more than two zones throughout the 1000 binary revolutions. Lastly, the cases marked with a double asterisk do not exhibit stable peaks at all, and the density distribution evolves throughout the entire simulation, which lasts for multiple viscous timescales at the radii of greatest interest.

Even if the condition from Equation (23) is satisfied for some configurations with $m < 6$, we do not consider time

Table 2
Summary of Configurations with Different Viscosity Coefficients

Run's Code	q	ι	t_*	$r_{\rm dT}$	$\tilde{r}_{\rm dT}$	ς $(r_{\rm dT})$	$r_{\rm max}$	$\tilde{r}_{\rm max}$	$\varsigma(r_{\rm max})$	r_{10} %
0.03-1-000*	1:1	0°	6000	1.9903	1.9764	0.0085	3.8927	3.9133	0.0226	2.0158
0.03-1-045	1:1	45°	6000	1.8822	1.8895	0.0071	4.3573	4.3791	0.0238	1.8850
0.03-1-090	1:1	90°	6000	1.5862	1.5860	0.0002	2.5471	2.5471	0.0000	1.3653
0.03-1-135	1:1	135°	6000	1.4503	1.4496	0.0007	4.1856	4.2282	0.0270	1.2135
0.03-1-180	1:1	180°	6000	⋯	⋯	⋯	4.7567	4.8023	0.0408	1.2135
0.03-4-000*	1:4	0°	6000	1.8991	1.8952	0.0059	3.6089	3.6023	0.0163	1.8324
0.03-4-045	1:4	45°	6000	1.7173	1.7158	0.0023	3.5294	3.5228	0.0161	1.5004
0.03-4-090	1:4	90°	6000	1.5251	1.5252	0.0001	3.9754	3.9754	0.0000	1.3478
0.03-4-135	1:4	135°	6000	1.4337	1.4338	0.0003	3.9340	3.9616	0.0214	1.2204
0.03-4-180	1:4	180°	6000	⋯	⋯	⋯	4.7114	4.7340	0.0248	1.2204
0.03-5-000*	1:10	0°	6000	1.7258	1.7251	0.0081	3.2175	3.2240	0.0289	1.5742
0.03-5-045	1:10	45°	6000	1.5568	1.5568	0.0000	3.3723	3.3983	0.0202	1.3103
0.03-5-090	1:10	90°	6000	1.4276	1.4276	0.0002	4.1009	4.1009	0.0000	1.2635
0.03-5-135	1:10	135°	6000	1.3859	1.3859	0.0000	4.0589	4.0869	0.0217	1.2197
0.03-5-180*	1:10	180°	6000	⋯	⋯	⋯	4.5769	4.6290	0.0337	1.2197
0.003-1-000	1:1	0°	24,000	2.1439	2.1537	0.0097	5.0808	5.0965	0.0243	2.5877
0.003-1-045*	1:1	45°	24,000	2.1471	1.9636	0.1073	4.8479	4.9875	0.1141	2.4247
0.003-1-090	1:1	90°	24,000	1.7018	1.7015	0.0007	2.3672	2.3672	0.0000	1.6434
0.003-1-135	1:1	135°	24,000	1.5761	1.5794	0.0057	2.4030	2.4030	0.0000	1.3441
0.003-1-180	1:1	180°	24,000	1.2282	1.2334	0.0047	2.4750	2.4869	0.0185	1.3358
0.003-4-000	1:4	0°	24,000	2.0817	2.0819	0.0023	4.5771	4.5696	0.0182	2.3912
0.003-4-045*	1:4	45°	24,000	2.4531	2.3322	0.1556	4.0169	3.9552	0.1194	2.1754
0.003-4-090	1:4	90°	24,000	1.5736	1.5735	0.0014	4.6662	4.6662	0.0000	1.7244
0.003-4-135	1:4	135°	24,000	1.7151	1.7095	0.0065	2.9897	2.9584	0.0153	1.4500
0.003-4-180	1:4	180°	24,000	⋯	⋯	⋯	2.5472	2.5291	0.0199	1.3357
0.003-5-000	1:10	0°	24,000	1.9492	1.9549	0.0036	3.8106	3.8175	0.0167	2.0753
0.003-5-045*	1:10	45°	24,000	1.6155	1.7187	0.1012	4.3142	4.2435	0.1487	1.6343
0.003-5-090	1:10	90°	24,000	1.5747	1.5713	0.0020	4.1432	4.1362	0.0173	1.4968
0.003-5-135	1:10	135°	24,000	1.6248	1.6250	0.0045	2.9147	2.9147	0.0000	1.4014
0.003-5-180	1:10	180°	24,000	⋯	⋯	⋯	2.4390	2.4570	0.0197	1.3324

Note. The first column contains the run code now in a Z-X-YYY format, where Z is the viscosity coefficient α, and X and YYY remain unchanged from Table 1. The other columns are the same as in Table 1.

samples prior to $m = 6$ to ensure that, once $\langle\sigma\rangle_m^{\rm max}$ is reached, the value of the density maximum does not increase. This precaution is taken because a quasi-steady state can still involve mass loss at the inner boundary or due to truncation error. However, this should not significantly impact the density, particularly near the peak.

The two plots on the left-hand side of Figure 3 demonstrate the application of the aforementioned definition using examples. The upper left plot depicts the evolution of the maximum and its eventual stabilization at a specific time. Conversely, the lower left plot emphasizes that not only the location of the maximum is significant but also the value of the density and its temporal evolution is significant. While the location of the maximum is established early on, the matter continues to flow toward the center and accumulate. Furthermore, during the initial stages of the evolution, the maxima are local. For instance, one can look at the blue lines in the right panels for t_1 and the dark-blue line for t_2 as illustrative examples.

The same single- and double-asterisk notation is applied for the additional cases of different viscosity; however, the timescales differ. In the case of $\alpha = 0.03$, we report solutions after $t = 3000\delta$, while, for $\alpha = 0.003$, we report them after $t = 12{,}000\delta$.

3.2. Density Distribution

We analyze the density distributions by focusing on two characteristic radii, denoted $r_{\rm max}$ and $r_{\rm dT}$. $r_{\rm max}$ corresponds to the location of the density maximum. We define $r_{\rm dT}$ as the radius where the averaged viscous torque density $\frac{dT_\nu}{dr} = \frac{d}{dr}\left(2\pi r^3 \nu \langle\sigma\frac{d\Omega}{dr}\rangle\right)$ and the averaged dynamical torque density $\frac{dT_{\rm d}}{dr} = -2\pi r\langle\sigma\frac{d\Phi}{d\phi}\rangle$ balance each other:

$$r_{\rm dT} = r\colon \frac{dT_{\rm d}}{dr} = \frac{dT_\nu}{dr}. \tag{24}$$

We observe that $r_{\rm dT}$ fluctuates/oscillates between time samples, so we report not only the radii $r_{\rm dT}$ at $t = t_6$ but also the average

$$\tilde{r}_{\rm dT} = \frac{1}{6}\sum_{m=6}^{11} r_{\rm dT}^m. \tag{25}$$

We present the variations observed in the above radii in two ways. We first present the variation in the mass density distribution for different inclinations at fixed mass ratios. Second, we present the variation of the mass density distribution for different mass ratios at fixed inclination angles.

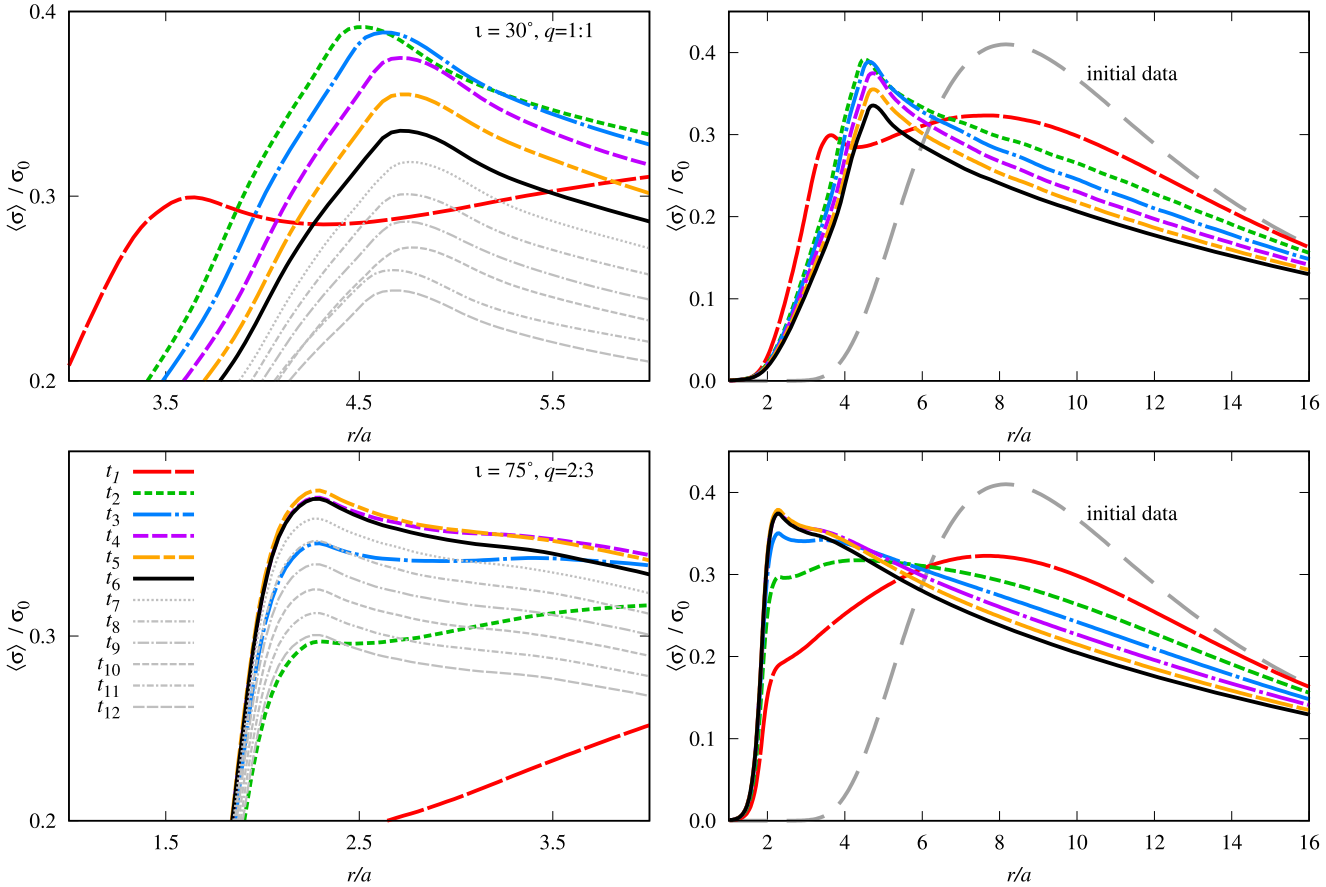

Figure 3. The averaged mass density as a function of radius is shown for two arbitrarily chosen cases: $q = 1/1$, $\iota = 30°$ (upper row) and $q = 2/3$, $\iota = 75°$ (lower row). In the left panels, the colored lines represent times prior to reaching the quasi-steady state. The final color curve (black) represents the density distribution at the quasi-steady state, specifically at $t = 6000\delta$, while the gray lines with different patterns correspond to later stages. The left panels are zoomed in on the regions closer to the peak density, while right panels present the same data over a wider range of radii to show the evolution of the averaged density distribution throughout the disk.

Figure 4 presents a comparison between azimuthally averaged, instantaneous density profiles ("snapshots") and the "quasi-steady-state" density profile, which is averaged both azimuthally and over time. The averaging is done over a period of 50 binary revolutions for viscosity $\nu = 0.01$, which is chosen to accurately depict the quasi-steady behavior while smoothing out fluctuations in $\langle\sigma\rangle(r)$. This is consistent with A. I. MacFadyen & M. Milosavljević (2008) and D. J. D'Orazio et al. (2013), who adopted averaging intervals of 45 and 50 binary revolutions, respectively.

The study of the density distribution for $\iota \in \{0°, 15°, 30°, 45°, 60°, 75°, 90°, 105°, 120°, 135°, 150°, 165°, 180°\}$ shows an interesting dependence on the inclination angle. Plots in Figure 5 present the locations of $r_{\max}$ and $r_{\rm dT}$, along with the density distributions. They suggest the existence of at least two ranges of inclination angles in which the system behaves in two distinct ways. We first focus on equal-mass cases where the effects are the strongest. In the low inclination sector with ι less than or equal to 45°, the density increases moderately with increasing radius, and at a distance of approximately four binary separations from the center, it approaches a maximum value close to the initial data maximum (red, green, dark-violet, and blue curves on the upper left panel). In the case of high inclinations with ι greater than 45°, the density has a very steep gradient inside the maximum and grows to values substantially higher than the initial data maximum. In these cases, the locations of $r_{\max}$ are roughly half of their values for the low inclinations (dark-green, orange, and violet curves on the upper left panel). This simple division into two regimes of inclinations begins to blur, but is still identifiable, as we decrease the mass ratio. In Figure 4, on the left, we can see the azimuthally and time-averaged distributions, and on the right, only the azimuthally averaged distribution at a certain time. The observed division strongly manifests in the equal and nearly equal mass cases, but closer to the extreme case of a binary with a central mass and a satellite, it is significantly weaker but still noticeable. As a matter of fact, we observe intermediate states as we vary the mass ratio from the equal-mass value of $q = 1/1$ to the lowest considered value of $q = 1/10$.

This division into two domains of behavior is further muddied by the fact that there were configurations, for example 1–045 and 4–045, that were unable to reach a quasi-steady state. Following this observation, we decided to look closer and simulate additional systems with $\iota \in \{20°, 25°, 35°, 40°, 50°, 55°\}$. The majority of the additional configurations reached a quasi-steady state according to our requirements, and only some cases, those between $35° \leqslant \iota \leqslant 50°$, failed to do so. In principle, a larger inclination makes the quasi-steady state more difficult to reach, because the time needed to fulfill the condition from Equation (23) is getting longer. However, for inclinations at or above 55°, all configurations are equally well behaved and clearly reach a quasi-steady state. We present a sample of those density plots in Figure 4.

