APS physics

CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Evolution of massive black hole binaries in rotating stellar nuclei: Implications for gravitational wave detection

Alexander Rasskazov and David Merritt

Phys. Rev. D **95**, 084032 — Published 19 April 2017

DOI: 10.1103/PhysRevD.95.084032

Evolution Of Massive Black Hole Binaries In Rotating Stellar Nuclei: Implications For Gravitational Wave Detection

Alexander Rasskazov and David Merritt[1]

[1]*School of Physics and Astronomy and Center for Computational Relativity and Gravitation, Rochester Institute of Technology, Rochester, NY 14623*

We compute the isotropic gravitational wave (GW) background produced by binary supermassive black holes (SBHs) in galactic nuclei. In our model, massive binaries evolve at early times via gravitational-slingshot interaction with nearby stars, and at later times by the emission of GWs. Our expressions for the rate of binary hardening in the "stellar" regime are taken from the recent work of Vasiliev et al., who show that in the non-axisymmetric galaxies expected to form via mergers, stars are supplied to the center at high enough rates to ensure binary coalescence on Gyr timescales. We also include, for the first time, the extra degrees of freedom associated with evolution of the binary's orbital plane; in rotating nuclei, interaction with stars causes the orientation and the eccentricity of a massive binary to change in tandem, leading in some cases to very high eccentricities ($e > 0.9$) before the binary enters the GW-dominated regime. We argue that previous studies have overestimated the mean ratio of SBH mass to galaxy bulge mass by factors of 2 – 3. In the frequency regime currently accessible to pulsar timing arrays (PTAs), our assumptions imply a factor 2 – 3 reduction in the characteristic strain compared with the values computed in most recent studies, removing the tension that currently exists between model predictions and the non-detection of GWs.

PACS numbers: 04.30.-w, 04.30.Tv, 97.60.Lf

I. INTRODUCTION

Pulsar timing arrays (PTAs [6]) are designed to detect the low-frequency (∼nHz) gravitational wave (GW) background that would be generated by a population of binary supermassive black holes (SBHs). In the simplest model – a cosmologically homogeneous and isotropic population of massive binaries on circular orbits, which evolve solely due to GW emission – the characteristic strain of the GW-induced distortions has a frequency dependence given by [24]

$$h_c(f) = A_{\rm yr}\left(\frac{f}{f_{\rm yr}}\right)^{-2/3} \tag{1}$$

where $f_{\rm yr} \equiv 1/{\rm yr}$. The corresponding energy density per unit logarithmic frequency, expressed in terms of the cosmological critical density $\rho_c = 3H_0^2/(8\pi G)$, is [24]

$$\Omega_{\rm GW}(f) \equiv \frac{1}{\rho_c}\frac{d\rho_{\rm GW}}{d\ln f} = \frac{2\pi^2}{3H_0^2}f^2h_c^2(f) \tag{2}$$

where $H_0 \approx 7.2\times10^{-11}$ yr^{-1} is the Hubble constant. The parameter $A_{\rm yr}$, the predicted strain at a frequency of one inverse year, depends in a possibly complicated way on the astrophysical parameters that characterize the binary population, including the mass function of SBHs; the distribution of binary mass ratios; the galaxy merger rate; and the rate at which binaries attain separations small enough ($\lesssim 10^{-2}$ pc) that GW emission can dominate their evolution. Theoretical estimates of $A_{\rm yr}$ typically lie in the range $\sim 10^{-15}-10^{-14}$ [20, 27, 32]. Detection of GWs in this frequency regime would provide robust evidence for the existence of binary SBHs and would allow the astrophysical parameters that determine the frequency spectrum to be constrained [33].

The peak sensitivity of a PTA occurs at a frequency that is roughly the inverse of the time over which pulsar timing data has been collected [3]. That time is now roughly one decade. At these lower frequencies, $f \ll f_{\rm yr}$, the characteristic strain is expected to differ from the prediction of Eq. (1). The semimajor axis of a binary SBH with orbital period P is

$$a = 1.0\times10^{-2}\left(\frac{M_{12}}{10^8\mathcal{M}_\odot}\right)^{1/3}\left(\frac{P}{10\ {\rm yr}}\right)^{2/3}{\rm pc} \tag{3}$$

with $M_{12} = M_1+M_2$ the mass of the binary. At separations corresponding to orbital frequencies less than $\sim(10\ {\rm yr})^{-1}$, i.e. $a \gtrsim 10^{-2}$ pc, a massive binary is expected to evolve primarily via interactions with ambient stars and gas in the galactic nucleus rather than by GW emission [21, chapter 8]. Furthermore there is no compelling reason why binaries at these large separations should be on circular orbits; it is only at smaller separations that GW emission becomes

effective at reducing eccentricities. Both considerations would predict a reduction in h_c below a certain frequency, compared with Eq. (1). Sampson et al. [11] suggested a simple parametrization of the GW spectrum describing a population of binaries that have been "environmentally" influenced:

$$h_c(f) = A\frac{(f/f_{\rm yr})^{-2/3}}{\left[1+(f_{\rm bend}/f)^{\kappa}\right]^{1/2}}, \tag{4a}$$

$$A = A_{\rm yr}\left[1+(f_{\rm bend}/f_{\rm yr})^{\kappa}\right]^{1/2} \tag{4b}$$

where $f_{\rm bend}$ is understood as the orbital frequency below which environmental interactions dominate the binary's evolution and κ is determined by the type of interaction; in the case of stellar scattering considered in this paper, $\kappa = 10/3$. Simple evolutionary models suggest $f_{\rm bend} \approx 10^{-9}$ Hz [33], a frequency regime that is beginning to be probed by PTAs.

Analysis of pulsar timing data by three groups has so far succeeded in placing only upper limits on $A_{\rm yr}$: 3.0×10^{-15} (EPTA [8]); 1.0×10^{-15} (PPTA [37]); and 1.5×10^{-15} (NANOGrav [1]). These values are generally interpreted as being "in tension with" the predictions of some theoretical models; for instance, McWilliams et al. [20] predict $A_{\rm yr}\approx10^{-14.5}$. According to Shannon et al. [37, Table S8], even the models with the lowest predicted $A_{\rm yr}$ ([32] and [27]) have only 9% and 21% probability, respectively, of being consistent with the limit derived from their observations.

In the present paper we present a new calculation of $h_c(f)$. Our treatment differs in three important ways from previous ones.

1. *Binary hardening rates.* The "final-parsec problem" [9] refers to the possibility that massive binaries might stall at separations much greater than required for the emission of detectable GWs. Here we make use of recent work [43, 44] which shows that even in "collisionless" (gas-free, long-relaxation-time) nuclei, like those of massive galaxies, a modest departure from exact axisymmetry is sufficient to keep a massive binary shrinking. In such a nucleus, the binary hardening rate decreases with time, but interactions with stars are nevertheless able to drive the binary to coalescence on a timescale of order 1 Gyr or less. In an accurately axisymmetric nucleus, hardening rates are low enough that coalescence is not likely in a Hubble time, but a binary can still enter into the PTA frequency regime. Guided by the observations, we characterize high-luminosity bulges as triaxial and low-luminosity bulges as axisymmetric, then use the expressions derived in the cited papers to compute binary hardening rates for the different galaxy populations.

2. *Eccentricity evolution.* In the regime where binary hardening is driven by interaction with stars, as opposed to GW emission, eccentricity evolution has been shown to be modest, at least in spherical nonrotating nuclei [25]. On this basis, most discussion of the stochastic GW spectrum have assumed zero eccentricities. The situation can be very different in the case of nuclei with significant rotation, particularly if the angular momentum of the massive binary is initially misaligned with that of the nucleus [26]. Given such initial conditions, the binary's orbital plane rotates to bring its angular momentum vector more in alignment with the nuclear rotation axis, and the binary's eccentricity simultaneously increases, sometimes to very large values ($e > 0.9$). As the binary aligns fully with the nucleus, its eccentricity returns again to lower values. Our models are the first to include these additional degrees of freedom.

3. *SBH demographics.* Arzoumanian et al. [1] adopted prior probability distributions for $A_{\rm yr}$ from the modeling studies of Sesana [32, S13] and McWilliams et al. [20, MOP] and used them to infer posterior distributions of the parameters A, $f_{\rm bend}$ and κ in Eq. (4). MOP assumed a mean ratio of SBH mass to bulge mass of ~ 0.003, and they further augmented the SBH mass function to account for a putative population of "overmassive" SBHs in giant galaxies. The resulting estimate of $A_{\rm yr}\approx10^{-14.4}$ was found by Arzoumanian et al. to be difficult to reconcile with the PTA data. In order to limit the predicted contribution at low frequencies, a value $f_{\rm bend}\gtrsim10^{-8}$ Hz was required, substantially larger than the value expected physically unless nuclear densities in giant galaxies exceed $\sim10^{-3}\mathcal{M}_\odot{\rm pc}^{-3}$. This possibility was judged unlikely by Arzoumanian et al., and those authors suggested that the MOP prior might be in error (too large), either because binary SBHs typically stall at separations outside the PTA band, or because the characterization adopted by MOP for the parent population of SBHs was somehow incorrect. Shannon et al. [37] reached a similar conclusion. We argue in fact that both S13 and MOP substantially overestimated the mean ratio of SBH mass to bulge mass. A more conservative (in the sense of being based on more compelling data) estimate of this ratio is 0.001 which is the fiducial value we adopt here.