This motivates us to make a more general division where we have *moderately inclined* cases with $\iota \in \langle 0°, 30°\rangle$, *highly inclined* cases with $\iota \in \langle 55°, 90°\rangle$, and, for completeness,

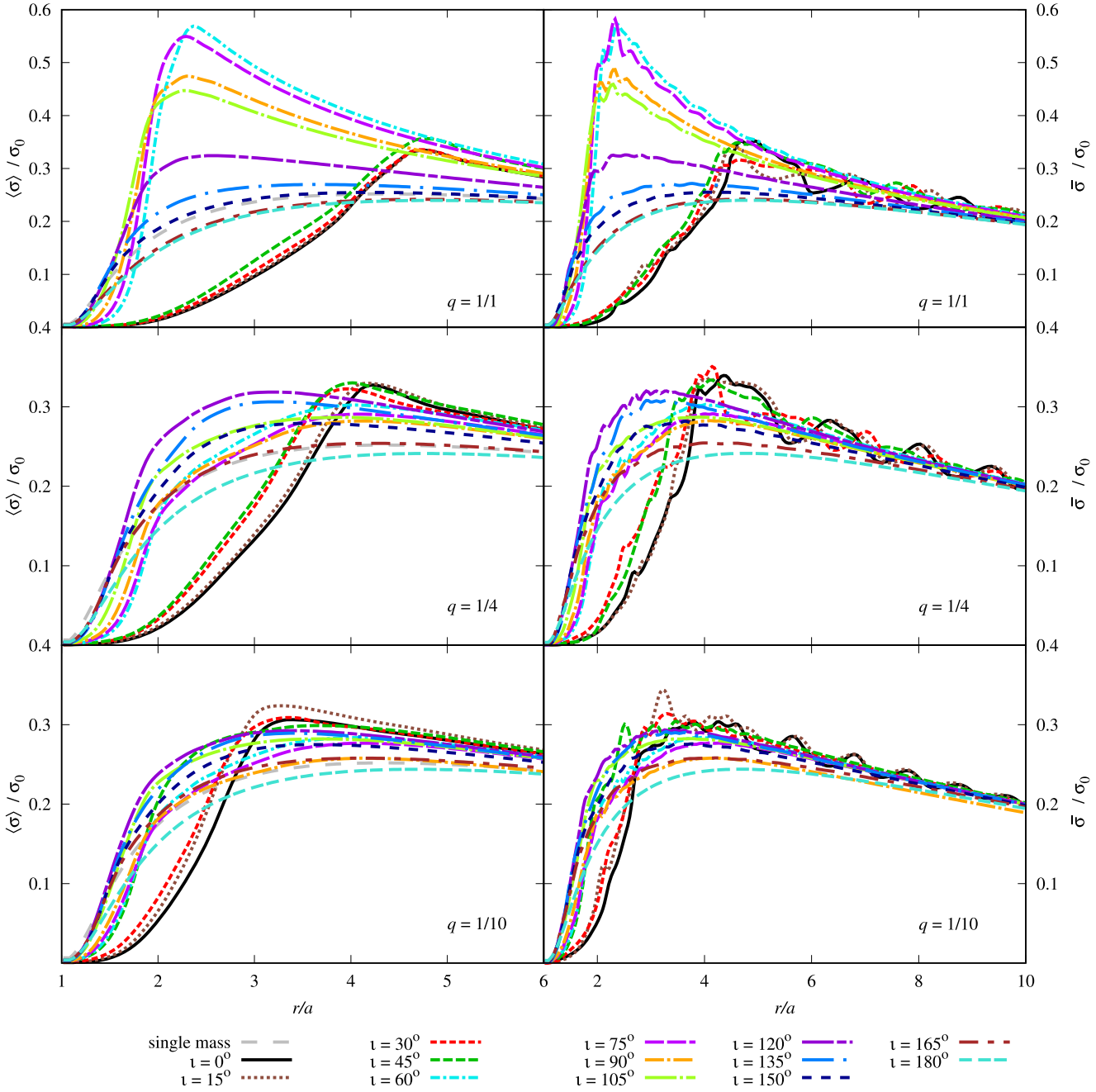

Figure 4. Left column: an averaged density distribution in the central part of the system for cases with $q = 1, 1/4, 1/10$. Right column: instantaneous density (averaged over the azimuthal angle) for the same cases.

counterrotating cases with $\iota \in (105°, 180°\rangle$. We refer to the subset of cases with $\iota \in \langle 0°, 30°\rangle \cup \langle 55°, 180°\rangle$ as the *stable sector* because they reach the quasi-steady state in a relatively short time. On the other hand, we refer to the configurations with $\iota \in \langle 35°, 50°\rangle$ as the *unstable sector*. Some of these cases do not settle into a quasi-steady state; others have reached this state in the sense of the condition from Equation (23), but could still potentially become unstable over longer timescales. To explore this possibility, we continued to evolve these cases for a much longer time, up to 48,000 binary revolutions (see Tables 1 and 2), with varying results in terms of stability.

To illustrate the dynamics of these two sectors, we compare two cases, both with $q = 1/4$, but one with inclination $\iota = 60°$ (in the stable sector) and the other with $\iota = 45°$ (in the unstable sector). The upper left panel of Figure 6 confirms the fact that the density peak is not moving for $\iota = 60°$. On the other hand, the behavior for $\iota = 45°$ presented in the upper right panel is quite different, with the location of the density peak continuing to evolve steadily. In addition, the torque balance presented in the bottom row reinforces this picture, since the dynamical torque in particular continues to change substantially. The reason for this varying behavior is unclear to us, and will be the subject of future studies.

The differences in the density distribution are noticeable not only in the average density but also in single time samples. We analyzed two-dimensional snapshots of the density profiles, and found that the division between not inclined/moderately inclined and highly inclined configurations also manifests in the spatial pattern created

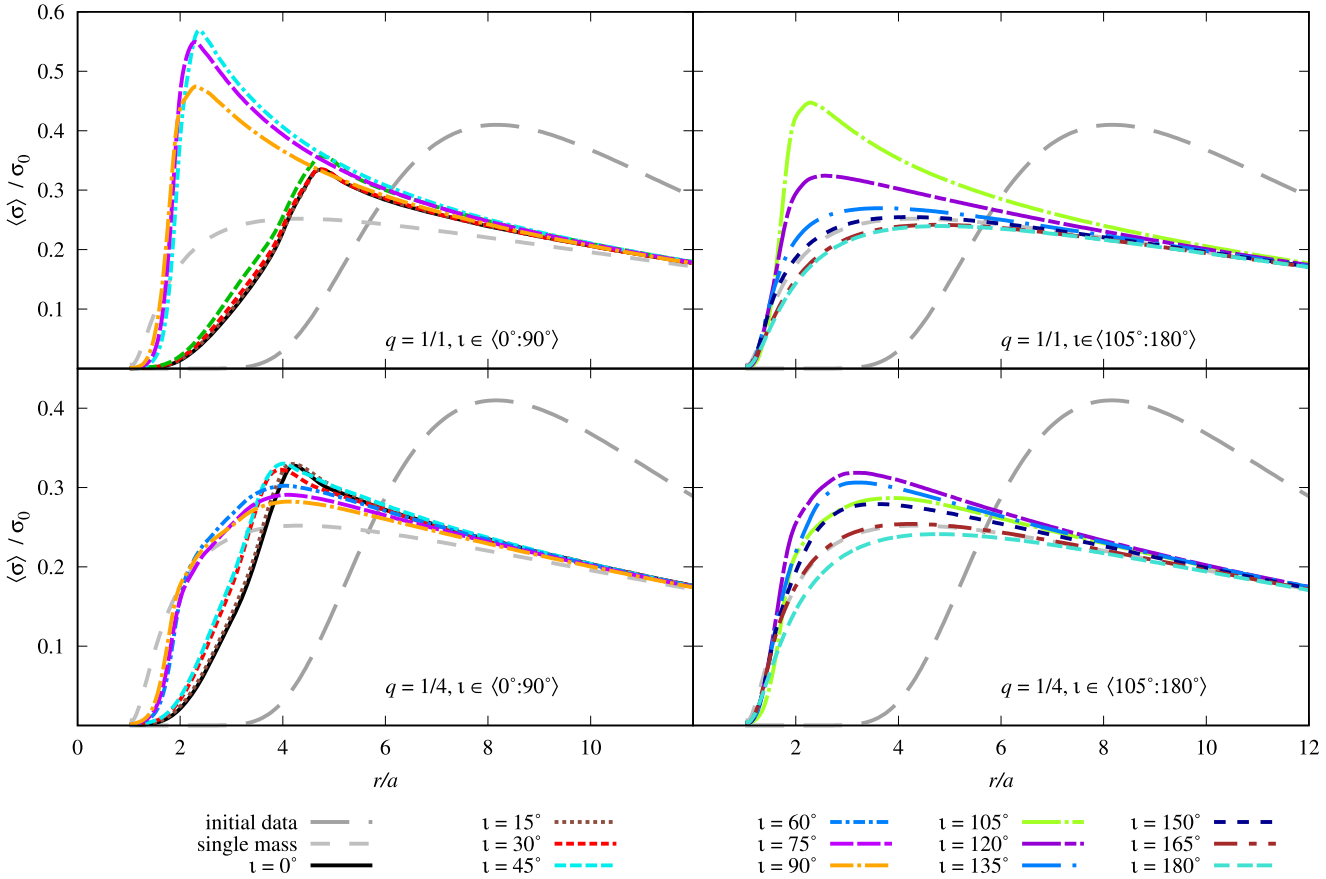

Figure 5. Averaged density distribution in the central part of the system for the cases with $q = 1$ and $q = 1/4$ together with the initial data and single mass case for reference.

by the fluid. For configuration with $\iota \leqslant 45°$, the density distribution forms the characteristic concentric and closed rings composed of the density's local maxima/minima, as in the upper left panel of Figure 7. On the other hand, the *highly inclined* cases, as represented in the upper right panel, form a spiral pattern instead of rings. As we go to still larger inclinations, through the perpendicular and into the counterrotating scenario, we see that the matter distribution in the disk becomes increasingly similar to the single mass case, and the effect of the periodic changes in the gravitational potential is not mirrored in the density distribution.

Our simulations focus on the mutual inclination between the disk and a unequal-mass binary. However, it may be useful to compare our baseline coplanar, equal-mass configuration with the results from previous studies, particularly the three-dimensional magnetohydrodynamical simulations reported by J.-M. Shi et al. (2012) and the fully relativistic MHD simulations by S. C. Noble et al. (2012). These studies provide valuable one-dimensional and two-dimensional profiles of the density and torque distributions, offering a useful reference for interpreting our results.

As an additional illustration of these two inclination-dependent regimes, we study the azimuthal mode decomposition of the density distribution,

$$D'_m = \frac{1}{2\pi^2}\int_0^{2\pi} d\phi \int_0^{2\pi} d(\Omega_b t)\,\rho e^{im(\phi-\Omega_b t)}, \quad \text{if } m \neq 0 \tag{26}$$

$$D_0 = \frac{1}{4\pi^2}\int_0^{2\pi} d\phi \int_0^{2\pi} d(\Omega_b t)\,\rho, \tag{27}$$

which are numerically calculated as

$$D_m = \frac{D'_m}{D_0} = \frac{\sqrt{\left[\frac{1}{n_r n_t}\Sigma_{j=0}^{n_t}\Sigma_{i=0}^{n_r} \quad \rho_i \cos(m(\phi_i - t))\right]^2 + \left[\frac{1}{n_r n_t}\Sigma_{j=0}^{n_t}\Sigma_{i=0}^{n_r} \quad \rho_i \sin(m(\phi_i - t))\right]^2}}{\frac{1}{n_r}\Sigma_{j=0}^{n_t}\Sigma_{i=0}^{n_r}\rho_i}. \tag{28}$$

We present the results of the azimuthal mode decomposition in Figure 8. As in Paper I, we observe that there is no unambiguous decay with increasing mode number beyond $m = 1$ at any length scale across the next three m modes. Furthermore, odd-m modes at $q = 1/1$ should be zero since the corresponding harmonic of the gravitational potential is also zero. This is indicative of the fact that, as in the coplanar case, the evolution of the azimuthal modes of the density distribution

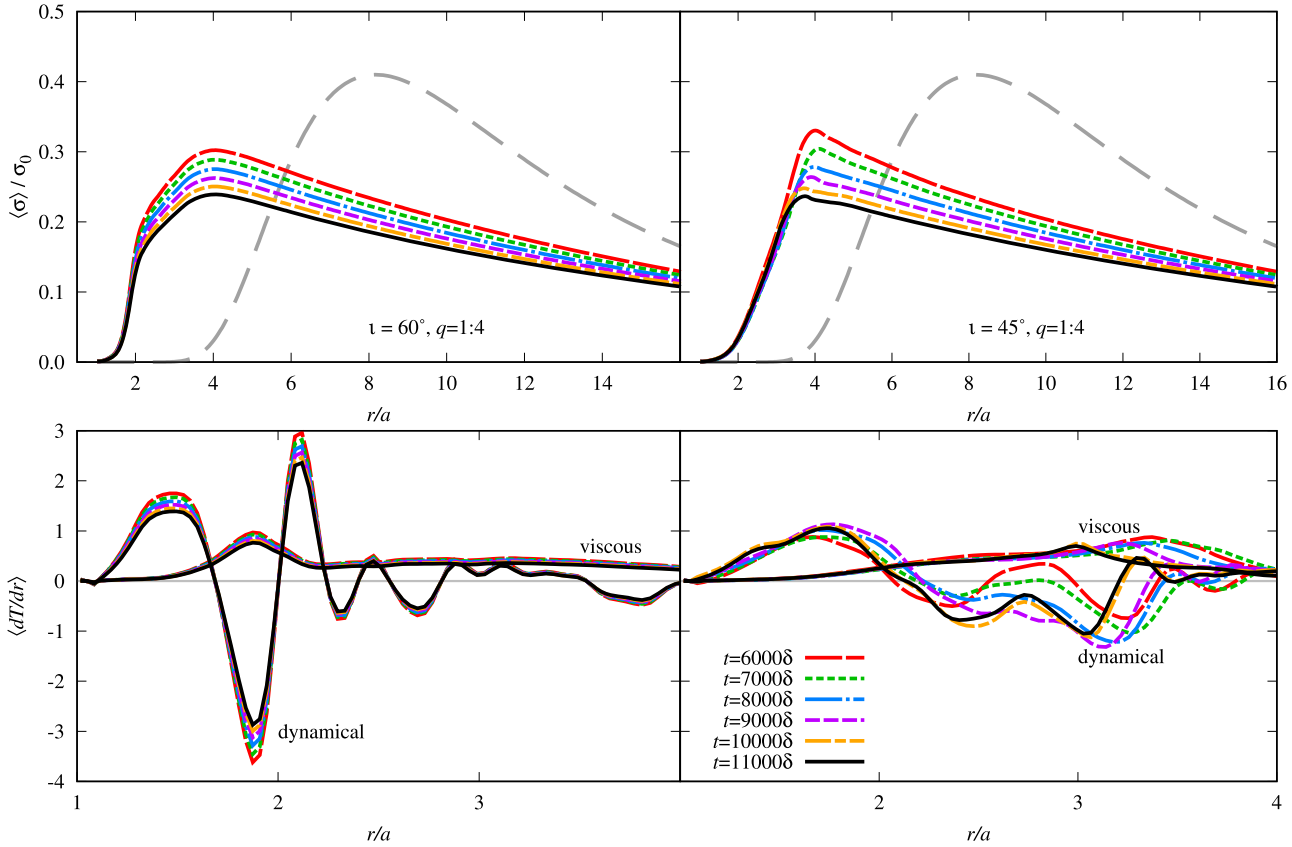

Figure 6. Density (upper row) and torque density (lower row) for different time samples for two cases with $q = 1/4$: stable one with $\iota = 60°$ (left column) and unstable one with $\iota = 45°$ (right columns) as an example of what we understand as "inability to settle down in quasi-steady state." The color coding is consistent with the color palette in the upper left plot. In the case of torque densities, solid and broken lines correspond to dynamical and viscous torques respectively.

is not dictated by a linear coupling to the gravitational potential. Therefore, the assertion that there is a dynamical irrelevance of resonant torquing in the opening of the circumbinary gap can be extended to cases where there is an inclination between the disk and binary planes. In our orbital stability picture of gap opening, we sum over multiple modes of the gravitational potential and do not assert that any particular harmonic mode is responsible for the opening of the gap as is the case in the resonant torquing picture.

Another interesting observation is related to mass ratios. In Figure 9, we compare the density distributions for our full range of mass ratios for a given inclination angle, and provide separate panels for each inclination angle. For *not-inclined* and *moderately inclined* configurations in the top row, we see a continuous evolution from the equal-mass case down through the lower mass ratios. The situation for larger inclination angles is quite different. From $\iota = 50°$ through 105°, we can see the $q = 1$ and $q = 2/3$ cases are very different from the other three cases. Interestingly, these three cases, $q \in \{3/7, 1/4, 1/10\}$, are quite similar to the *highly inclined* systems, which all begin to resemble the single mass case.

Let us define an additional reference location, the *isodensity radius* $r_{10\%}$, as the radius at which the density has a value of 10% of the final density maximum. We now focus on the location of the three radii $r_{\rm dT}$, $r_{\rm max}$, and $r_{10\%}$, as different ways of characterizing the density distributions more generally. The top panels in Figures 10, 11, and 12 present these radii as functions of a mass ratio for each inclination angle, while the bottom panels show the radii as functions of an inclination angle for each mass ratio. We find that the locations of the density maxima for small mass ratio cases $q \in \{3/7, 1/4, 1/10\}$ vary weakly with inclination angle. On the other hand, for comparable mass cases, i.e $q \in \{1/1, 2/3\}$, the results are completely different from the other three and almost identical with each other—the location of the density maximum behaves almost like a step function, with the maximum moving sharply inward at high inclinations. For *moderately inclined* cases, the density maxima occur at relatively large radii. For *highly inclined* configurations, with $\iota \geqslant 55°$, the maxima occur at roughly half the radius of the low inclination cases.

The trends observed in the behavior of the other two reference radii are less striking and more continuous across mass ratios and inclinations. Both $r_{\rm dT}$ and $r_{10\%}$ are large for small inclinations and comparable masses and, with a few exceptions, decrease with decreasing mass ratio. They also decrease with increasing inclination angle, although there are still more exceptions to this monotonicity. All of these exceptions occur in or close to the unstable sector.

The behavior of the dynamical and viscous torques also bears further study. The bottom row of Figure 6 displayed how the values of $r_{\rm dT}$ stabilize over time. In this paper, we do not focus on the nature or the apparent behavior of the torques, aside from noting that some simulated configurations never reach a consistent value for $r_{\rm dT}$. This instability may manifest as completely irregular or secular changes over time, or may oscillate within some restricted range of values.

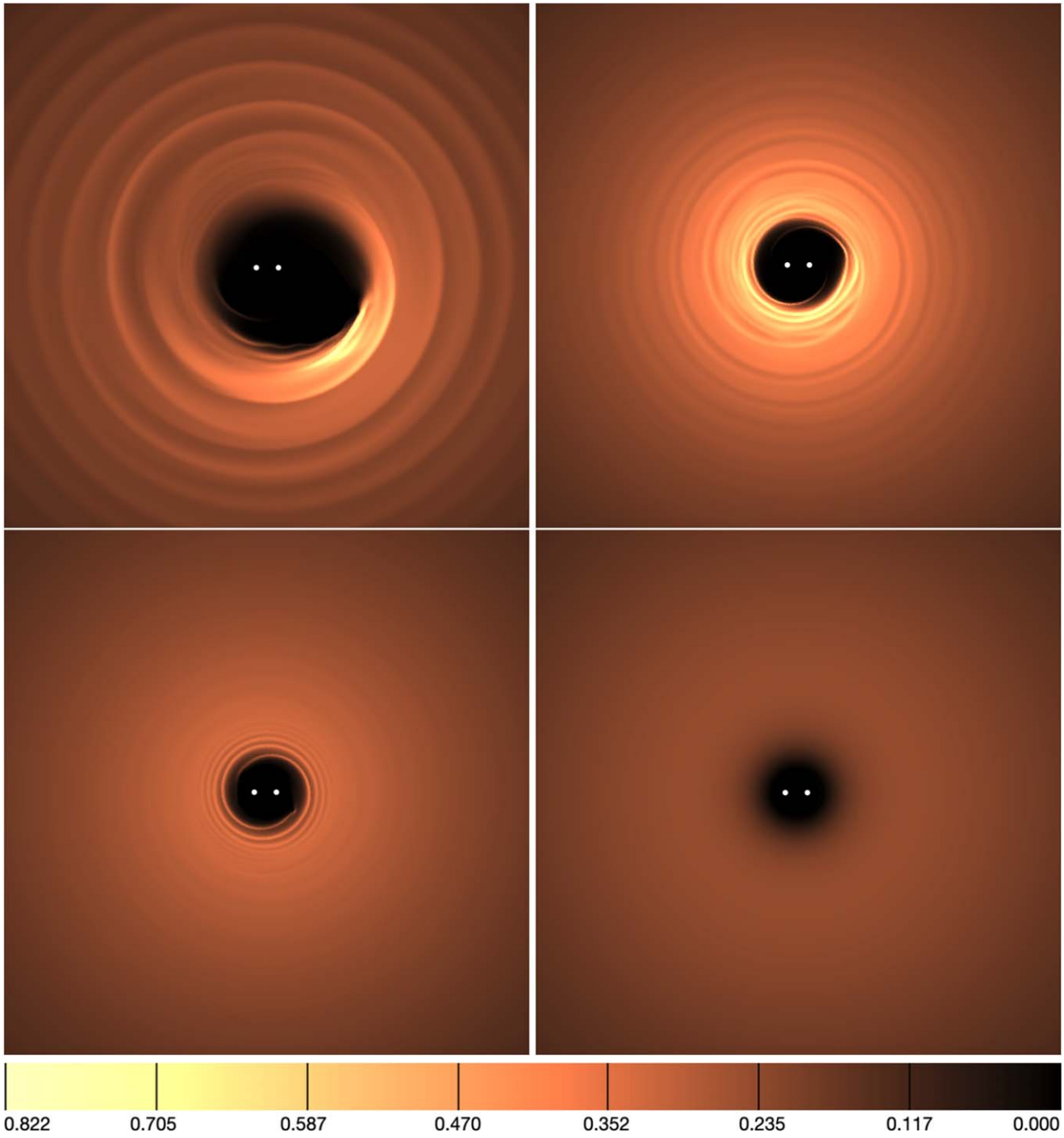

Figure 7. Two-dimensional contours of density at $t = 6000\delta$ for the equal mass case with $\iota = 0°$ (top left), 60° (top right), 120° (bottom left), and 180° (bottom right).

In the second case, the amplitude and period of these oscillations seem to be completely unpredictable and do not depend in any simple way on the model parameters. These effects are present for the entire simulation, long after $t = 6000\delta$, which we used to report our data. We pushed some unstable configurations to 24,000δ, 36,000δ, or even, in the case of $\iota = 45°$, 48,000δ. Only *not-inclined* or *highly inclined* ($\iota \gtrsim 55°$) systems are free from this effect.

Figure 13 illustrates this instability with some examples of r_{dT} calculated at time samples spaced by 1000δ. The data here were intentionally chosen to illustrate these qualitatively different behaviors. For inclined cases with $\iota \in \{15°, 20°, 25°, 30°, 35°, 40°\}$, all investigated systems presented some periodicity. We observed that the amplitude was largest and the period was longest for $\iota = 30°$.

The inclination $\iota = 45°$ behaved qualitatively differently and requires separate discussion. In the case of $\iota = 45°$, we observed periodicity for $q = 1/4$, stability for $q = 2/3$, and neither stability nor periodicity for $q = 1/1$ and $q = 3/7$.

Lastly, we must emphasize that our conclusions regarding the physical mechanism that dominates the gap and our observations regarding instabilities may depend on the range of

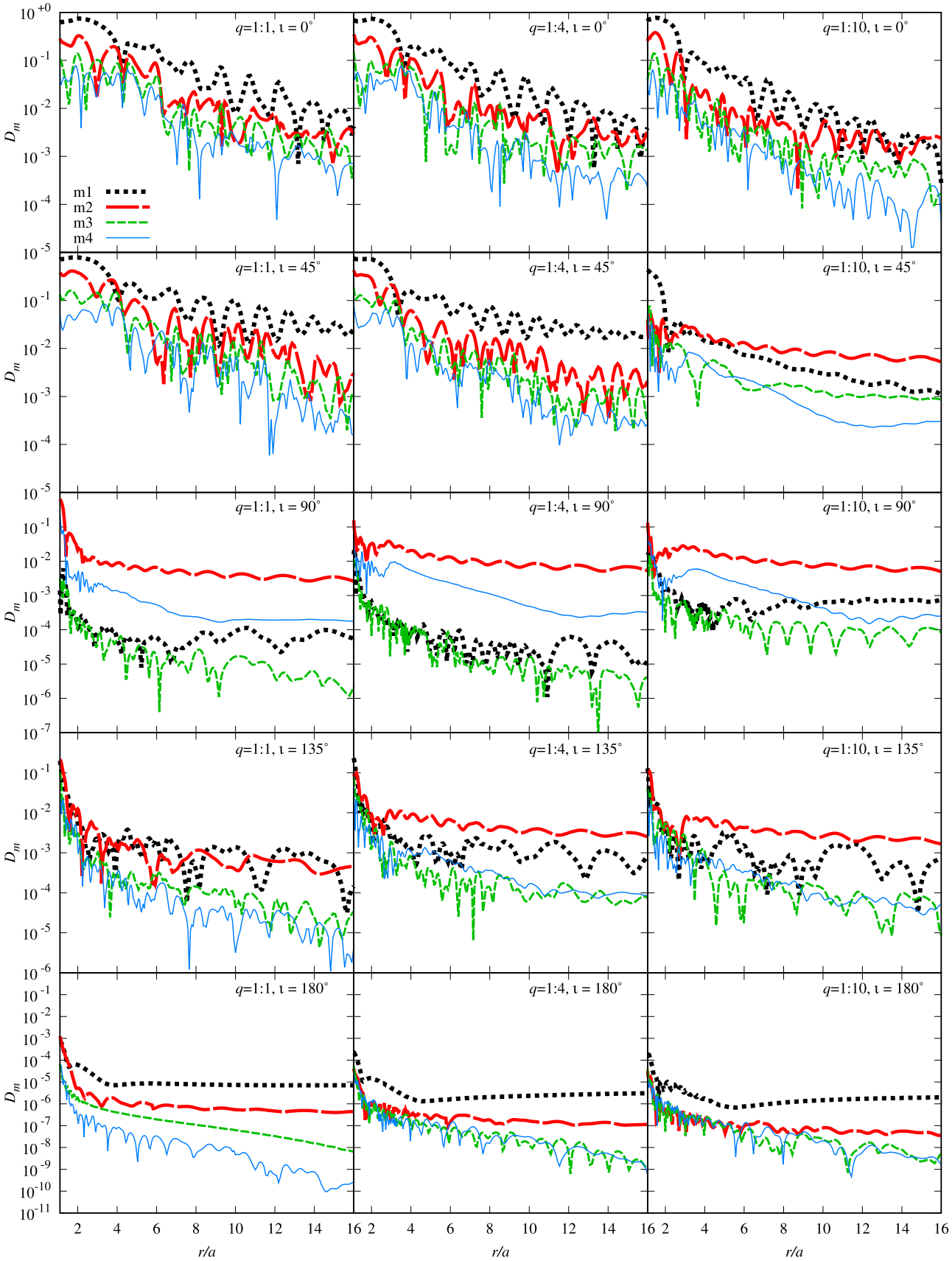

Figure 8. Azimuthal mode decomposition for $m \in \{1, 2, 3, 4\}$ corresponding to the density profiles presented in Figure 9. Columns from the left correspond to mass ratio $q \in \{1/1, 1/4, 1/10\}$, and rows from the top correspond to $\iota \in \{0°, 45°, 90°, 135°, 180°\}$.