The last of these assumptions is most important at setting the predicted amplitude of $h_c(f)$ at frequencies that lie in the current range of PTA sensitivity. Our models have $A_{\rm yr} < 10^{-15}$, a factor of at least two lower than in most other recent calculations of $h_c(f)$ [20, 27, 32]. Although we make no attempt to model the PTA data in the manner of Arzoumanian et al. [1] or Shannon et al. [37], such a low value for $A_{\rm yr}$ would presumably (1) be consistent with a physically more plausible range of parameters $\{A, f_{\rm bend}, \kappa\}$ in Eq. (4); (2) remove the "tension" between the non-detection of a stochastic GW signal by the various groups and the predictions of MOP; and (3) unfortunately,

imply that a PTA detection of the stochastic GW background from inspiralling SBHs is not likely to occur in the immediate future.

Our physical model for the formation and evolution of massive binaries is presented in §II; this section also includes the derivation of a formula for $h_c(f)$ that, for the first time, allows for any possible functional form of the time-dependence of the binary hardening rate. §III presents estimates of the characteristic strain spectrum and its dependence on the parameters that define the initial population of binaries and their host galaxies. §IV sums up and discusses the implications of our results for the detection of isotropic gravitational wave background via PTAs.

II. METHOD

We assume that shortly after two galaxies merge, the two SBHs form a "hard binary"[1] at the center of the merger product. The components of the binary have masses M_1 and M_2; $M_{12} = M_1 + M_2$, $\mu_{12} = M_1 M_2/(M_1 + M_2)$ is the binary's reduced mass and $q \equiv M_2/M_1 \leq 1$ its mass ratio (or $\mathcal{Q} = \mu/M_{12} = q/(1+q)^2$ its symmetric mass ratio). The initial semimajor axis is the "hard-binary separation" a_h:

$$a_h \equiv \frac{G\mu}{4\sigma^2} \approx 2.7\,\mathcal{Q}\,\frac{M_{12}}{10^8\mathcal{M}_\odot}\left(\frac{\sigma}{200\,\mathrm{km\,s^{-1}}}\right)^{-2}\mathrm{pc} \tag{5}$$

where σ is the 1d stellar velocity dispersion in the nucleus, and the initial orbital frequency is

$$\frac{1}{P} = \frac{\sqrt{GM_{12}}}{2\pi a_h^{3/2}} = \frac{4\sigma^3}{\pi\mathcal{Q}^{3/2}GM_{12}} = \frac{4\sigma}{\pi\mathcal{Q}^{3/2}r_h} \tag{6a}$$

$$\approx 0.8\times10^{-12}\mathcal{Q}^{-3/2}\left(\frac{\sigma}{200\ \mathrm{km\ s^{-1}}}\right)\left(\frac{r_h}{10\ \mathrm{pc}}\right)^{-1}\mathrm{Hz} \tag{6b}$$

where $r_h \equiv GM_{12}/\sigma^2$ is the binary's gravitational influence radius. For reasonable values of the parameters in Eq. (6), this frequency is below the limit detectable by PTAs and so we ignore the contribution of binaries with $a > a_h$ to the GW background. In what follows, we identify the galaxy merger rate with the rate of formation of binaries having $a = a_h$.

A. Gravitational wave background from a population of massive binaries

Consider the set of binaries that form, at any cosmological time, but with the same values of $\{M_1, M_2, \boldsymbol{L}\}$, where $L^2 = GM_{12}a_h(1-e^2)$ and e is the binary's eccentricity at formation ($a = a_h$). After formation, the separation evolves as $a = a(t), t_h \leq t \leq t_c$ where t_h is the formation time and t_c is the time at which the two SBHs coalesce. Define $N(a,t)da$ to be the number of binaries from this set, per unit comoving volume, that have separations in the interval a to $a+da$ at time t (N can also depend on $M_1, M_2, \boldsymbol{L}$ and the properties of the host galaxy). The function N obeys a continuity equation:

$$\frac{\partial N}{\partial t} + \frac{\partial}{\partial a}\left(N\dot{a}\right) = 0 \tag{7}$$

with solution

$$N(a,t)\,da = \frac{\dot{\mathcal{N}}_m(t_h)}{\left|\dot{a}(a)\right|}\,da \tag{8}$$

with $\dot{\mathcal{N}}$ the galaxy merger rate:

$$\dot{\mathcal{N}}_m(t_h) \equiv -N(a_h,t_h)\,\dot{a}(a_h) \tag{9}$$

i.e. the rate, per comoving volume, at which galaxies are merging at time t_h (we neglect the time it takes for a binary to become hard after the galaxy merger).

---

[1] Defined as a binary that ejects passing stars at typical velocities greater than the escape velocity from the nucleus [72].

By analogy with Ravi et al. [27], the specific intensity of GWs at the Earth from binaries with semimajor axes between a and $a+da$ at redshifts between z and $z+dz$ is

$$dI = \frac{L(f_r,a)}{4\pi d_L^2}\frac{df_r}{df}N(a,z)\frac{dV_c}{dz}da\,dz, \tag{10}$$

where f is the observed GW frequency, $f_r=(1+z)f$ is the rest-frame frequency, d_L is the luminosity distance,

$$\begin{aligned}
L(f_r,a) &= \frac{32}{5}\frac{G^{7/3}}{c^5}\mathcal{M}^{10/3}(2\pi f_{\rm orb})^{10/3}\sum_{n=1}^{\infty}g(n,e)\,\delta(f_r-nf_{\rm orb}), &(11a)\\
\mathcal{M} &= M_{12}\mathcal{Q}^{3/5}, &(11b)\\
f_{\rm orb} &= \frac{1}{2\pi}\sqrt{\frac{GM_{12}}{a^3}}, &(11c)\\
g(n,e) &= \frac{n^4}{32}\Bigg[\left\{J_{n-2}(ne)-2eJ_{n-1}(ne)+\frac{2}{n}J_n(ne)+2eJ_{n+1}(ne)-J_{n+2}(ne)\right\}^2 &(11d)\\
&+ \left(1-e^2\right)\left\{J_{n-2}(ne)-2J_n(ne)+J_{n+2}(ne)\right\}^2+\frac{4}{3n^2}J_n^2(ne)\Bigg]
\end{aligned}$$

is the GW luminosity per unit rest-frame frequency of a binary with semimajor axis a and eccentricity e (which is determined by a since we are assuming here the same values for the other initial parameters) and

$$\frac{dV_c}{dz}=\frac{4\pi c d_L^2}{H(z)(1+z)^2} \tag{12}$$

is the comoving volume per unit z. Eq. (11) is taken from [23].

Now, the energy density in GWs at the Earth per logarithmic frequency unit is

$$\begin{aligned}
\Omega_{\rm GW}\rho_c c^2 &= \frac{f}{c}\int_0^{z_{\rm max}}\int_{a_c}^{a_h}\frac{dI}{dadz}dadz &(13a)\\
&= f\int_0^{z_{\rm max}}\int_{a_c}^{a_h}\frac{L(f_r,a)N(a,z)}{H(z)(1+z)}dadz &(13b)\\
&= f\int_{t_r(z_{\rm max})}^{0}dt_r\int_{a_c}^{a_h}L(f_r,a)N(a,t_r)da &(13c)
\end{aligned}$$

where a_c is the orbital separation at which BHs coalesce and

$$\begin{aligned}
t_r(z) &= \int_z^0\frac{dz'}{H(z')(1+z')}, &(14a)\\
H(z) &= H_0\left[\Omega_{\rm M}(1+z)^3+\Omega_\Lambda\right]^{1/2}, &(14b)\\
H_0 &= 67.7\,{\rm km/s/Mpc},\quad \Omega_{\rm M}=0.31,\quad \Omega_\Lambda=0.69 &(14c)
\end{aligned}$$

is the proper time. Substituting the expressions (8) for $N(a,t_r)$ and (11) for $L(f_r,a)$ into Eq. (13c) and using the following property of the δ function:

$$\begin{aligned}
\int y(t)\delta(x(t))dt &= \frac{y(t_0)}{dx/dt(t_0)}, &(15a)\\
x(t_0)&=0, &(15b)
\end{aligned}$$

we can carry out the integration over a, yielding

$$\Omega_{\rm GW}\rho_c c^2 = \frac{32}{5}\frac{G^{7/3}}{c^5}\mathcal{M}^{10/3}f\int_{t_r(z_{\rm max})}^{0}dt_r\sum_{a_c<a_n<a_h}\left(n\left|\frac{df_{\rm orb}}{da}(a_n)\right|\right)^{-1}\left(\frac{2\pi f_r}{n}\right)^{10/3}g(n,e(a_n))\frac{\dot{\mathcal{N}}_m(t_h(a_n,t_r))}{\left|\dot{a}(a_n)\right|} \tag{16}$$

where $t_h(a,t_r)$ is the formation time of a binary that has semimajor axis a at time t_r, and a_n is the binary semimajor axis at which the orbital frequency is f_r/n, so that the n-th harmonic is contributing to the total energy radiated at frequency f_r:

$$f_{\rm orb}(a_n) = \frac{1}{2\pi}\sqrt{\frac{GM_{12}}{a_n^3}} = \frac{f_r}{n} = \frac{f(1+z(t_r))}{n} \tag{17}$$

It is straightforward to show that

$$\left(n\left|\frac{df_{\rm orb}}{da}(a_n)\right|\right)^{-1} = \frac{4\pi}{3n}\left(\frac{n}{2\pi f_r}\right)^{5/3}(GM_{12})^{1/3}, \tag{18}$$

which gives us

$$\Omega_{\rm GW}\rho_c c^2 \;=\; \frac{32}{5}\frac{G^{7/3}}{c^5}\mathcal{M}^{10/3}(GM_{12})^{1/3} f\int_{t_r(z_{\rm max})}^{0} dt_r \sum_{a_c<a_n<a_h}\frac{4\pi}{3n}\left(\frac{2\pi f_r}{n}\right)^{5/3} g(n,e(a_n))\frac{\dot{\mathcal{N}}_m(t_h(a_n,t_r))}{\left|\dot{a}(a_n)\right|} \tag{19}$$