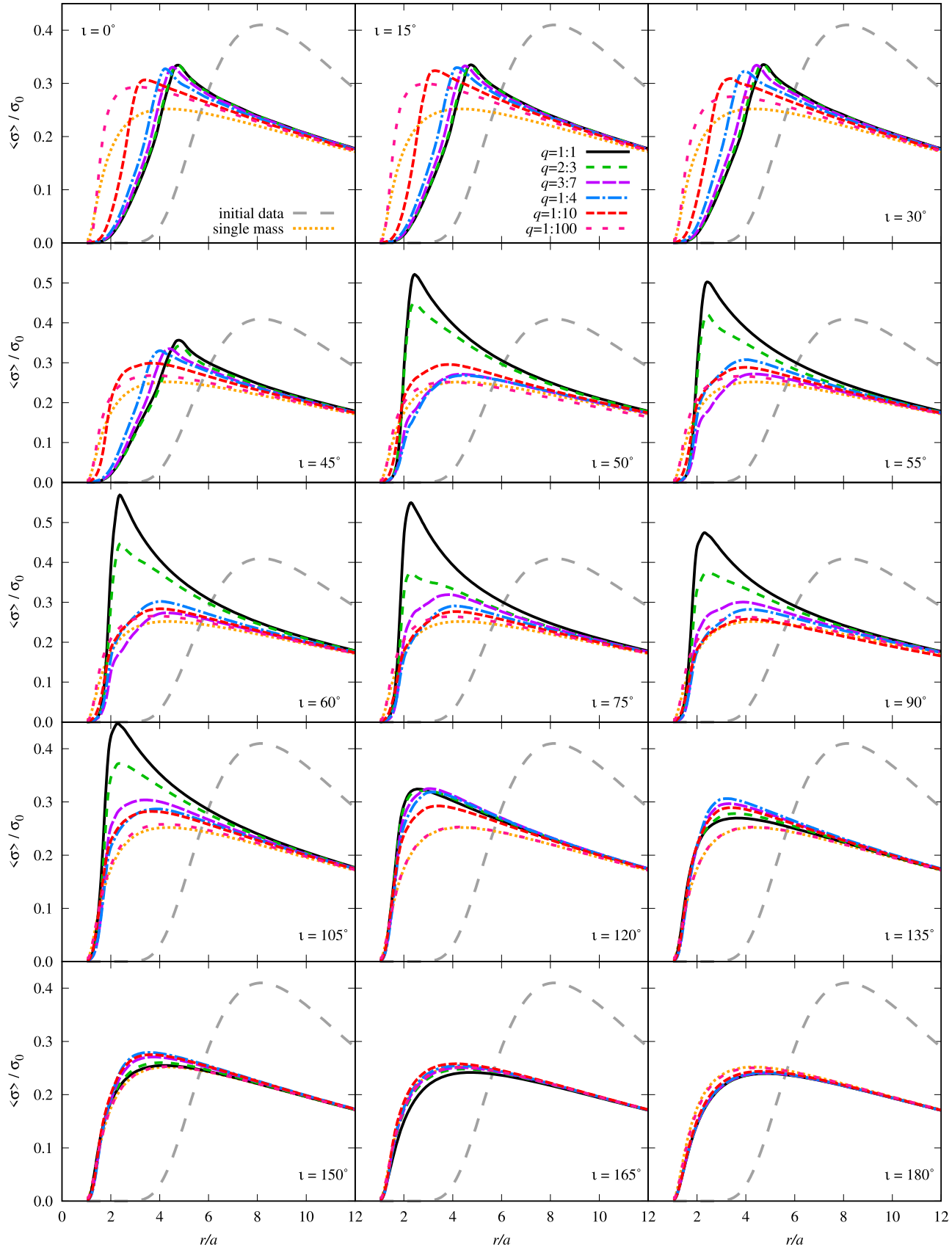

Figure 9. Azimuthally and time-averaged density distributions. Each panel shows all mass ratio configurations for a given inclination angle, with panels for all the inclination angles we considered.

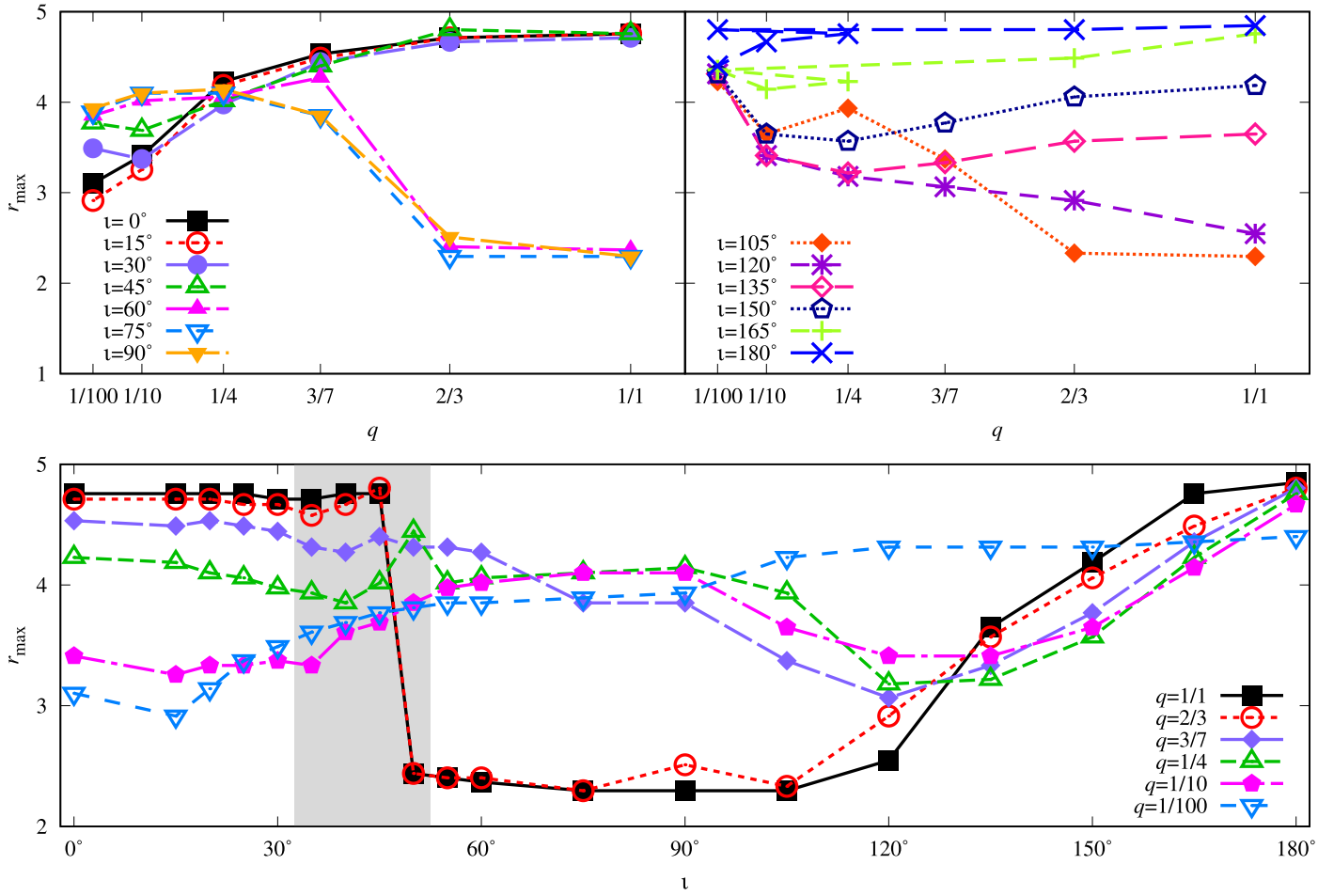

Figure 10. The location of r_{max}. The top panels show different inclination angles as a function of mass ratio, with corotating cases on the left and counterrotating cases on the right (cases $\iota \in \{20°, 25°, 35°, 40°, 50°, 55°\}$ are omitted for clarity). The bottom panel shows different mass ratios $q \in \{1/1, 2/3, 3/7, 1/4, 1/10\}$ as a function of inclination angle. The gray region indicates the unstable sector.

viscosities under consideration. In Figure 14, we show $r_{10\%}$ as a function of inclination for $q \in \{1{:}1, 1{:}4, 1{:}10\}$, for three different viscosities $\alpha \in \{0.003, 0.01, 0.03\}$. While the general trend of decreasing gap size with increasing inclination is universal, the gap shrinks with increased viscosity at lower inclinations, and is independent of viscosity at larger inclinations. In addition, the unstable sector that we have discussed appears to apply for lower viscosity, but is qualitatively different for the larger viscosity case, where only the not-inclined and counterrotating cases manifest instability. We leave the study of larger viscosity cases to future work.

Since we are analyzing three parameters simultaneously, mass ratio q, inclination angle ι, and viscosity ν, we present most of our results in a compact form in Figures 10–14, which display the values of key parameters that describe the density and torque distributions. However, we also present certain specific plots to facilitate direct comparison with the existing literature. For the coplanar configuration, a valuable reference is D. J. D'Orazio et al. (2013). Specifically, for the density distribution, the top left panel of Figure 9 can be compared with Figure 4 in D. J. D'Orazio et al. (2013). Regarding torque, a useful comparison of the viscous and dynamical torque intersections can be found by comparing the data from Figure 11 with the torque intersections explicitly presented in Figure 5 of D. J. D'Orazio et al. (2013).

4. Analytical Calculations

In this section, as with Paper I, we compute the size of the circumbinary gap under two regimes. The first, which is valid at small timescales, relates to the stability of the epicyclic orbits and defines the gap size as the radius where orbital instabilities propagate at timescales equal to the binary orbital period. This has proven fruitful in explaining the observed trends in the gap sizes in the coplanar case from numerical simulations. The second method that assumes the gap is maintained by the balancing of resonant and fluid torques defines the gap size as the radius of the resonance where the corresponding torque densities balance.

4.1. Orbital Stability and Gap Size

To recap from Paper I, we study orbital stability of the epicyclic orbits using Lyapunov exponents. Each Lyapunov exponent corresponds to a particular eigenmode through which a perturbation from an exact solution to the Hamilton's equations of motion propagates. The eigenvector provides the dynamical nature of the perturbation while the eigenvalue, known as the Lyapunov exponent, provides the inverse timescale over which the perturbation evolves exponentially. That is, it is a complex number whose real part determines the inverse timescale over which the perturbation decays or blows

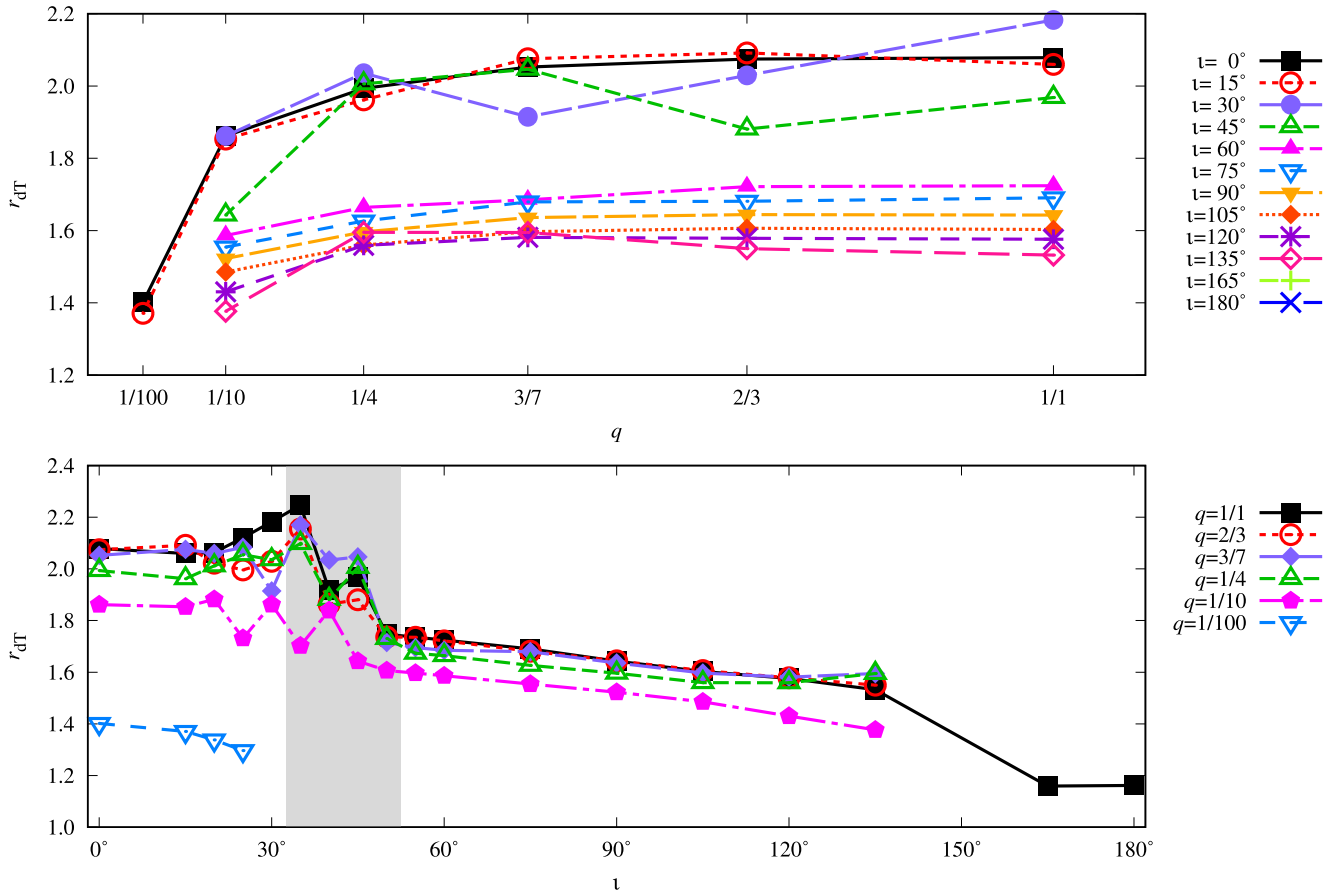

Figure 11. The location of $r_{\rm dT}$. The top panel shows different inclination angles as a function of mass ratio, but unlike Figure 10, all inclinations are included in a single panel. The bottom panel again shows different mass ratios $q \in \{1/1, 2/3, 3/7, 1/4, 1/10\}$ as a function of inclination angle. The gray region indicates the unstable sector.

up while the imaginary part provides the frequency of the oscillatory component of the perturbation.

In this case, we are interested in the leading or largest Lyapunov exponent as that corresponds to the instability that blows up over the shortest timescale. We then define the gap radius prediction due to orbital instabilities as the radius, $r_{\rm L}$, where the following condition is met

$$\Re\mathfrak{e}\left(\lambda^{\rm max}(r_{\rm L})\right) = \frac{1}{\delta} = \frac{1}{2\pi}. \tag{29}$$

Here, $\lambda^{\rm max}$ is the largest Lyapunov exponent, and $\delta = 2\pi$ is the binary orbital period in our units of choice.

We present results for $\tau = 1/\Re\mathfrak{e}(\lambda^{\rm max})$, the instability timescale corresponding to the largest Lyapunov exponent, in Figure 15. We note that the instability timescale at each inclination increases with decreasing mass ratio and increasing inclination. This is a result of the weakening of the perturbing potential in those limits. As we approach the counterrotating case, the instability timescales become larger than δ.

4.2. Resonant Torquing Picture

To again briefly recap from Paper I, we study the dynamics of the disk in the fluid picture where the gap is created by the balance of torques applied by the gravitational force of the binary as well as the torque due to viscous dissipation. An analytical description of the torques is based on the WKB approximation of the hydrodynamic equations, where angular momentum is deposited into the disk by dynamical torques at Lindblad resonances and dissipated away by viscous torques. The circumbinary gap is said to be opened at the outermost Lindblad resonance where the dynamical torque is greater than the viscous torque. In this description, we first determine the outermost Lindblad resonance where the resonant torque and the viscous torque can balance each other, referred to as the gap opening resonance. Upon determining the gap opening resonance, we define the size of the gap, $r_{\rm T}$, as the outermost location where the amplitude of the epicycles, $A(r_{\rm T})$, given by Equation (3), can extend to the location of the gap opening resonance, r_ζ. That is, we solve the equation

$$f(r_{\rm T}) \equiv \log\left(\frac{A(r_{\rm T})}{|r_{\rm T} - r_\zeta|}\right) = 0. \tag{30}$$

We present the results of the gap size according to the resonant torquing picture in Figure 16. We note that, as with Paper I, the equal mass case yields a gap size smaller than other mass ratios since the symmetry in masses destroys the $(m, n) = (1, 1)$ resonance. Additionally, we note that, at each mass ratio, there is a critical inclination at which the gap opening resonance transitions from (1, 1) to (2, 2). This is due to the weakening of the gravitational potential with increasing inclination. Finally, we note that, when approaching the counterrotating case, the gap size shows an increase due to

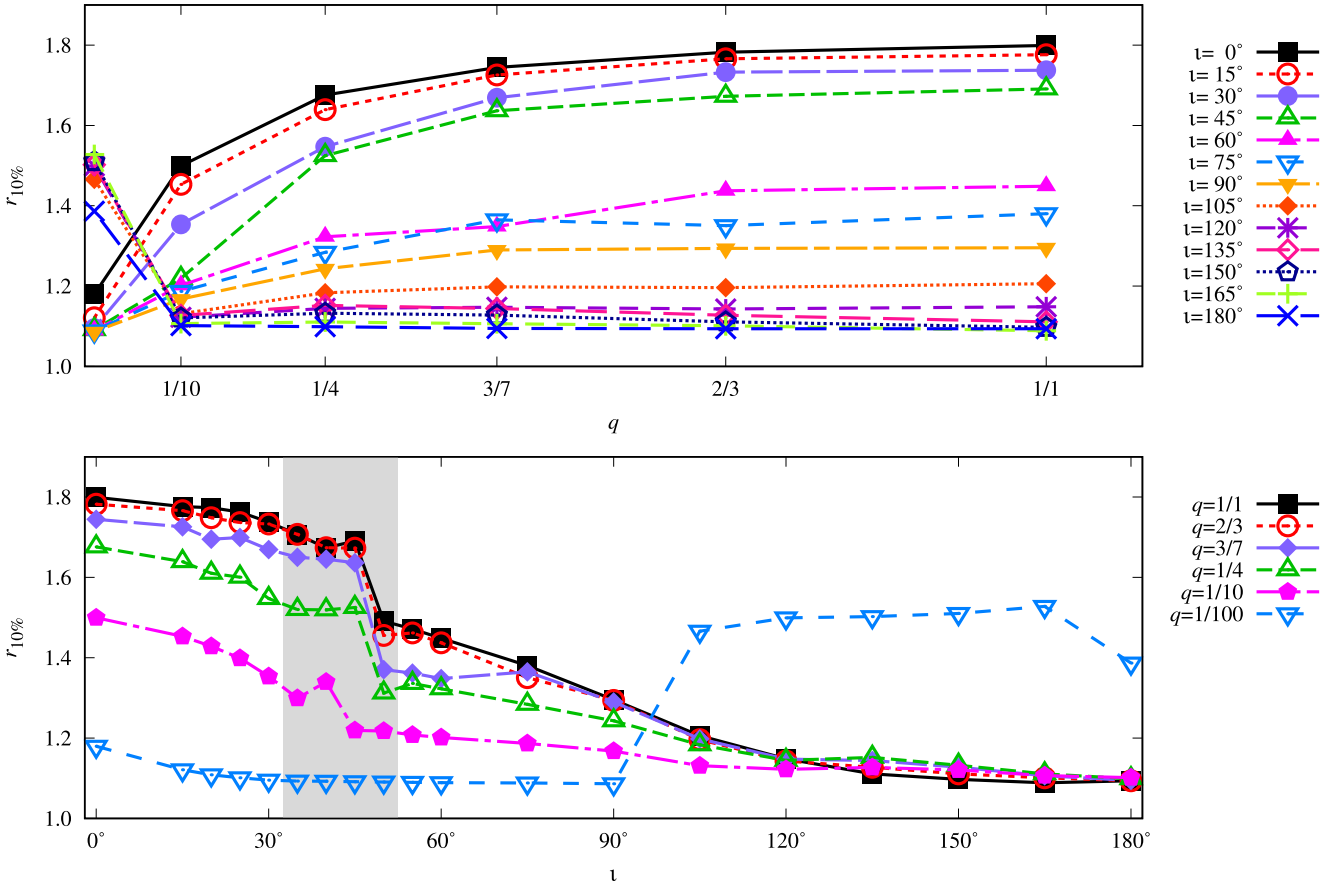

Figure 12. The location of $r_{10\%}$. The presentation of different cases is the same as in Figure 11.

the weakening of the resonant torque at these inclinations leading to noncircular ($m \neq n$) gap opening resonances.

5. Comparison of Analytical and Numerical Results

As with Paper I, we wish to compare the two gap opening mechanisms, but with the added dependence on inclination angle. We emphasize as with Paper I that we wish to compare the trends in the gap size as opposed the exact values at any given mass ratio or inclination, given the flexibility in each definition of gap size. We normalize the numerical and analytical results to clearly identify the trends they make when we overlay them graphically. In this case, we define the quantities $\bar{r}_X$ such that

$$\bar{r}_X = \frac{r_L}{r_X}\Big|_{\mu=0.5,\iota=0} r_X, \tag{31}$$

where the subscript X denotes the subscript of the relevant radius from the numerical or analytical computations, e.g., $r_{10\%}$ as defined in Section 3. The vertical line denotes that the values are taken at the mass ratio $\mu = 0.5$ and inclination $\iota = 0$, so that the gap size as defined by the instability timescale and the numerical/analytical measure of gap size are normalized to agree for equal masses and a coplanar disk–binary system. Finally, we summarize all the definitions of the gap size in Table 3. We also plot the results of the three different definitions in Figure 17. The black scatter plot data indicate the behavior of the numerical gap size $r_{10\%}$ over an inclination,

Table 3
Summary of All Gap Size Definitions

Label	Definition	Defined
$r_{10\%}$	Radius where surface density is 10% of the final maximum	Table 1
r_L	Radius where the Lyapunov timescale is equal to the binary orbital period	Equation (29)
r_T	Radius where the epicyclic amplitude reaches the gap opening resonance	Equation (30)

with each plot corresponding to a different mass ratio, and the unstable sector again highlighted as the gray region. The orange dashed line shows the gap size r_T derived from the resonant torque picture, and the blue solid line shows the gap size r_L derived from the orbital stability picture. We note that, at all mass ratios, the behavior of $r_{10\%}$ shows better agreement with the orbital stability picture than the resonant torque picture. This is particularly true at lower mass ratios where the instability picture predicts gap sizes much closer to the numerical values at inclinations beyond the step function drop. Additionally, at some mass ratios and inclinations, we see an increase in r_T due to the dominance of noncircular resonances, which is not reflected in the numerical data. However, we also observe that r_L does not display the step function-like drop in the gap size while r_T does. That being said, the step function behavior of r_T does not appear to match the step function

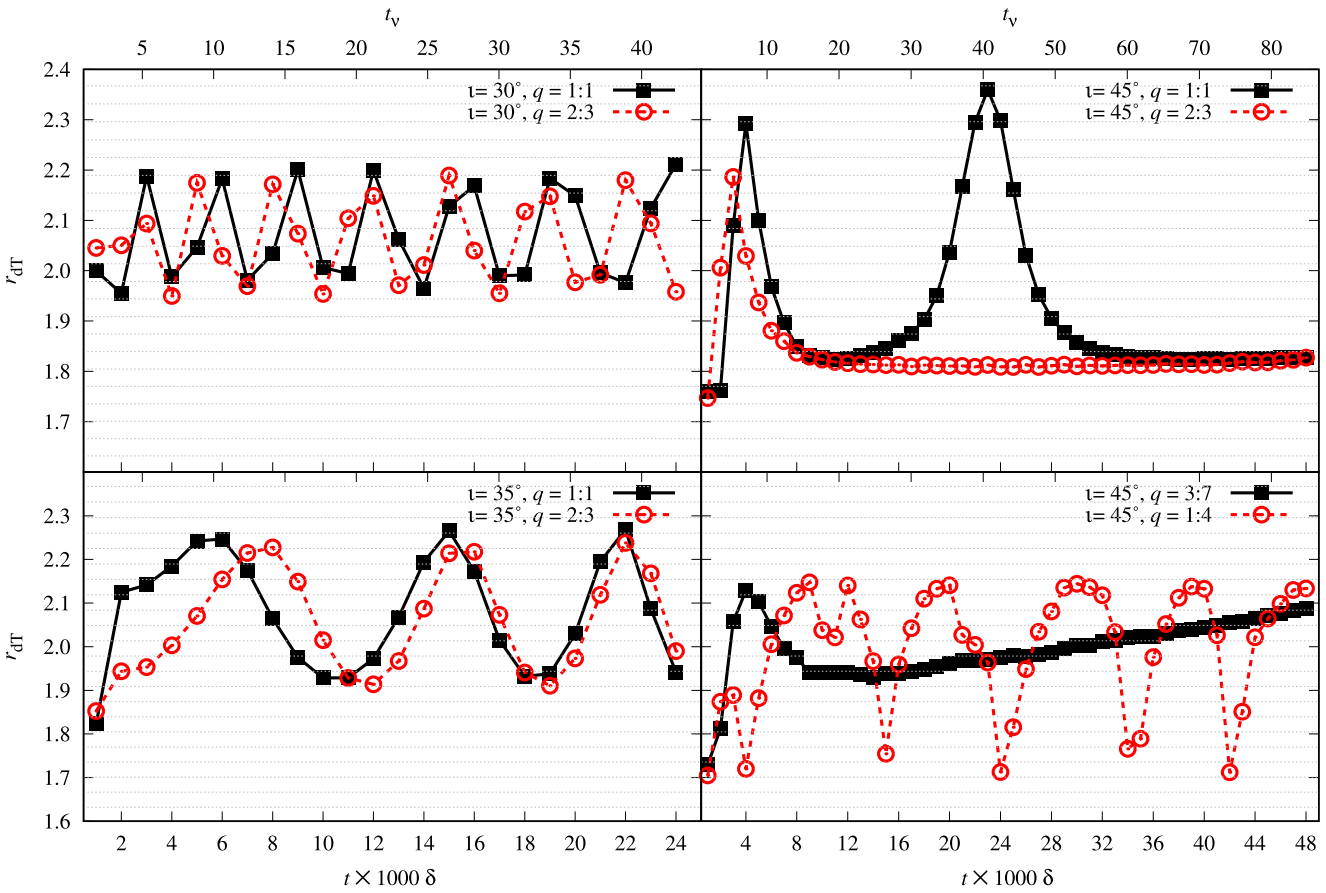

Figure 13. The location where the dynamical and viscous torque densities balance, $r_{\rm dT}$, for a selection of unstable cases simulated over long times. The gray dotted horizontal lines correspond to the numerical grid's radial nodes.

behavior of $r_{10\%}$. For instance, we note that the critical inclination where the numerical results for $r_{10\%}$ drop does not depend on mass ratio while the size of the drop is larger as we approach the equal mass case. However, in the resonant torquing picture, the drop in gap size $r_{\rm T}$ happens at lower inclinations as we approach the equal mass case, which has no step function-like behavior at all, because the drop in $r_{\rm T}$ is due to the transition of the gap opening resonance from the (1, 1) Lindblad resonance to the (2, 2) Lindblad resonance. On the other hand, the drop in the numerical results is due to the fact that, in two different ranges of inclinations, the quasi-steady state is arrived at differently. It is unclear exactly why this is the case in our two-dimensional simulations, and these inclinations are high enough that the role of the gravitational force from the binaries directed out of the disk plane may need to be included to gain a complete picture; three-dimensional simulations could therefore provide a better understanding of how the quasi-steady state is approached in this regime. Additionally, while the orbital stability picture seems to provide better agreement than the resonant torque picture at higher mass ratios (closer to equal mass), the effect of vertical instabilities will need to be incorporated at higher mass ratios where the strength of the gravitational force is higher for larger inclinations. Overall, the results of this paper indicate that the formation and maintenance of the gap at binary timescales through the propagation of orbital instabilities appear to be the dominant mechanism over the parameter space of mass ratios and inclinations at the viscosities we consider.