We define the binary hardening rate, S, in the usual way as

$$S \equiv \frac{d}{dt}\left(\frac{1}{a}\right) \tag{20}$$

and S_h as the hardening rate in full loss-cone approximation (one that would occur if the distribution of stars in phase space were not affected by the presence of the binary; see Eq. 39 and 42). The initial hardening timescale (again, in full loss-cone approximation) is

$$t_h \equiv \left|\frac{a}{\dot{a}}\right|_{a=a_h} = \frac{1}{a_h S_h} \tag{21}$$

so

$$\dot{a} = \frac{d(a/a_h)}{d(t/t_h)}\frac{a_h}{t_h} = \frac{d(a/a_h)}{d(t/t_h)}a_h^2 S_h. \tag{22}$$

Eq. (19) can then be written

$$\begin{aligned}\Omega_{\rm GW}\rho_c c^2 \;&=\; \frac{32}{5}\frac{G^{7/3}}{c^5}\mathcal{M}^{10/3}(GM_{12})^{1/3}(a_h^2 S_h)^{-1} f \times \\ &\times\; \int_{t_r(z_{\rm max})}^{0} dt_r \sum_{a_c<a_n<a_h}\frac{4\pi}{3n}\left(\frac{2\pi f_r}{n}\right)^{5/3} g(n,e(a_n))\,\dot{\mathcal{N}}_m(t_h(a_n,t_r))\left|\frac{d(a/a_h)}{d(t/t_h)}(a_n)\right|^{-1}.\end{aligned} \tag{23}$$

It is convenient to return to z as the integration variable, in terms of which

$$\begin{aligned}\Omega_{\rm GW}\rho_c c^2 \;&=\; \frac{32}{5}\frac{G^{7/3}}{c^5}\mathcal{M}^{10/3}(GM_{12})^{1/3}(a_h^2 S_h)^{-1} f \\ &\times\; \int_0^{z_{\rm max}} \frac{dz}{H(z)(1+z)} \sum_{a_c<a_n<a_h}\frac{4\pi}{3n}\left(\frac{2\pi f_r}{n}\right)^{5/3} g(n,e(a_n))\,\dot{\mathcal{N}}_m(z_h(a_n,z))\left|\frac{d(a/a_h)}{d(t/t_h)}(a_n)\right|^{-1} \qquad (24a) \\ &=\; \frac{64\,(2\pi f)^{8/3}}{15\,Gc^5 H_0}(GM_{12})^{11/3}\mathcal{Q}^2(a_h^2 S_h)^{-1} \\ &\times\; \int_0^{z_{\rm max}} \frac{(1+z)^{2/3}}{\left[\Omega_{\rm M}(1+z)^3+\Omega_\Lambda\right]^{1/2}}dz \sum_{a_c<a_n<a_h} n^{-8/3} g(n,e(a_n))\,\dot{\mathcal{N}}_m(z_h(a_n,z))\left|\frac{d(a/a_h)}{d(t/t_h)}(a_n)\right|^{-1} \qquad (24b)\end{aligned}$$

where $z_h(a,z)$ is the formation redshift of a binary which has semimajor axis a at redshift z.

According to Eqs. (5) and (42),

$$a_h^2 S_h = \left(\frac{G\mathcal{Q}M_{12}}{4\sigma^2}\right)^2 \times 4\sqrt{\frac{GM_{12}}{(GM_{12}/\sigma^2)^5}} = \frac{1}{4}\mathcal{Q}^2\sigma \tag{25}$$

so

$$\Omega_{\rm GW}\rho_c c^2 = \frac{256\,(2\pi f)^{8/3}}{15\,Gc^5H_0}(GM_{12})^{11/3}\sigma^{-1}(M_{12}) \times \int_0^{z_{\rm max}}\frac{(1+z)^{2/3}}{\left[\Omega_{\rm M}(1+z)^3+\Omega_\Lambda\right]^{1/2}}dz\sum_{a_c<a_n<a_h} n^{-8/3}g(n,e(a_n))\,\dot{\mathcal{N}}_m(z_h(a_n,z))\left|\frac{d(a/a_h)}{d(t/t_h)}(a_n)\right|^{-1}, \quad (26a)$$

$$\sigma(M_{12}) = 180\left(\frac{M_{12}}{10^8\mathcal{M}_\odot}\right)^{1/5}\,{\rm km\,s^{-1}}. \quad (26b)$$

(The second equation is the $M-\sigma$ relation, as given below in Eq. 41). We now relax our assumption of a single set of values $\{M_{12}, q, e_0\}$ and generalize Eq. (26) to consider a population of binaries with different initial parameters. Furthermore we add two parameters related to nuclear rotation: θ_0, the binary's initial inclination, and η, a parameter that determines the degree of ordered rotation of a nucleus; both parameters are defined and discussed in more detail in §II D. We redefine the merger rate as the rate per unit comoving volume, and per unit of M_{12}, q, e_0, θ_0 and η. The resulting expression is

$$\Omega_{\rm GW}\rho_c c^2 = \frac{256\,(2\pi f)^{8/3}}{15\,Gc^5H_0}\int_{1/2}^1 d\eta\int_{q_{\rm min}}^1 dq\int_{M_{12,\rm min}}^{M_{12,\rm max}} dM_{12}\int_0^\pi d\theta_0\int_0^1 de_0\int_0^{z_{\rm max}} dz\,(GM_{12})^{11/3}\sigma^{-1}(M_{12}) \times \frac{(1+z)^{2/3}}{\left[\Omega_{\rm M}(1+z)^3+\Omega_\Lambda\right]^{1/2}} \times \sum_{a_c<a_n<a_h} n^{-8/3}g(n,e(a_n))\,\dot{\mathcal{N}}_m(z_h(a_n,z),M_{12},q,e_0,\theta_0,\eta)\left|\frac{d(a/a_h)}{d(t/t_h)}(a_n)\right|^{-1}. \quad (27)$$

Note our additional assumption that host galaxy properties, such as σ, are determined by M_{12}. The characteristic strain, $h_c(f)$, is given in terms of $\Omega_{\rm GW}(f)$ by Eq. (2), so that

$$h_c(f) = \left[\frac{1024\pi\,(2\pi f)^{2/3}}{15c^7H_0}\int_{1/2}^1 d\eta\int_{q_{\rm min}}^1 dq\int_{M_{12,\rm min}}^{M_{12,\rm max}} dM_{12}\int_0^\pi d\theta_0\int_0^1 de_0\int_0^{z_{\rm max}} dz\,(GM_{12})^{11/3}\sigma^{-1}(M_{12}) \times \frac{(1+z)^{2/3}}{\left[\Omega_{\rm M}(1+z)^3+\Omega_\Lambda\right]^{1/2}} \times \sum_{a_c<a_n<a_h} n^{-8/3}g(n,e(a_n))\,\dot{\mathcal{N}}_m(z_h(a_n,z),M_{12},q,e_0,\theta_0,\eta)\left|\frac{d(a/a_h)}{d(t/t_h)}(a_n)\right|^{-1}\right]^{1/2}. \quad (28)$$

In what follows, we adopt the following limits on the integrals that appear in Eq. (28): $z_{\rm max}=3$, $q_{\rm min}=1/10$, $M_{12,\rm min}=10^6\mathcal{M}_\odot$, $M_{12,\rm max}=10^{10}\mathcal{M}_\odot$. As Fig. 1 shows, in this way, we account for more than 95% of the total signal.

B. Galaxy merger rate

Galaxy merger rates are customarily expressed "per galaxy", i.e., $\dot{N}_{\rm mergers}(M_{\rm gal},z)$ is the rate at which a "primary" galaxy, of mass $M_{\rm gal}$, experiences mergers with other galaxies at redshift z. We convert such an expression into the merger rate per unit galaxy mass, per unit mass ratio, by multiplying it by (i) the galaxy mass function at a given redshift $\phi(M_{\rm gal},z)$; and (ii) the distribution of galaxy mass ratios $q_{\rm gal}$, which we assume to be $\propto 1/q_{\rm gal}$ following [48] and [32]:

$$\dot{\mathcal{N}}_m\,dq\,dM_{12}\,de_0\,d\theta_0 = \frac{dN_{\rm mergers}}{dt}(M_{\rm gal},z)\frac{1}{q_{\rm gal}(-\ln q_{\rm lim})}\phi(M_{\rm gal},z)\frac{dM_{\rm gal}}{dM_{12}}\mathcal{F}(e_0,\theta_0,\eta)\,dq\,dM_{12}\,de_0\,d\theta_0. \quad (29)$$

Here $q_{\rm lim}$ is the minimum mass ratio selected in counting galaxy pairs and $\mathcal{F}(e_0,\theta_0,\eta)$ is the joint distribution of binary initial parameters e_0, θ_0 and of η. Unfortunately, little is known about $\mathcal{F}$. With regard to its dependence on M_{12}, q and z, we take a "maximally-uninformed" stance and posit no dependence.