6. Conclusions

We have studied black hole binary–disk systems within Newtonian gravity and Newtonian hydrodynamics. We have considered configurations with different mass ratios of the binary and two-dimensional, non-self-gravitating, viscous, locally isothermal disks inclined with respect to the binary's orbital plane. Black hole binaries are modeled by point masses, which move on fixed circular orbits.

We investigated the influence of mass ratio and inclination angle on the disk's density distribution. We focused on the location of three radii, which give a general description of the matter distribution in the disk, namely, $r_{\rm dT}$, the radius where the dynamical torque density equals viscous torque density; $r_{\rm max}$, the radius where the density reaches its maximum; and $r_{10\%}$, the radius where the density reaches 10% of the final density maximum. Analyzing these radii together with the general density distributions, we present a broad perspective on the matter density distribution as a function of our model parameters.

Finally, we investigated the effect of different theoretical explanations for the opening and maintenance of the circumbinary gap. In Paper I, we noted that the propagation of orbital instabilities over the timescale of the binary period provides a markedly better estimate of the gap size. In this paper, we have

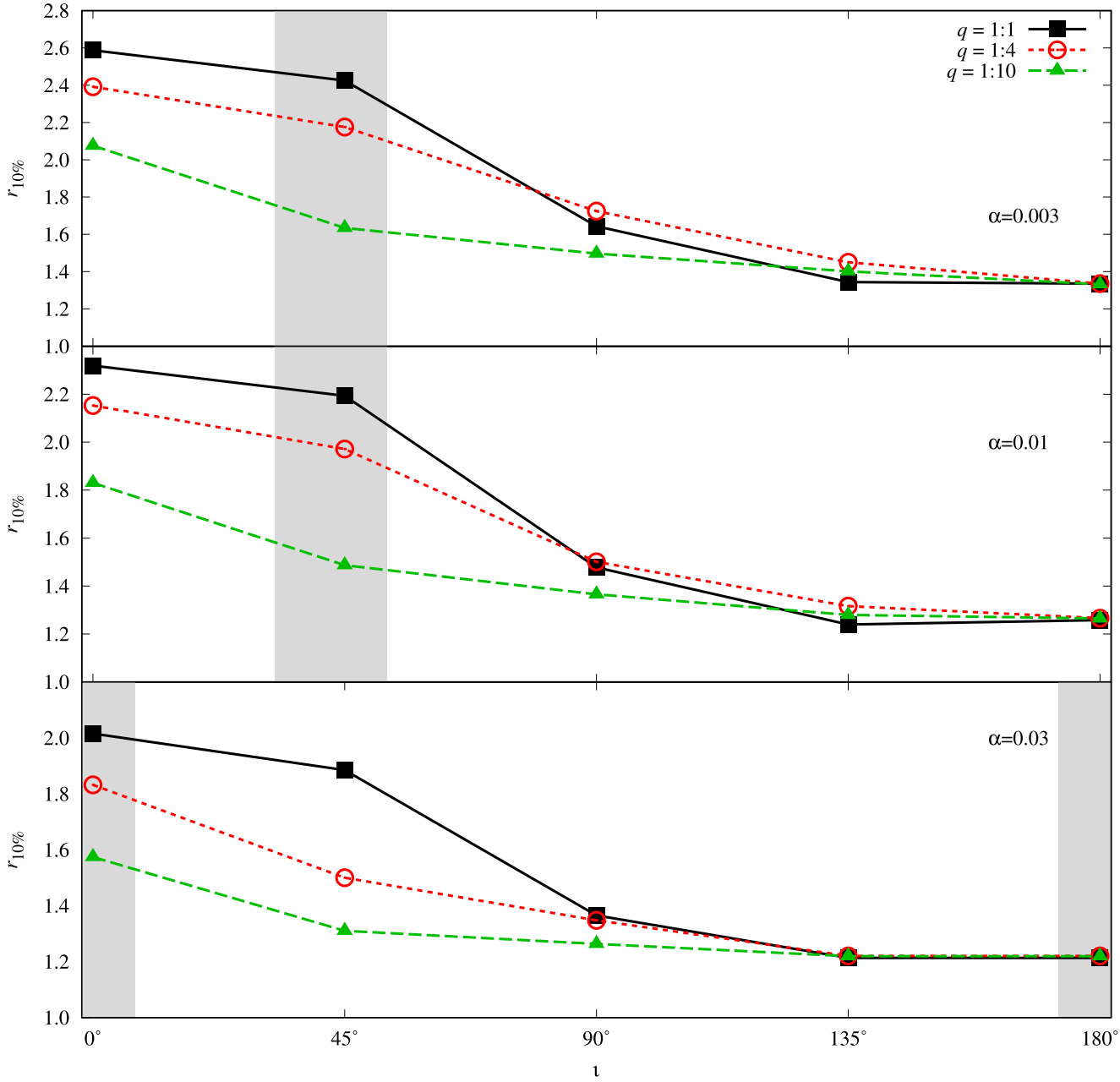

Figure 14. The location of $r_{10\%}$ for different inclinations. Each subplot represents a different choice of the viscosity parameter α.

found that we can reasonably extend this explanation to the case of inclined binary–disk systems. The theoretical explanations and numerical studies approximate the circumbinary disk as a two-dimensional plane; however, a three-dimensional treatment of the disk may be able to clarify the nature of the "unstable sector" of inclination angles, where the disk never reaches a quasi-steady state. We leave this for future study.

Our explanation for the opening of the circumbinary gap over orbital timescales also motivates future studies of this nature with two other model parameters: the binary orbital eccentricity, which has been shown to grow due to the back-reaction of disk dynamics onto the binary (see M. Siwek et al. 2023) or trigger Kozai–Lidov oscillations in highly inclined disks (see W. Fu et al. 2017); and the binary orbital precession, which is seen in binary black hole systems due to spin–orbit interactions (E. Poisson & C. M. Will 2014).

The effect of the Kozai–Lidov mechanism in circumbinary disks has not been clearly studied for hydrodynamic systems. A test particle of mass m in the circumbinary disk forces Kozai–Lidov oscillations on the binary with the total mass M with a period $P_{\rm Kozai} \propto \frac{M}{m}$ (Section 3.4.2 of E. Poisson & C. M. Will 2014). For a typical disk-to-binary mass ratio of $\mathcal{O}(10^{-5})$, this period would be $>10^5\delta$ even at the gap radius, which is much longer than the viscous timescale. J. J. Zanazzi & D. Lai (2017) and W. Fu et al. (2017) focus on the Kozai–Lidov mechanism for circumstellar discs, where the perturbing mass is the secondary mass of the binary and therefore can be relevant even for a massless circumstellar disk.

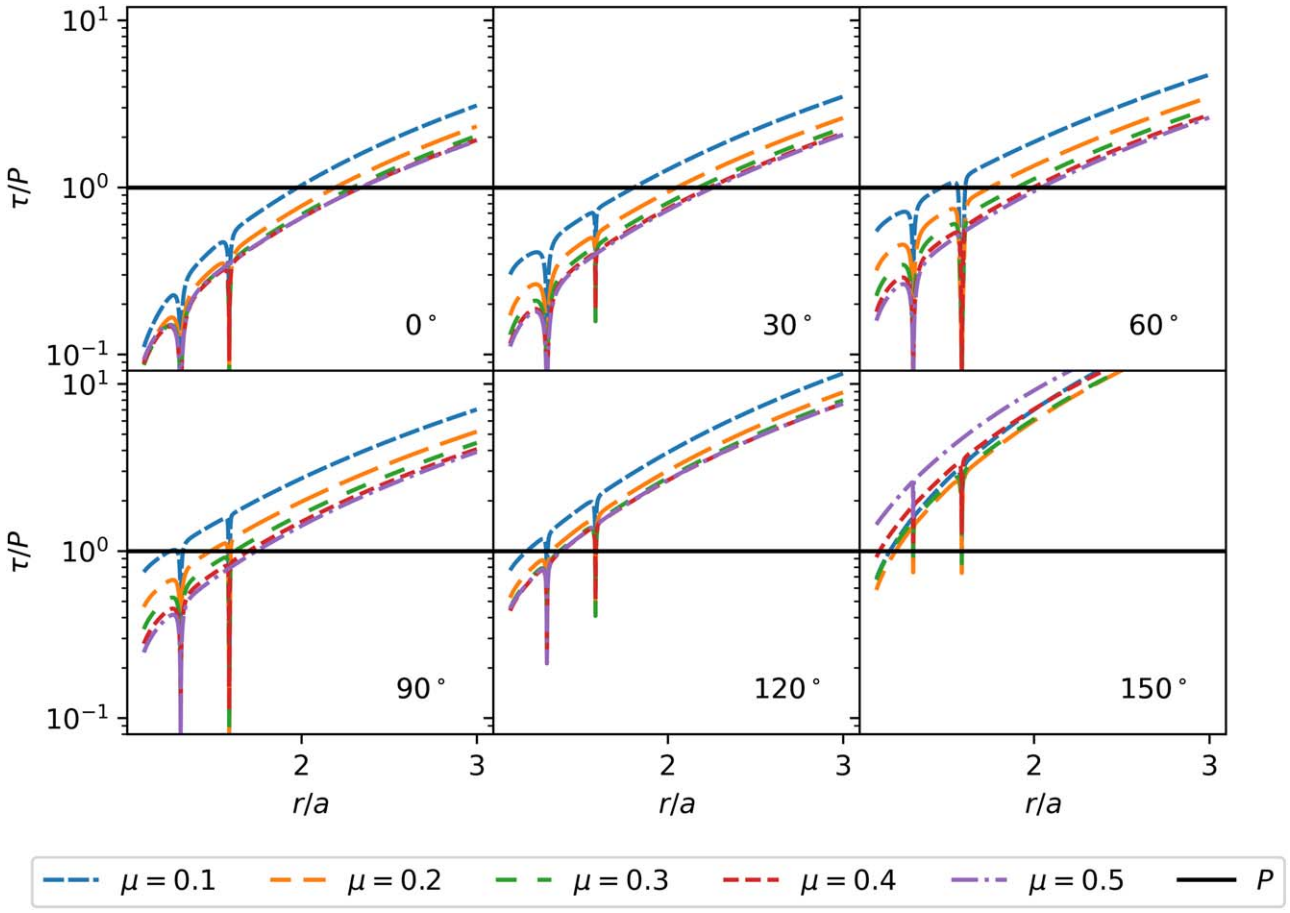

Figure 15. The timescale associated with the largest Lyapunov exponent for all mass ratios at inclinations $\iota \in \{15°, 30°, 45°, 60°, 75°, 90°\}$. The timescales have been normalized to the binary orbital period as opposed to our choice of units where the binary frequency is set to unity. The black line corresponds to the binary orbital period.

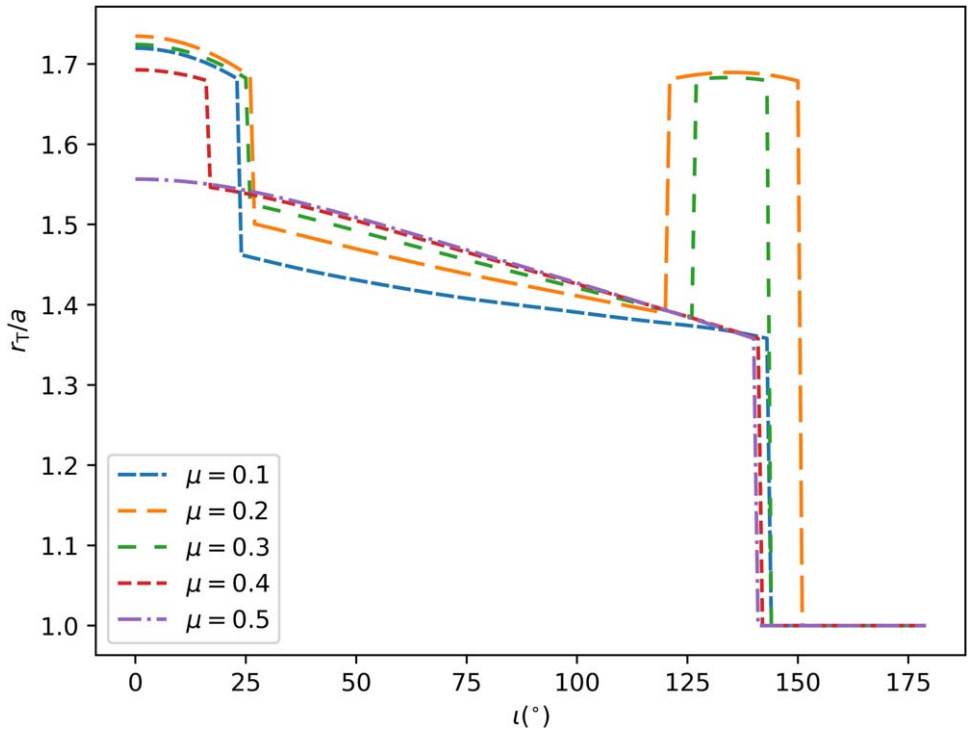

Figure 16. The value of r_T given by the balancing of the resonant and viscous torques in the WKB approximation.

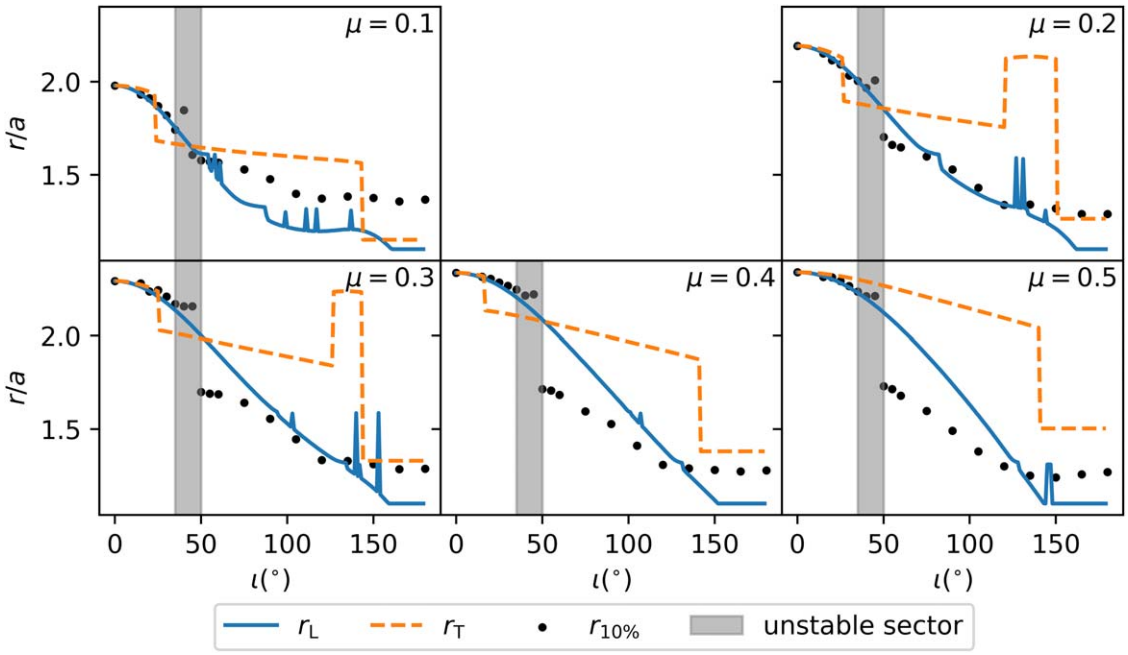

Figure 17. The behavior of the gap sizes listed in Table 3. The black scatter points denote the numerical results, $r_{10\%}$, with the gray region denoting the unstable sector specified in Section 3. The solid blue line denotes the gap size r_L from the orbital instability picture, and the dashed orange line refers to the gap size r_T in the resonant torque picture.

Future studies of this type, along with the physics discussed in this paper, will have broad implications for these astrophysical systems, which should be regularly observed by gravitational-wave and electromagnetic observatories in the coming decades.

Acknowledgments

S.M. and S.T.M. were supported in part by a National Science Foundation CAREER grant No. PHY-1945130. The authors acknowledge the computational resources provided by the WVU Research Computing Spruce Knob HPC cluster, which is funded in part by NSF EPS-1003907, and the Thorny Flat HPC cluster, which is funded in part by NSF OAC-1726534. This research was supported in part by the National Science Foundation under grant No. NSF PHY-1748958 and NASA grants 22-LPS22-0022 and 24-2024EPSCoR-0010.

ORCID iDs

Michal Pirog https://orcid.org/0000-0002-3164-5923
Siddharth Mahesh https://orcid.org/0000-0002-8340-614X
Sean T. McWilliams https://orcid.org/0000-0003-2397-8290

References

Artymowicz, P., & Lubow, S. H. 1996, ApJ, 467, L77
Bartos, I., Kocsis, B., Haiman, Z., & Mrka, S. 2017, ApJ, 835, 165
Bate, M. R., Bonnell, I. A., & Price, N. M. 1995, MNRAS, 277, 362
Bowen, D. B., Mewes, V., Noble, S. C., et al. 2019, ApJ, 879, 76
Chiang, E. I., & Murray-Clay, R. A. 2004, ApJ, 607, 913
Combi, L., Armengol, F. G. L., Campanelli, M., et al. 2021, PhRvD, 104, 044041
Combi, L., Lopez Armengol, F. G., Campanelli, M., et al. 2022, ApJ, 928, 187
Cuadra, J., Armitage, P. J., Alexander, R. D., & Begelman, M. C. 2009, MNRAS, 393, 1423
de Val-Borro, M., Gahm, G. F., Stempels, H. C., & Peplinski, A. 2011, MNRAS, 413, 2679
Dittmann, A., Dempsey, A., & Li, H. 2024, ApJ, 964, 61
D'Orazio, D. J., Haiman, Z., & MacFadyen, A. 2013, MNRAS, 436, 2997
Duffell, P. C. 2016, ApJS, 226, 2
Duffell, P. C., D'Orazio, D., Derdzinski, A., et al. 2020, ApJ, 901, 25
Duffell, P. C., & MacFadyen, A. I. 2011, ApJS, 197, 15
Duffell, P. C., Dittmann, A. J., DOrazio, D. J., et al. 2024, ApJ, 970, 156
D'Orazio, D. J., Haiman, Z., Duffell, P., MacFadyen, A., & Farris, B. 2016, MNRAS, 459, 2379
Escala, A., Larson, R. B., Coppi, P. S., & Mardones, D. 2005, ApJ, 630, 152
Facchini, S., Lodato, G., & Price, D. J. 2010, A&A, 510, A23
Farris, B. D., Duffell, P., MacFadyen, A. I., & Haiman, Z. 2014, ApJ, 783, 134
Farris, B. D., Gold, R., Paschalidis, V., Etienne, Z. B., & Shapiro, S. L. 2012, PhRvL, 109, 221102
Foucart, F., & Lai, D. 2014, MNRAS, 445, 1731
Frank, J., King, A., & Raine, D. 2002, Gas Dynamics (3rd ed.; Cambridge: Cambridge Univ. Press), 822
Fu, W., Lubow, S. H., & Martin, R. G. 2017, ApJL, 835, L29
Goldreich, P., & Tremaine, S. 1980, ApJ, 241, 425
Günther, R., & Kley, W. 2002, A&A, 387, 550
Hanawa, T., Ochi, Y., & Ando, K. 2010, ApJ, 708, 485
Hayasaki, K., Mineshige, S., & Sudou, H. 2007, PASJ, 59, 427
Hioki, T., Itoh, Y., Oasa, Y., Fukagawa, M., & Hayashi, M. 2011, PASJ, 63, 543
Lines, S., Leinhardt, Z. M., Baruteau, C., Paardekooper, S. J., & Carter, P. J. 2015, A&A, 582, A5
Lubow, S. H., & Ogilvie, G. I. 1998, ApJ, 504, 983
MacFadyen, A. I., & Milosavljević, M. 2008, ApJ, 672, 83
Mahesh, S., McWilliams, S. T., & Pirog, M. 2024, ApJ, 973, 18
McKernan, B., Ford, K. E. S., Bellovary, J., et al. 2018, ApJ, 866, 66
Milosavljević, M., & Phinney, E. S. 2005, ApJL, 622, L93
Miranda, R., & Lai, D. 2015, MNRAS, 452, 2396
Miranda, R., Muoz, D. J., & Lai, D. 2017, MNRAS, 466, 1170
Noble, S. C., Mundim, B. C., Nakano, H., et al. 2012, ApJ, 755, 51
Pelupessy, F. I., & Portegies Zwart, S. 2013, MNRAS, 429, 895
Poisson, E., & Will, C. M. 2014, Gravity (Cambridge: Cambridge Univ. Press)
Ragusa, E., Lodato, G., & Price, D. J. 2016, MNRAS, 460, 1243
Roedig, C., Sesana, A., Dotti, M., et al. 2012, A&A, 545, A127
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 500, 33
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
Siwek, M., Weinberger, R., & Hernquist, L. 2023, MNRAS, 522, 2707
Stone, N. C., Metzger, B. D., & Haiman, Z. 2017, MNRAS, 464, 946
Tagawa, H., Haiman, Z., & Kocsis, B. 2020, ApJ, 898, 25
Tang, Y., MacFadyen, A., & Haiman, Z. 2017, MNRAS, 469, 4258
Toro, E. F. 2013, Riemann Solvers and Numerical Methods for Fluid Dynamics (Berlin: Springer)
Zanazzi, J. J., & Lai, D. 2017, MNRAS, 467, 1957
Zettili, N. 2009, Quantum Mechanics: Concepts and Applications (New York: Wiley)

The analytical and numerical explorations presented in this chapter provide a comprehensive picture of how black hole binaries dynamically shape their surrounding environments. The resulting insights lay the groundwork for predicting electromagnetic counterparts to gravitational wave events. The next chapter will conclude the thesis by summarizing these findings and outlining future directions for research in multimessenger astrophysics.

Chapter 4

Conclusion: Dynamical Instabilities in Black Hole Binary Modeling

4.1 Summary of the central thesis

The above chapters have detailed key contributions to the fields of gravitational waveform modeling and the study of black hole binary environments. This dissertation has developed an instability-based approach to two central problems in black hole binary astrophysics: the modeling of gravitational waveforms from binary mergers and the modeling of circumbinary disk dynamics. Although these problems involve different physical regimes, both can be understood by identifying the dynamical instability that controls the observable of interest. In the waveform problem, the merger-ringdown signal is governed by the transition from binary dynamics to remnant black-hole perturbations. In the disk problem, the circumbinary gap is governed by the stability of perturbed orbits in the binary potential. The main result of this dissertation is that these instability-based descriptions provide physically interpretable alternatives to purely phenomenological modeling.

4.2 Vacuum binary dynamics and waveform modeling

The waveform-modeling component of this dissertation introduced SEBOB, a framework that combines EOB inspiral dynamics with a BOB description of the merger-ringdown. This construction

shows that a significant part of the late-time waveform can be modeled using the remnant-centered dynamics of the final black hole, reducing the need for phenomenological calibration in the merger-ringdown sector. The model remains competitive with highly calibrated EOB models for aligned-spin binaries while making the physical assumptions of the merger-ringdown attachment more transparent.

4.3 Circumbinary disk dynamics

Having summarized the advances in modeling the vacuum dynamics of the binary, we now turn to the conclusions drawn regarding their astrophysical environments. The study of disk-binary interactions in Chapter 3 provides a new theoretical framework for understanding the dynamics of disk-binary systems, particularly the formation and maintenance of the circumbinary gap. The framework was able to provide predictions for the gap size that were in good agreement with numerical simulations, and generalized well to the case of misaligned disk-binary systems. In order to develop the framework, the gravitational potential of the binary on the disk was chosen to be Newtonian and the analysis of the orbits was limited to linearized perturbation about those due to single point mass.

The disk-dynamics component showed that the standard resonant-torque picture is not sufficient to explain the circumbinary gap seen in numerical simulations. In the coplanar case, the gap location is better predicted by the onset of epicyclic instability in the binary potential. In inclined systems, the same framework predicts an unstable sector at intermediate inclinations and explains persistent time-dependent oscillations in the disk. These results shift the interpretation of gap maintenance from a long-timescale viscous balance to a short-timescale dynamical response.

The fundamental quantity that the framework predicted was the radius of the circumbinary gap. Of course, this quantity assumes that the gap is approximately circular. On the other hand, the morphology of the circumbinary gap can be highly eccentric [6] and while the framework does not provide an explicitly eccentric treatment of the orbits, the relative amplitude of the epicycles sourced by the time-dependent part of the binary potential provides a semi-accurate estimate of this eccentricity.

4.4 Future directions

Waveform Modeling

Future work in waveform modeling should focus on improving the amplitude behavior of BOB-informed NQC corrections, extending SEBOB to higher modes and precessing systems, developing a more self-consistent inspiral-to-remnant map, and further reducing the model's reliance on numerical-relativity fits.

The SEBOB waveform model as described in Chapter 2 provides accurate $(2, 2)$ spin-weighted spherical harmonic modes for aligned-spin quasi-circular black hole binaries. To make SEBOB a viable candidate for future GW detectors like LISA, we must extend the model to include higher order modes, spin precession and other parts of the predicted parameter space. Additionally, introducing BOB to the EOB paradigm requires careful examination of the choice of calibration parameters and biases observed during parameter estimation.

In the latter direction, progress has been made towards developing a new calibration strategy for SEBOB that reproduces the NR merger-ringdown signal with higher fidelity. In the original SEOBNRv5 model, the calibration is performed by performing parameter estimation on the NR waveform to determine the optimal values of the calibration parameters, given the model parameters. The calibration parameters are chosen as the median of the estimated posterior distribution. This flavor of calibration can be computationally expensive to repeat with new NR waveforms and is potentially susceptible to introducing biases from artifacts in the simulated noise. The choice of waveforms used, in terms of coverage of parameter space and numerical configuration, can also introduce biases when performing parameter estimation on untrained NR datasets. While the SEOBNRv5 model showcases state-of-the-art accuracy, there is no open source implementation of the calibration procedure, making it difficult for the broader community to adopt and validate the model, especially given the replacement of the merger-ringdown signal in our work.

The proposed strategy does away with the probabilistic nature of the original procedure and instead reformulates it as a constrained optimization problem. The calibration parameters are chosen as the values that minimize the mismatch between the NR and EOB waveforms, with the minimization performed by the Nelder-Mead algorithm [20]. Due to the low dimensionality of the calibration parameter space, the Nelder-Mead algorithm is able to efficiently find the optimal values of the calibration parameters while avoiding the computational expense of gradient-based optimization methods. Currently, this has been implemented for the non-spinning calibration parameters. The results of this calibration are shown in Figures 4.1 and 4.2.

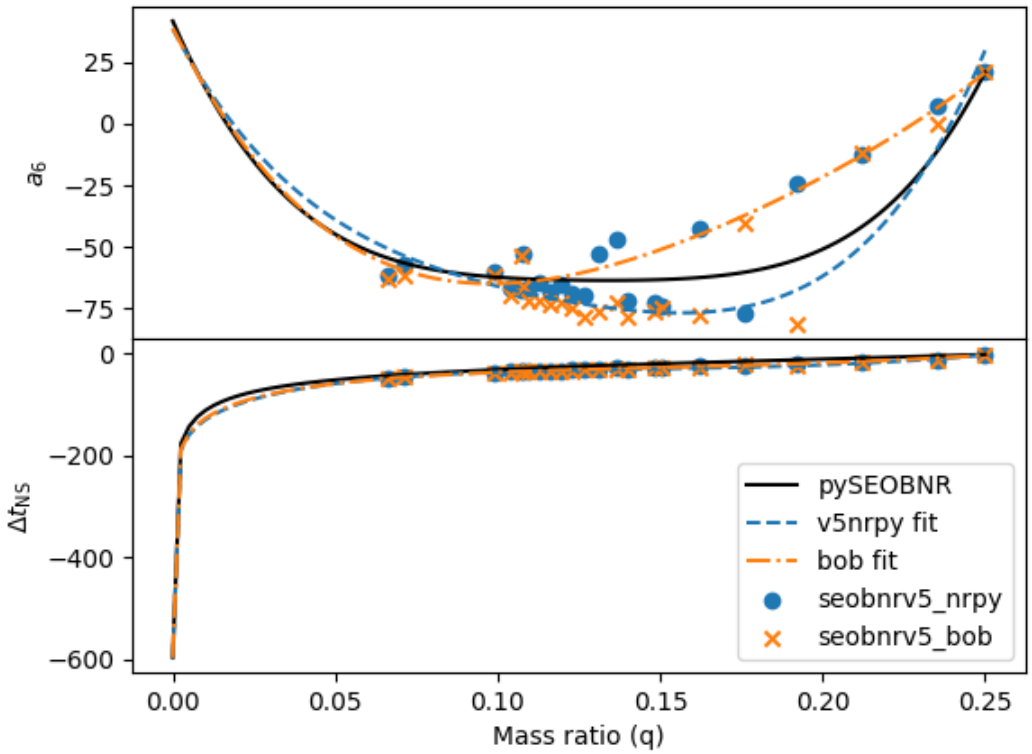

Figure 4.1: Calibration coefficients for SEBOB for the non-spinning case. The blue and orange scatters show the Nelder-Mead calibration results for SEOBNRv5 and SEBOB respectively. The black fit shows the published fit for SEOBNRv5, while the blue and orange dashed lines show fits to the respective scatters.