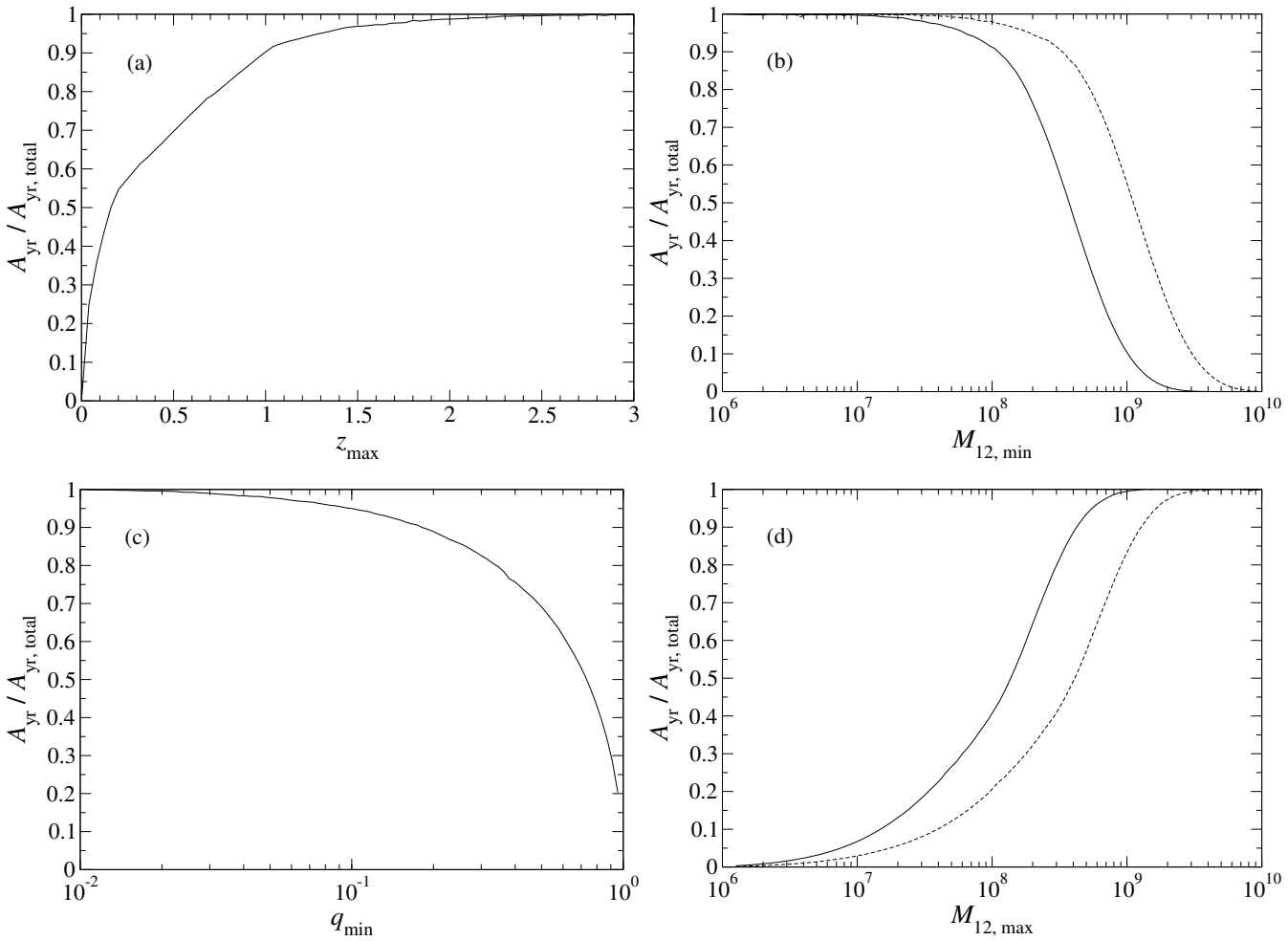

FIG. 1: Fraction of the total GW strain at $f = 1\mathrm{yr}^{-1}$ contributed by massive binaries with (a) $z < z_{\max}$, (b) $M_{12} > M_{12,\min}$, (c) $q > q_{\min}$ or (d) $M_{12} < M_{12,\max}$. If not otherwise specified, $z_{\max} = 4$, $q_{\min} = 1/100$, $M_{12,\min} = 10^6 \mathcal{M}_\odot$ and $M_{12,\max} = 10^{10} \mathcal{M}_\odot$. Plots assume circular-orbit binaries and SBH-bulge mass ratio $\beta = 0.001$ (straight lines) or $\beta = 0.003$ (dashed lines); for (a) and (c) both lines look identical.

In what follows, we consider the following possibilities for the dependence of $\mathcal{F}$ on e_0, θ_0 and η. In each case we normalize $\mathcal{F}$ so that

$$\int_{1/2}^{1} \int_{0}^{1} \int_{0}^{\pi} \mathcal{F}(e_0, \theta_0, \eta)\, d\theta_0\, de_0\, d\eta = 1.$$

1. e_0, θ_0 and η are the same for all binaries:

$$\mathcal{F}\, d\theta_0\, de_0\, d\eta = \delta(e_0 - e_{0,1})\, \delta(\theta_0 - \theta_{0,1})\, \delta(\eta - \eta_1)\, de_0\, d\theta_0\, d\eta. \tag{30}$$

   This includes the simplest case in which all binaries are initially circular, $e_{0,1} = 0$; in this special case, the distributions over θ_0 and η do not matter because a circular binary remains circular in the course of its evolution.

2. e_0 has a "thermal" distribution, $dN/de_0 = 2e_0$, and the binary is either corotating from the beginning ($\theta_0 = 0$) or its orbital plane has no preferred initial direction:

$$\mathcal{F}\, d\theta_0\, de_0 \;=\; \mathcal{F}_1(\eta) \cdot 2e_0\, de_0\, \delta(\theta_0)\, d\theta_0, \text{ or} \tag{31a}$$
$$\mathcal{F}\, d\theta_0\, de_0 \;=\; \mathcal{F}_1(\eta) \cdot e_0\, de_0\, \sin\theta_0\, d\theta_0. \tag{31b}$$

   The functional form of $\mathcal{F}_1(\eta)$ is discussed in Section II D.

For the galaxy (stellar) mass function $\phi(M_{\rm gal}, z)$, we adopt the analytic expressions of [40] (Schechter fit) for $z \leq 0.2$ and [14] (double Schechter fit) for $z > 0.2$. With the possible exception of the highest galaxy masses, both expressions agree well with determinations by other groups (see Fig. 7 in [40] and Fig. 6 in [41]). The galaxy merger rate and the relation between galaxy mass and SBH mass we use are discussed in the next two subsections.

As for the merger rate per galaxy, we adopt the analytic expression of [45]:

$$\frac{dN_{\rm mergers}}{dt} = 0.053\,{\rm Gyr}^{-1}\left(\frac{M_{\rm gal}}{10^{10.7}\mathcal{M}_\odot}\right)^{0.3}\frac{(1+z)^{2.2}}{1+z/8}. \tag{32}$$

Xu et al. [45] obtained this expression by dividing their observed galaxy pair fraction by the average major merger timescale ($\sim 300\,{\rm Myr}$) taken from Lotz et al. [17], who performed hydrodynamical simulations of disk galaxy mergers for a number of different masses, mass ratios and initial orbits.[2]

Given that the galaxy merger process is very complicated, there are undoubtedly many uncertain factors that influence the merger rate, both through the (observed) pair fraction and the (simulated) merger timescales. As shown in Figure 15 of [28], the pair-fraction estimates of Xu et al. [45] are consistent with the average value of the other studies, although a factor of ~ 2 discrepancy between different papers exists. Some of these discrepancies might be caused by differences in photometric completeness levels and blending issues in different wavebands used to identify mergers [69].

As noted above, we equate the galaxy merger rate, Eqs. (29) and (32), with the rate at which "hard" SBH binaries are forming – in other words, we have ignored the time for two SBHs to reach the center of the merger remnant. As shown by Merritt [71] and Dosopoulou & Antonini [70], this time can indeed be long for very low mass ratios q and $q_{\rm gal}$, approaching the Hubble time for $q \lesssim 10^{-3}$. However, for binaries of any mass with $q > 0.1$ (which contribute almost all of the GW background, see Fig. 1c) this time is always shorter than 100 Myr and can be ignored.

The merger timescales from [17] are actually "observability timescales" – the total amount of time a merging pair of galaxies spends at a certain observable merger stage; for example, in the case of [45] the observability criterion is a projected separation between 5 and 20 $h^{-1}\,{\rm kpc}$, and the timescales from [17] are chosen accordingly. This way, the uncertainty in the definitions of the beginning and the end of a merger is eliminated. There are, however, a number of other caveats in this approach:

1. In all of their simulations, Lotz et al. [17] assume both galaxies to be disklike while the galaxies contributing most of the GW signal are elliptical (Fig. 11a). The dependence of merger timescale on galaxy morphology might be one of the largest sources of uncertainty.

2. We use the same average timescale for all mergers of the same mass and redshift, ignoring the dependence on various parameters of a galaxy pair that could be correlated with GW emission, such as mass ratio or initial spin orientations of the galaxies. Lotz et al. found that orientations have little effect for close pairs observed at $5 < r_p < 20\,h^{-1}\,{\rm kpc}$. On the other hand, timescales for equal-mass mergers could be $\sim 30\%$ longer compared to $q = 1/3$. Since binaries with higher q contribute more signal, that could lead us to underestimate the merger timescale and, consequently, overestimate h_c, but only by $\lesssim 15\%$ since $h_c \propto \sqrt{\dot{N}_{\rm mergers}}$.

3. Merger timescales depend on gas fraction: for equal-mass mergers, they become considerably shorter when gas fractions are high [18]. The reason is that the disk galaxies having higher gas fractions are harder to deblend at close separations, which reduces the time interval over which two galaxies can be observed as a close pair. However, this effect vanishes for $q \lesssim 1/3$. Lotz et al. [19] have calculated the average timescale for $1/4 < q < 1$ and three different assumptions about gas fraction and found it to be almost the same as the value we use (330 Myr) for all three cases with a weak dependence on redshift (at least at $z \lesssim 1$).

4. The hydrodynamical simulations of Lotz et al. give a significantly lower merger timescale estimate than the semi-analytical model of Kitzbichler & White [15], who used a mock galaxy catalog derived from the Millennium simulation. Kitzbichler & White assume that the secondary galaxy moves in a circular orbit inside the constant potential of the primary, and this is probably not a good approximation for close pairs of nearly equal mass.

5. All of the simulation timescales adopted here assume highly eccentric orbits of galaxies with pericenter distances $\sim 0.01 - 0.05$ times the virial radii of the progenitors. However, galaxies merging on circular orbits or with larger impact parameters can be identified as close pairs for $15 - 40\%$ longer [16].

---

[2] The simulations assume that observable galaxies are surrounding by dark-matter halos; otherwise merging timescales would be much longer.

6. Also, at higher redshifts the mergers could proceed faster for the reason that the galaxy sizes for a given mass are smaller. According to Huertas-Company et al. [68], the characteristic radii of big ellipticals are ~ 2 times smaller at $z = 1$ compared to $z = 0$.

7. Finally, all of the galaxy merger simulations posit that the observed, luminous parts of galaxies are surrounded by extensive, dynamically-active, dark-matter haloes. If the dark-matter haloes are not present, or if the "dark matter" is not particle in nature, merger times would be much longer; indeed most of the observed interacting pairs would "pass in the night" and never merge [61]. It is not our intention here to stake out a position in the dark-matter debate [58]. But we do note the troubling lack of corroborative evidence for the merger hypothesis, and the fact that some observationally-based studies reach conclusions contradicting the predictions of ΛCDM cosmology, such as overabundance of bulgeless giant galaxies [56, 60] or that giant ellipticals couldn't have been formed from disk galaxy mergers [59].

The effects on $h_c(f)$ of systematic uncertainties in the merger rate are discussed in more detail in Section III (Eq. 51).

C. SBH demographics

We assume a strict proportionality between SBH mass and the mass of the stellar bulge, defining the parameter

$$\beta \equiv M_{\rm BH}/M_{\rm bulge}. \tag{33}$$

Estimates of this quantity have evolved over time; we identify three epochs.

1. Kinematical modeling of early-type galaxies by Magorrian et al. (1998) [49] favored a high value, $\beta \approx 0.006$. This value was immediately seen to be inconsistent with (i.e. larger than) the mean mass ratio in active galaxies, either as predicted by the Soltan argument, or as estimated via reverberation mapping in individual galaxies [52].

2. In 2000, the $M-\sigma$ relation was discovered by restricting the sample to galaxies with clear, prima facie evidence for a Keplerian velocity rise, leading to a much lower estimate, $\beta \approx 0.0012$. This smaller value eliminated the discrepancies with the other two methods [50].

3. Starting around 2006, and continuing until the present day, most authors have sought to be comprehensive, including in their samples essentially every published SBH mass without regard to the presence or absence of a kinematical signal. These studies (as summarized by [47]) find a larger value, $\beta \approx 0.003$, that is once again inconsistent with (i.e. larger than) SBH masses in AGN [51]. This is the value of β assumed in all recent calculations of $h_c(f)$ [20, 32].

We note here a worrisome phenomenon: when stellar (as opposed to gas) kinematical data for a galaxy are re-modeled independently, the results for $M_{\rm BH}$ are often very different than in the "discovery" paper: the best-fit SBH mass is found to be much lower; there is a range of equally-likely masses; or only an upper limit can be established. A recent example is NGC 1277, where claims of a $\sim 2\times10^{10}\mathcal{M}_\odot$ SBH [53] were subsequently found to be too large by factors of 3-5 [46, 66]. Indeed, beyond the Local Group, few if any galaxies show evidence for a central increase in the rms *stellar* velocities on the relevant spatial scales (see Figure 2.5 in [21]). Brightest cluster galaxies (BCGs), which are strongly represented among galaxies with "overmassive" SBHs [54], are particularly difficult cases due to their low central densities, so that stellar velocity dispersions measured near the projected center are strongly weighted by stars that are far from the SBH. For instance, M87, the BCG in the Virgo Cluster, exhibits no prima facie evidence for a central SBH in the stellar velocities, and the value of $M_{\rm BH}$ derived from the stellar data in M87 depends critically on what mass-to-light ratio is assigned to the stars [13, 62]; furthermore the value of $M_{\rm BH}$ derived from the stellar data is a factor ~ 2 greater than the value derived from the gaseous rotation curve [55]. (The latter *does* exhibit a clear Keplerian rise and the value of $M_{\rm BH}$ derived from it is much more robust, having remained essentially constant since its first determination in 1997 [63].) In effect, what is being measured in such galaxies is not the SBH mass, but rather the mass of the SBH plus the mass of the stars within some region the size of which is comparable to the resolution limit set by the telescope and which may be much larger than the SBH influence radius. Disentangling the two contributions can be extremely difficult [42]. Claims that the SBH influence radius has been "resolved" in such galaxies are always suspect, since they are based – not on an observed Keplerian velocity rise – but rather on the *assumption* that the influence radius is given by $\sim GM_{\rm BH,est}/\sigma^2$ with $M_{\rm BH,est}$ the *estimated* SBH mass. It is axiomatic that a "best-fit" value of $M_{\rm BH}$ will have an influence radius larger than the instrumental resolution, whether

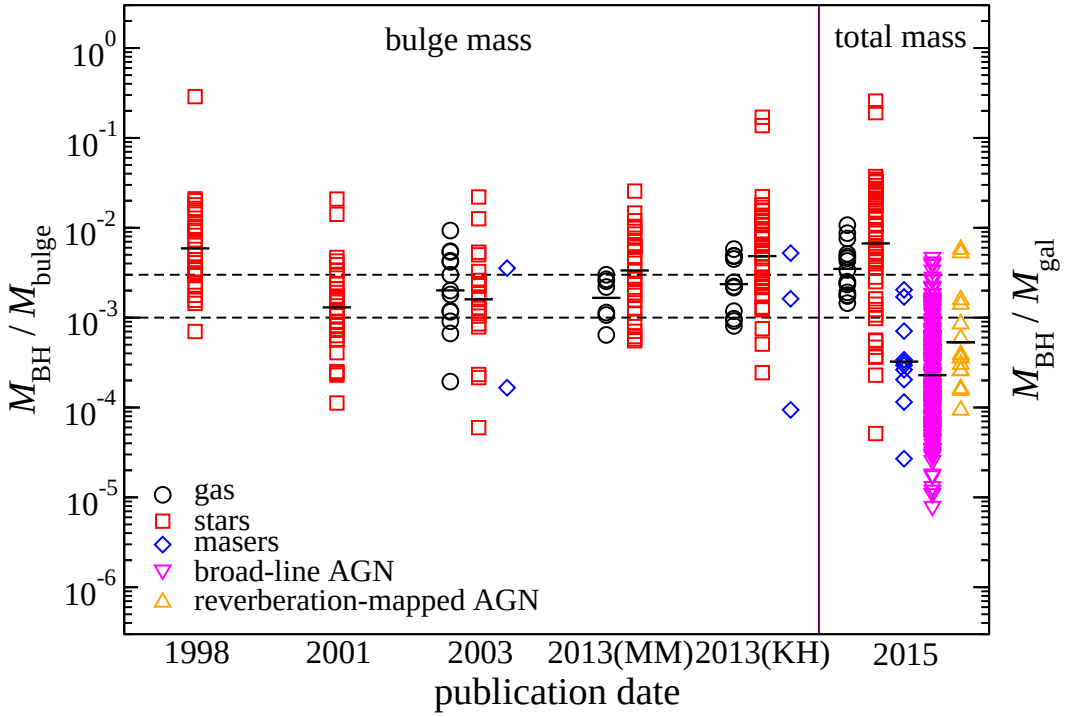

FIG. 2: Published estimates of the ratio $M_{\rm BH}/M_{\rm galaxy}$ (or $M_{\rm BH}/M_{\rm bulge}$) ordered by publication date and SBH mass measurement method used. Every point corresponds to a single galaxy; mean values are indicated with horizontal ticks. Horizontal dashed lines mark the values 0.003 (the currently accepted value) and 0.001 (the more conservative estimate considered in this paper). References: Magorrian et al. (1998) [49], Merritt & Ferrarese (2001) [50], Marconi & Hunt (2003) [64], McConnell & Ma (2013) [10], Kormendy & Ho(2013) [47], Reines & Volonteri (2015) [51].

or not the data from which $M_{\rm BH,est}$ was derived contain any useful information about the presence of a central mass concentration.

Figure 2 presents a compilation from the literature of estimates of $M_{\rm BH}/M_{\rm bulge}$. In line with the discussion in the previous paragraph, we make the following observations. i) Estimates of the mean $M_{\rm BH}/M_{\rm bulge}$ reached a minimum near 2000, following the winnowing of the stellar-based $M_{\rm BH}$ values by Ferrarese & Merritt [5]. Around this time, estimates based on stellar and gas data were consistent. ii) Estimates of $M_{\rm BH}/M_{\rm bulge}$ made since that time have crept back upward, particularly in the case of the stellar-based masses, and particularly in the most-massive galaxies (BCGs). iii) In any given study, the ordering of $\langle M_{\rm BH}/M_{\rm bulge}\rangle$ typically obeys stars > gas > AGN. This is consistent with the fact that the stellar data rarely exhibit a Keplerian rise, hence the $M_{\rm BH}$ values are likely to be biased upward, as discussed above. Estimates of $M_{\rm BH}$ in AGN, at the other extreme, use measured velocities of broad-emission-line gas that lies well inside the SBH influence sphere, hence is guaranteed to be responding almost entirely to the gravitational force of the SBH.

So far we have emphasized uncertainties in $M_{\rm BH}$. Table I points out a completely independent source of worry. Estimates of $M_{\rm bulge}$ in a given galaxy can exhibit wide variation from author to author. We are unable to give a reason for this, except to note that different authors base their estimates on luminosities measured in different passbands, carry out the bulge-disk decompositions differently, and make different assumptions about the stellar IMF and/or the mass-to-light ratio. (Marconi & Hunt [64] estimate bulge masses using the virial theorem, not measured luminosities.)