I have also been working on introducing the new SEBOB approximant into LISA waveform systematics projects for higher mode and intermediate mass ratio studies. The aim of these projects will be to bring together the theoretical and data-driven approaches to improve the accuracy of waveform models and make them ready for LISA data analysis.

The final aim of the upcoming SEBOBv2 approximant will be to complete the parameter space of SEBOB to include spin precession, higher modes, and improved calibration. The approximant will also take into account recent improvements to the BOB formalism in order to further reduce the amount of NR freedom needed in the merger regime[21].

Disk Dynamics

Future work in disk dynamics should test the instability picture in simulations that include additional physics beyond the current locally isothermal, non-self-gravitating treatment. Important

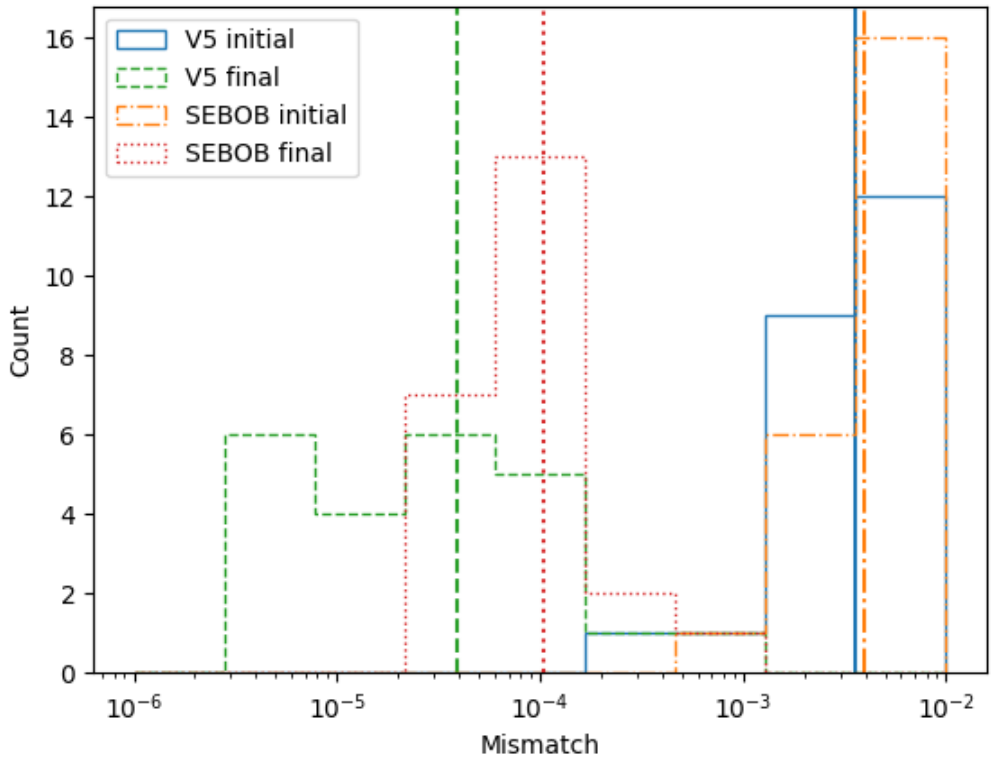

Figure 4.2: Pre and post calibration mismatch histograms for SEBOB and SEOBNRv5 for non-spinning black hole binaries. The blue(solid) and orange(dot-dashed) histograms show mismatches for SEBOB and SEOBNRv5 for the published SEOBNRv5 calibration constants. The green(dashed) and red(dotted) lines show the mismatches for SEBOB and SEOBNRv5 for the Nelder-Mead calibration constants. The results show that the Nelder-Mead calibration procedure is able to improve significantly upon the published SEOBNRv5 calibration constants.

extensions include evolving binaries with gravitational-wave-driven inspiral, including disk backreaction on the binary, exploring magnetohydrodynamic turbulence, adding radiative thermodynamics, and translating the predicted disk variability into electromagnetic observables.

A more rigorous assessment of the morphology of the circumbinary gap can be performed through the direct solution of bound orbits in the restricted 3-body problem. While the numerical integration of the orbits is a well-researched topic [22], they must be computed in a manner capable of performing the orbital stability analysis that was done as part of this thesis. More recently, Langford et al. [23] reduced the problem of finding orbits in the restricted 3-body problem to a root finding problem, including methods to analyze the stability of the resulting numerical orbits. Future work in this area focuses on reproducing that study, but taking advantage of improvements in auto-differentiable code so as to refine the orbital stability calculation by differentiating with respect to the numerical

trajectory solving techniques. The resulting approach can help with generalizing the test particle picture to cases such as highly eccentric or retrograde orbits, where dynamical chaos in the 3-body problem no longer allows for a purely linearized approximation.

Another advantage of adopting numerical techniques is to provide a concrete assessment of if, when, and how disk-binary decoupling manifests in the real system. In that case, the gravitational interaction of the disk and binary can be Newtonian, but the dynamics of the binary system itself can be simulated using PN-derived inspiral dynamics. Then, the evolution of the metastable orbits can be analyzed over the timescale of the inspiral and merger of the binary, providing an estimate of when the characteristic stability timescale of bound orbits in the restricted 3-body problem can no longer keep track of the relativistic merger of the system.

Another important future direction lies in assessing how the interaction between the disk and the binary influences the evolution of the binary itself. Numerically, the change in the semi-major axis and eccentricity of the binary system have been studied [24, 25]. These studies indicated that the circumbinary disk is strongly responsible for shaping the orbital evolution of binary systems. Analytically, Goldreich and Tremaine [16] provided an estimate of how circular binaries evolve by asserting that the angular momentum deposited on the disk at the resonances is the orbital angular momentum lost by the binary. It becomes important to assess how the orbital stability framework developed in this thesis contributes to this subject. More broadly, a formal understanding of why the test particle limit physics provides a good explanation of the dynamics of the circumbinary disk at the gap scale must be derived as a limit of the fluid dynamics itself.

4.5 A unified closing perspective

This thesis lays the groundwork for combining first principles approaches in the inspiral and merger-ringdown regime. The resulting model reduces overall reliance on NR while retaining the accuracy required for current generation gravitational wave detectors. The model is implemented using best practices in software development including unit testing and continuous integration, making it a robust and reliable tool for gravitational waveform modeling. It also explores the application of orbital stability and dynamics systems techniques to resolving an open problem in the description of circumbinary disks. The resulting picture provided substantial predictive power and opens the possibility of exploring the morphology of the circumbinary gap and disk-binary dynamics in a broad range of astrophysical scenarios.

The broader goal of this dissertation has been to move black hole binary modeling toward de-

scriptions that are not only accurate, but also physically interpretable. By connecting waveform morphology and disk structure to dynamical instabilities, this work provides a framework for understanding how black hole binaries generate both gravitational-wave and electromagnetic observables. This framework will become increasingly important in the multi-messenger era, where future detectors will require models that are accurate, computationally efficient, and grounded in the underlying dynamics of the source.

It is crucial to emphasize that this thesis does not claim that compact binary coalescences, or the massive black hole binary environments, are inherently unstable dynamical systems. Rather, the central claim is that dynamical instabilities exist *within the modeling* of these systems, and leveraging them bridges massive gaps in our theoretical understanding. These instabilities emerge from simplified physical models, but when exploited, they provide crucial interpretability and predictive power to otherwise opaque dynamics. For circumbinary disks, this interpretative power arises from the chaos intrinsic to the restricted three-body problem, which dictates disk dynamics in the absence of complex viscous or thermodynamic transport. For gravitational-wave modeling, the Kerr geometry exhibits unstable orbits for light-like systems that tie the properties of the merger emission to a surprisingly neat description. This mirrors the underlying philosophy of the Effective-One-Body framework, where the EOB metric is not an approximate spacetime of the combined binary system, but rather an abstract mathematical entity that imparts much-needed interpretability to the conservative post-Newtonian problem.

Ultimately, the successful resolution of these two distinct problems highlights the primary thesis objective: that black hole binary observables can often be understood by locating the boundary between stable and unstable dynamics. Whether evaluating the complex dynamics of black hole binary mergers (governed by the light ring) or their surrounding circumbinary disks (governed by epicyclic orbits), these systems can be elegantly understood and powerfully predicted through a common mathematical language.

Bibliography

[1] Abac, A. G. et al., "GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run," 8 2025.

[2] Blanchet, L., "Post-Newtonian Theory for Gravitational Waves," *Living Rev. Rel.*, Vol. 17, 2014, pp. 2.

[3] Yu, A. W. et al., "Overview of the NASA LISA laser system development," *Class. Quant. Grav.*, Vol. 43, No. 3, 2026, pp. 035016.

[4] Milosavljevic, M. and Phinney, E. S., "The Afterglow of massive black hole coalescence," *Astrophys. J. Lett.*, Vol. 622, 2005, pp. L93–L96.

[5] Tang, Y., Haiman, Z., and Macfadyen, A., "The late inspiral of supermassive black hole binaries with circumbinary gas discs in the LISA band," *Mon. Not. Roy. Astron. Soc.*, Vol. 476, No. 2, 2018, pp. 2249–2257.

[6] Dittmann, A. J., Ryan, G., and Miller, M. C., "The Decoupling of Binaries from Their Circumbinary Disks," *Astrophys. J. Lett.*, Vol. 949, No. 2, 2023, pp. L30.

[7] Misner, C. W., Thorne, K. S., and Wheeler, J. A., *Gravitation*, W. H. Freeman, San Francisco, 1973.

[8] Einstein, A. B., Infeld, L., and Hoffmann, B., "The Gravitational equations and the problem of motion," *Annals of Mathematics*, Vol. 39, 1938, pp. 65–100.

[9] Buonanno, A. and Damour, T., "Effective one-body approach to general relativistic two-body dynamics," *Phys. Rev. D*, Vol. 59, 1999, pp. 084006.

[10] Goldstein, H., Poole, C., and Safko, J., *Classical mechanics*, 2002.

[11] Needham, T., *Visual complex analysis*, Oxford Univ. Press, Oxford, 2002.

[12] Damour, T. and Nagar, A., "Faithful effective-one-body waveforms of small-mass-ratio coalescing black-hole binaries," *Phys. Rev. D*, Vol. 76, 2007, pp. 064028.

[13] Pompili, L. et al., "Laying the foundation of the effective-one-body waveform models SEOBNRv5: Improved accuracy and efficiency for spinning nonprecessing binary black holes," *Phys. Rev. D*, Vol. 108, No. 12, 2023, pp. 124035.

[14] Mahesh, S., McWilliams, S. T., and Etienne, Z., "Combining effective one-body inspirals and backwards one-body merger-ringdowns for aligned spin black hole binaries," *Phys. Rev. D*, Vol. 112, Dec 2025, pp. 124065.

[15] Mészáros, P., Fox, D. B., Hanna, C., and Murase, K., "Multi-Messenger Astrophysics," *Nature Rev. Phys.*, Vol. 1, 2019, pp. 585–599.

[16] Goldreich, P. and Tremaine, S., "Disk-satellite interactions." *The Astrophysical Journal*, Vol. 241, Oct. 1980, pp. 425–441.

[17] Mahesh, S., McWilliams, S. T., and Pirog, M., "Analytical and Numerical Analysis of Circumbinary Disk Dynamics. I. Coplanar Systems," *Astrophys. J.*, Vol. 973, No. 1, 2024, pp. 18.

[18] Poisson, E. and Will, C. M., *Gravity*, Cambridge University Press, Cambridge, England, June 2014.

[19] Pirog, M., Mahesh, S., and McWilliams, S. T., "Analytical and Numerical Methods for Circumbinary Disk Dynamics. II. Inclined Disks," *Astrophys. J.*, Vol. 980, No. 1, 2025, pp. 130.

[20] Nelder, J. A. and Mead, R., "A Simplex Method for Function Minimization," *The Computer Journal*, Vol. 7, No. 4, 01 1965, pp. 308–313.

[21] Kankani, A. and McWilliams, S. T., "BOB the (Waveform) Builder: Optimizing Analytical Black-Hole Binary Merger Waveforms," 10 2025.

[22] Szebehely, V., *Theory of orbits. The restricted problem of three bodies*, 1967.

[23] Langford, A. and Weiss, L. M., "A Dynamical Systems Approach to the Theory of Circumbinary Orbits in the Circular Restricted Problem," *The Astronomical Journal*, Vol. 165, No. 4, April 2023, pp. 140.

[24] Siwek, M., Weinberger, R., and Hernquist, L., "Orbital evolution of binaries in circumbinary discs," *Mon. Not. Roy. Astron. Soc.*, Vol. 522, No. 2, 2023, pp. 2707–2717.

[25] Dittmann, A. J., Ryan, G., and Combi, L., "Eccentric Binaries Accreting from Thin Disks: Orbital Evolution," *Astrophys. J.*, Vol. 1001, No. 2, 2026, pp. 132.