Based on these arguments, we adopt $\beta \approx 0.001$ (the value in 2001) as our preferred estimate of this ratio. We note that such a value is lower than what previous authors have assumed when estimating the stochastic GW background and thus implies a lower PTA signal than in the earlier studies (all else being equal). But given the sources of uncertainty discussed above, we present results for other (higher) values of β as well in what follows.

As for the fraction of mass of the galaxy contained in the bulge $f_{\rm bulge}$, we adopt the prescription of Simon & Burke-Spolaor [39]: for quiescent (elliptical) galaxies $f_{\rm bulge} = 0.9$ for $M_{\rm gal} > 10^{11}\mathcal{M}_\odot$, declining log-linearly to $f_{\rm bulge} = 0.25$ at $M_{\rm gal} = 10^{10}\mathcal{M}_\odot$, and for all star-forming (spiral) galaxies $f_{\rm bulge} = 0.25$. We do not allow for any scatter in these

TABLE I: Bulge mass estimates (Solar masses)

Reference	M87	NGC4459	NGC3377
Marconi & Hunt [64, 2003]	6.2×10^{11}	3.6×10^{11}	7.8×10^{10}
McConnell & Ma [10, 2013]	1.3×10^{12}	—	2.4×10^{10}
Scott et al. [65, 2013]	2.3×10^{11}	2.0×10^{10}	2.0×10^{10}
Kormendy & Ho [47, 2013]	5.3×10^{11}	7.6×10^{10}	3.2×10^{10}
Reines & Volonteri [51, 2015]	2.4×10^{11}	3.6×10^{10}	1.4×10^{10}
Savorgnan et al. [29, 2016]	2.6×10^{11}	2.9×10^{10}	4.0×10^{10}

parameters which makes our predictions for $h_c(f)$ somewhat lower than those of [32] or [39].

Given such large discrepancies between different aluthors and different methods, we consider most of the observed scatter in the $M-\sigma$ relation to be caused by measurement errors and ignore any possible intrinsic scatter. That makes our predictions for $h_c(f)$ somewhat lower than those of [32] or [39]. Figure 7 of [39] shows exactly how much we would underestimate the GW amplitude given intrinsic scatter; for example, a value of 0.3–0.4 dex reported by [10] and [47] implies a factor of ~ 1.5 difference in amplitude compared to zero scatter.

D. Nuclear rotation

In the simple galaxy models adopted here, rotation is implemented by supposing that some fraction of the stars on any given orbit have had the direction of their orbital angular momentum flipped compared with a nonrotating (isotropic) model. The parameter η is defined as the fraction of stars having a positive angular momentum component along the assumed axis of rotation; $\eta = 1/2$ corresponds to a nonrotating galaxy, $\eta = 1$ to a maximally-rotating one.

Sesana et al. [34] present a compilation from the literature of values of V/σ: the ratio of mean (streaming) velocity to velocity dispersion. They find that the following functions (normalized here to unit total number) are good representations of the observed distribution of $x \equiv V/\sigma$ for elliptical galaxies and for the bulges of spiral galaxies respectively:

$$\text{Ellipticals}: \quad N(x)dx = 4.27\,\exp\left(-3.98\,x^{0.96}\right)dx \tag{34a}$$

$$\text{Bulges}: \quad N(x)dx = 2.18\,x^{0.24}\exp\left\{-0.5\left[(x-0.47)/0.24\right]^2\right\}dx. \tag{34b}$$

Both functions are effectively zero for $V/\sigma > 1$.

In order to map η onto the observed V/σ, we investigated the observable properties of our models. Monte-Carlo representations were constructed and the projected, line-of-sight mean velocity and velocity dispersion were computed along the equatorial plane. Figure 3 (left) shows $\langle V\rangle/\sigma$ computed at two projected radii: $R = r_{\rm infl}$, the SBH influence radius; and $R = R_{\rm eff}$, the effective (projected half-mass) radius. The latter quantity, which is most directly comparable to the V/σ values tabulated by Sesana et al. [34], is well described by

$$\left|\frac{\langle V\rangle}{\sigma}\right|_{R=R_{\rm eff}} \approx 0.9\,(2\eta-1)^{1.25}\,. \tag{35}$$

We used Eq. (35) to express the relations (34) in terms of η, so that the η part of the merger rate distribution function (Eqs. 31) is

$$\mathcal{F}_1(\eta) = N(x)\frac{dx}{d\eta}, \qquad x = 0.9\,(2\eta-1)^{1.25} \in [0, 0.9] \tag{36}$$

This function is shown in Fig. 3 (right).

E. Dynamical evolution of the binary

Interaction of a massive binary with stars in a galactic nucleus causes changes in the binary's semimajor axis a and eccentricity e, as well as its orbital plane; the latter is characterized by the angle θ between the binary's angular momentum vector and the rotation axis of the nuclear cluster (the latter assumed fixed). The only stars we take into account are the ones initially unbound to the binary but with a small enough pericenter distance (of the order of the

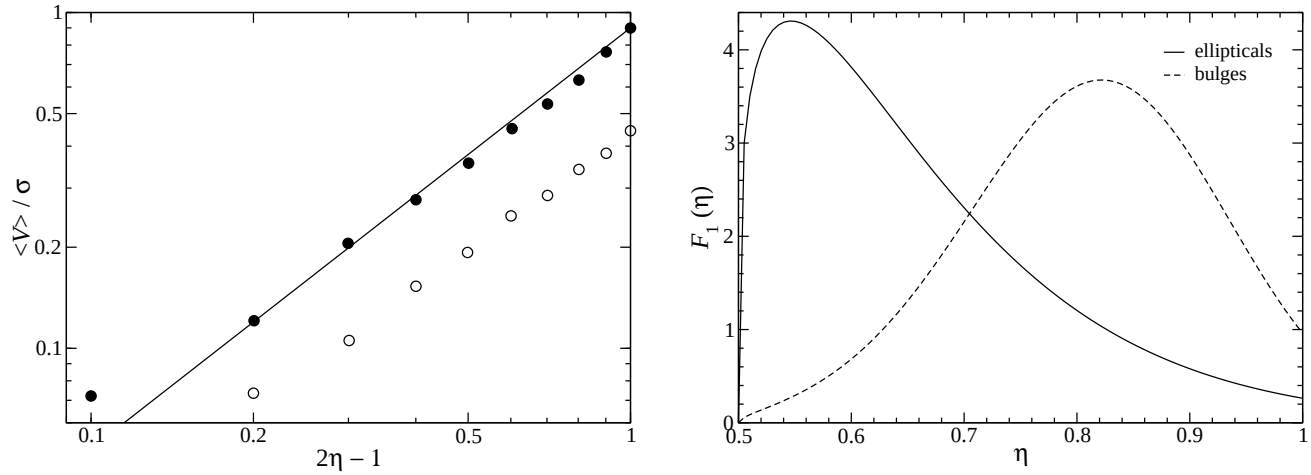

FIG. 3: Left: rotational properties of the galaxy models as a function of the parameter η. Filled and open circles show values measured at the galaxy half-mass radius and the SBH influence radius respectively, the line is Eq. (35). The nonrotating model from which the rotating models were generated by orbit-flipping was described by Dehnen's [4] density law with $\gamma = 1$ and with an assumed SBH mass of $0.002 M_{\rm gal}$. Right: two distributions of η used in this paper (see Eq.36 and 34).

binary semimajor axis a) that they can experience a close interaction with it. The equations describing the coupled changes in (a, e, θ) can be written [26]

$$\frac{da}{dt} = \left(\frac{da}{dt}\right)_\star + \left(\frac{da}{dt}\right)_{\rm GR} = -a^2 S - \frac{64}{5}\frac{\mathcal{Q} G^3 M_{12}^3}{c^5 a^3}\, F(e), \tag{37a}$$

$$\frac{de}{dt} = \left(\frac{de}{dt}\right)_\star + \left(\frac{de}{dt}\right)_{\rm GR} = aKS - \frac{304}{15}\frac{\mathcal{Q} G^3 M_{12}^3}{c^5 a^4}\, G(e), \tag{37b}$$

$$\frac{d\theta}{dt} = \left(\frac{d\theta}{dt}\right)_\star = aDS, \tag{37c}$$

$$F(e) = \frac{1 + (73/24)e^2 + (37/96)e^4}{(1-e^2)^{7/2}}, \tag{37d}$$

$$G(e) = e\frac{1 + (121/304)e^2}{(1-e^2)^{5/2}} \tag{37e}$$

where $S \equiv (d/dt)(1/a)$ is the binary hardening rate defined above, and K and D are dimensionless rate coefficients.[3] All three rate equations may depend on a, e and θ as well as on η and q. In deriving the rate coefficients, averages were taken over the binary argument of periapsis, ω, whose evolution is ignored. The binary's nodal angle Ω can be expected to evolve in a deterministic way; however we ignore that evolution here since it does not affect any other orbital element, and since Ω itself seems to be of little practical importance.

In terms of the hard-binary separation a_h (Eq. 5) and the initial hardening timescale (Eq. 21), we can write Eqs. (37) as

$$\frac{d(a/a_h)}{d(t/t_h)} = -\left(\frac{a}{a_h}\right)^2\left(\frac{S}{S_h}\right) - \left(\frac{a_{\rm GR,0}}{a_h}\right)^5\left(\frac{a}{a_h}\right)^{-3} F(e), \tag{38a}$$

$$\frac{de}{d(t/t_h)} = \left(\frac{a}{a_h}\right)\left(\frac{S}{S_h}\right)K - \frac{19}{12}\left(\frac{a_{\rm GR,0}}{a_h}\right)^5\left(\frac{a}{a_h}\right)^{-4} G(e), \tag{38b}$$

$$\frac{d\theta}{d(t/t_h)} = \left(\frac{a}{a_h}\right)\left(\frac{S}{S_h}\right)D, \tag{38c}$$

[3] Expressed in terms of quantities defined in [26], $D = D_{\theta,1}/H$.

$$a_{\rm GR,0} \equiv \left(\frac{64}{5}\frac{\mathcal{Q}G^3M_{12}^3}{c^5 S_h}\right)^{1/5}. \tag{38d}$$

The physical meaning of $a_{\rm GR,0}$ is the binary separation at which the hardening rate due to GW emission equals to that due to stellar interactions for a circular-orbit binary assuming $S = S_h$ (cf. Fig. 5).

In the case of an infinite homogenous distribution of stars with density ρ and velocity dispersion σ,

$$S_h = H\frac{G\rho}{\sigma}, \tag{39}$$

where $H \approx 15$ has been determined by scattering experiments [25]. In a real (inhomogeneous) galaxy, ρ and σ in Eq. (39) should be the density and velocity dispersion at (approximately) the influence radius of the binary, $r_{\rm infl}$ (as shown in [26]). Eqs. (38d) and (39) combine to give

$$\frac{a_{\rm GR,0}}{a_h} \approx 4.9\times10^{-3}\mathcal{Q}^{-4/5}\left(\frac{M_{12}}{10^8\mathcal{M}_\odot}\right)^{1/25}\left(\frac{\rho}{10^3\mathcal{M}_\odot{\rm pc}^{-3}}\right)^{-1/5}. \tag{40}$$

Here we have used the $M-\sigma$ relation [67]:

$$\frac{M_{12}}{10^8\mathcal{M}_\odot} \approx 1.66\left(\frac{\sigma}{200\,{\rm km\,s^{-1}}}\right)^5. \tag{41}$$

Since $\rho \sim M_{12}/r_{\rm infl}^3$ and $\sigma^2 \sim GM_{12}/r_{\rm infl}$,

$$S_h = S_{\rm infl} = b\sqrt{\frac{GM_{12}}{r_{\rm infl}^5}}, \tag{42}$$

where $b = 3\ldots5$ depending on galaxy structure [44]; in what follows we set $b = 4$. We can therefore write another expression for $a_{\rm GR,0}$:

$$\frac{a_{\rm GR,0}}{a_h} \approx 2.6\times10^{-3}\mathcal{Q}^{-4/5}\left(\frac{M_{12}}{10^8\mathcal{M}_\odot}\right)^{-1/10}\left(\frac{r_{\rm infl}}{10\,{\rm pc}}\right)^{1/2}. \tag{43}$$

To further simplify this expression, we can assume $r_{\rm infl} = GM_{12}/\sigma^2$ with σ related to M_{12} through Eq. (41), which yields

$$r_{\rm infl} \approx 13.2\,{\rm pc}\left(\frac{M_{12}}{10^8\mathcal{M}_\odot}\right)^{3/5}, \tag{44}$$

and substituting (44) into (43),

$$\frac{a_{\rm GR,0}}{a_h} \approx 3.0\times10^{-3}\mathcal{Q}^{-4/5}\left(\frac{M_{12}}{10^8\mathcal{M}_\odot}\right)^{1/5}. \tag{45}$$

Henceforth we define $a_{\rm GR,0}$ via Eq. (45).

Since a decreases monotonically with time, we can adopt $x \equiv a_h/a$ as a new time variable (x increases with time), starting our simulations at $x = 1$. The evolution equations for e and θ become

$$\frac{de}{dx} = \left[K - \frac{19}{12}\left(\frac{S_h}{S}\right)\left(\frac{a_{\rm GR,0}}{a_h}\right)^5 x^5 G(e)\right]\left[x + \left(\frac{S_h}{S}\right)\left(\frac{a_{\rm GR,0}}{a_h}\right)^5 x^6 F(e)\right]^{-1}, \tag{46a}$$

$$\frac{d\theta}{dx} = D\left[x + \left(\frac{S_h}{S}\right)\left(\frac{a_{\rm GR,0}}{a_h}\right)^5 x^6 F(e)\right]^{-1}. \tag{46b}$$

We adopt the following expressions from [26] for K and D:

$$K = 1.5\,e\,(1-e^2)^{0.7}\,[0.15 - (2\eta-1)\cos\theta], \tag{47a}$$

$$D = -0.3\,(2\eta - 1)\sqrt{\frac{1+e}{1-e}}\,\sin\theta. \tag{47b}$$

These expressions have been calculated assuming a hard binary ($a < a_h$), hence they do not depend on a or t. The hard-binary assumption is justified here because we choose $a = a_h$ as the initial separation.

In an irrotational nucleus ($\eta = 1/2$), the rate coefficients (47) become

$$K = 0.225\,e\,(1-e^2)^{0.7}, \qquad D = 0. \tag{48}$$

A few previous papers [22, 25, 31] have calculated K for a non-rotating nucleus; as shown in Fig. 4, our expressions are consistent with those of Sesana et al. [31] in the $a/a_h \to 0$ limit.

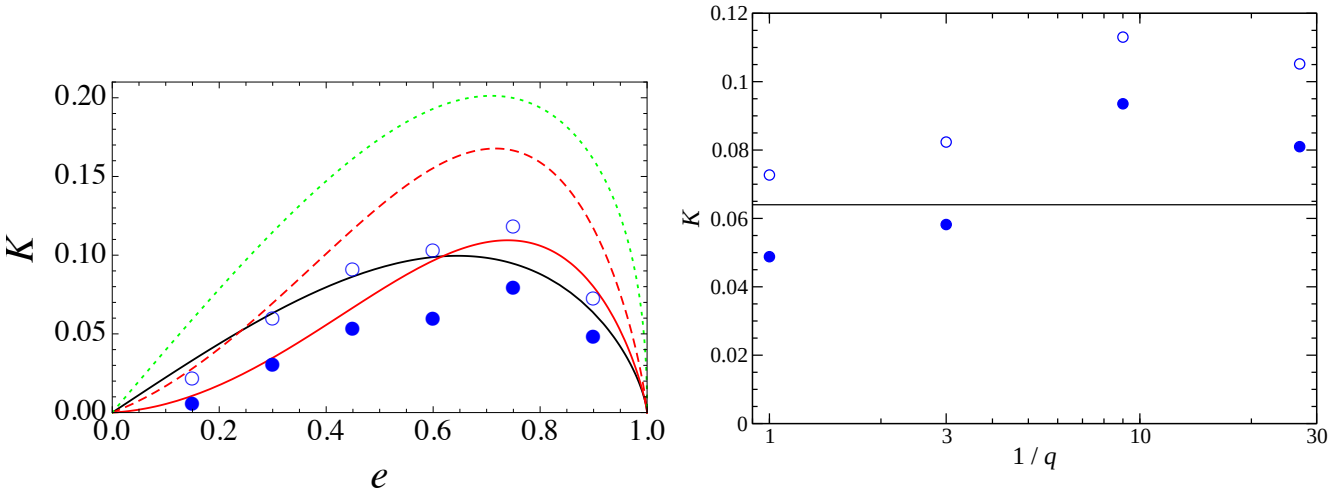

FIG. 4: Dimensionless eccentricity growth rate K (Eq. 47a) for equal-mass binaries and varying eccentricity (left) or $e = 0.9$ and varying mass ratio (right) in nonrotating nuclei. Black curve is our expression (48). Green dotted curve: the expression of Mikkola & Valtonen [22] in $a \to 0$ limit. Red curves: the results of Quinlan [25] for $a/a_h = 0.16$ (solid) and $a/a_h = 0.018$ (dashed). Blue circles: the results of Sesana et al. [31] for $a/a_h = 0.16$ (filled) and $a/a_h = 0.018$ (empty).

Fig. 5 presents solutions to the coupled Eqs. (46), for two different degrees of nuclear rotation, $\eta = 0.6$ (low rotation) and $\eta = 1$ (maximal rotation). The binary's orbital inclination always decreases, so that initially counterrotating binaries ($\theta \approx \pi$) tend to become corotating ($\theta \approx 0$). For corotating binaries, e almost always decreases with time: first due to stellar encounters, and later due to GW emission. In the case of counterrotating binaries, e generally increases at early times but eventually starts to decrease – either because the binary has become corotating, or because of GW emission. For binaries in maximally-rotating galaxies, reorientation of the orbital plane takes place quickly, and as a result, the binary enters the GW-dominated regime with low eccentricity. On the other hand, in slowly-rotating galaxies, both reorientation and circularization are much less pronounced, allowing the binary to enter the GW-dominated regime with high eccentricity.

By interacting with stars, a massive binary tends to decrease the number of stars on orbits that can interact with it ("loss-cone depletion"). This depletion is accounted for by letting S depend on a or t; the "full-loss-cone" approximation corresponds to $S = \text{const}$. Here we assume that loss-cone depletion has no effect on the mean energy or angular momentum carried away by a single stellar interaction; such an assumption is justified considering the chaotic nature of a binary-star interaction when the final velocity and orbital momentum of a star are weakly correlated with the initial ones. Then the change in hardening rate S is due only to the change in the rate of stellar interactions with the binary: $S \propto dn/dt$. Since all of the orbital parameters change at rates that are proportional to dn/dt, they are all proportional to S as well. That is why we can allow S to depend on a (or t) while keeping K and D time-independent.

The rate of loss-cone depletion depends strongly on the "geometry", i.e. the shape, of the galaxy. We adopt the following expressions for $S(a)$ from Vasiliev et al. [44]:

$$S(a) = kS_{\rm infl}\left(\frac{a}{a_h}\right)^{\alpha}, \tag{49a}$$

$$k = 0.4, \quad \alpha = 0.3 \ \text{ for triaxial nuclei}, \tag{49b}$$

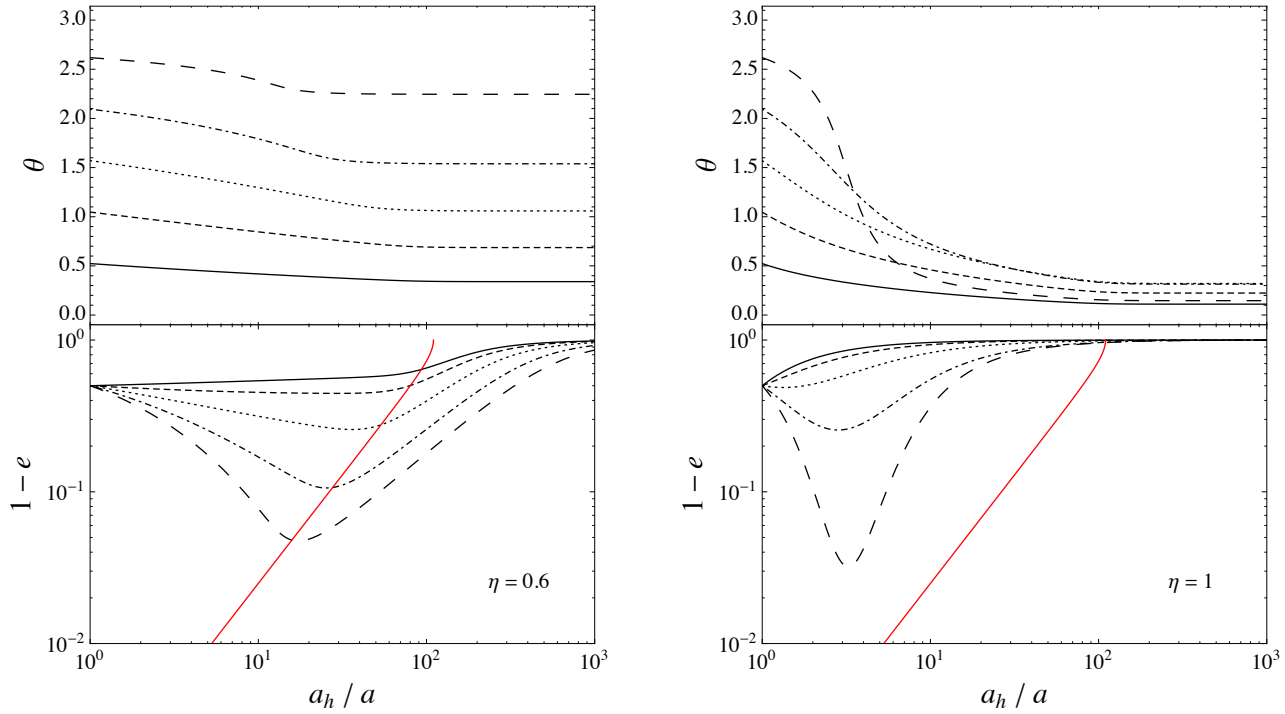

FIG. 5: Evolution of orbital inclination θ and eccentricity e for an equal-mass binary in a slowly-rotating nucleus ($\eta = 0.6$, left) and (b) a maximally-rotating nucleus ($\eta = 1$, right), computed using Eqs. (46) and (47) with $M_{12} = 10^8 \mathcal{M}_\odot$ and $S = S_h$. Different line styles correspond to different initial values of θ. The initial eccentricity is always $e_0 = 0.5$. The red curve separates the regimes where the hardening of the binary is dominated by stellar encounters (to the left) and GW emission (to the right); its equation is $a(e) = a_{\rm GR,0} F^{1/5}(e)$ (see Eq. 38a).

$$k = (N_\star/10^5)^{-1/2}, \quad \alpha = 0 \ \text{ for axisymmetric nuclei,} \tag{49c}$$
$$k = (N_\star/10^5)^{-1}, \quad \alpha = 0 \ \text{ for spherical nuclei} \tag{49d}$$

where $N_\star = M_{\rm gal}/M_\odot$ is the number of stars in the galaxy. The $N_\star$-dependence in Eqs. (49c,d) reflects the fact that in spherical and axisymmetric geometries, conservation of angular momentum (spherical symmetry) or its component along the symmetry axis (axisymmetry) fixes the minimum periapsis distance accessible to a star. Once all the stars on an orbit with given periapsis have been removed, continued supply of stars to the binary is only possible after new stars have been scattered onto the orbit by gravitational encounters, at rates that are $N_\star-$dependent. In triaxial galaxies, much of the phase space corresponds to orbits with no minimum periapsis; the time for a star on such an orbit to reach the binary depends much more on torques from the large-scale mass distribution than on two-body relaxation, hence the lack of an appreciable $N_\star$ dependence in the expression for the "triaxial" hardening rate. We anticipate the discussion in the next section by mentioning that mergers between luminous, gas-poor galaxies are expected to result in triaxial merger remnants, hence in efficient hardening of the binary.

Throughout this paper, we ignore the effect of torques from any ambient gas on the evolution of the binary SBH [73]. Gas is expected to be present, in dynamically significant densities, in low-mass systems (bulges of disk galaxies; dwarf elliptical galaxies), and more generally in galaxies at high redshift. One justification for our neglect of gas-dynamical torques is the recent realization, embodied here in Eqs. (49), that stellar-dynamical interactions can be much more effective than had previously been thought at evolving binary SBHs to separations $\ll a_h$. Nevertheless, there is a body of work, as summarized by [33], that argues that gaseous torques could shorten the time spent by a massive binary in the later stages of evolution, when GW emission competes with stellar-dynamical interactions.

III. RESULTS

We first consider the case in which all binary orbits are initially circular (and remain so). This assumption leaves only two important parameters: galaxy geometry (that is, the binary hardening law of Eq. 49) and β (the ratio of SBH mass to bulge mass). In Fig. 6a we plot $h_c(f)$ as predicted by our model after setting $\beta = 0.003$ and assuming "triaxial" (i.e. efficient) binary hardening. Also plotted there are the results of Sesana [33] and Ravi et al. [27], who made similar assumptions about β and binary hardening rates. Given the uncertainties quoted by those authors – 95% confidence intervals in h_c are said to be ± 0.5 dex [32] – we conclude that our model is consistent with both of them.

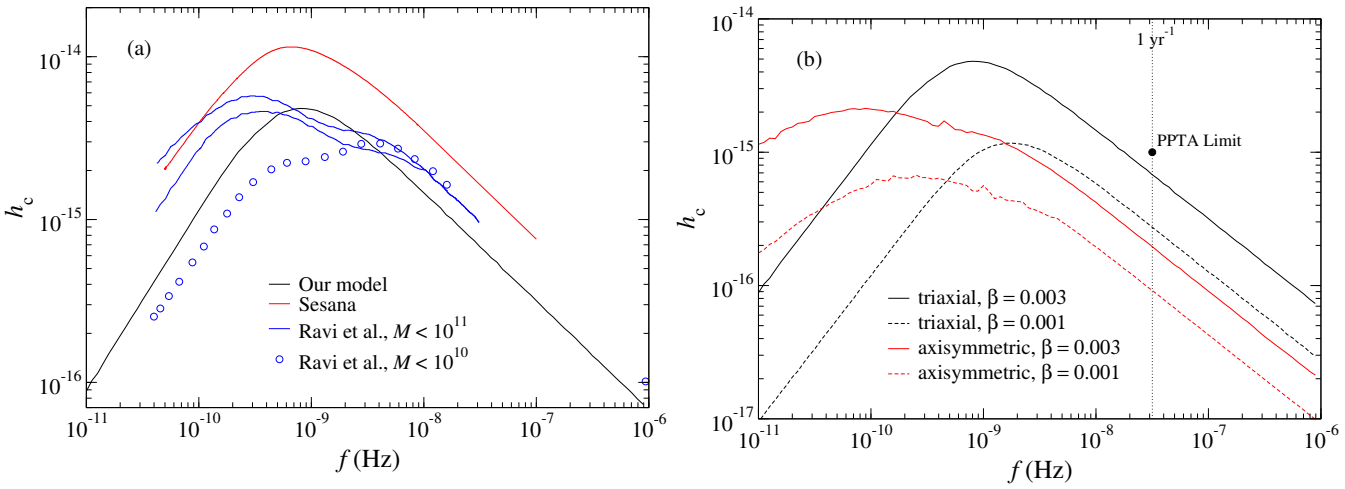

FIG. 6: (a) Predicted GW strain for circular binaries. Black: our model described in § II assuming the "triaxial" (efficient) hardening law and $\beta = 0.003$. Red: model from Sesana [33]. Blue: models from Ravi et al. [27]; solid lines correspond to $M_{12}/\mathcal{M}_\odot = 10^{6.5}...10^{11}$ and two different assumption about the stellar density profile; circles correspond to $M_{12}/\mathcal{M}_\odot = 10^{6.5}...10^{10}$. (b) Predicted GW strain for different assumptions about β (SBH mass) and binary hardening law. Black dot indicates the 95%-confidence upper limit from PPTA [37].

Fig. 6b illustrates the dependence of $h_c(f)$ on galaxy morphology (i.e. binary hardening law) and β. Decreasing the assumed SBH masses reduces the GW emission at all frequencies and shifts the peak of the spectrum to higher frequencies, since less-massive binaries enter the GW-dominated regime at higher orbital frequencies, i.e., smaller semimajor axes (Eq. 45). Changing the assumed galaxy morphology from triaxial to axisymmetric implies significant reduction in binary hardening rates (Eq. 49). There are two consequences. As shown in Fig. 7a-b, binaries in axisymmetric galaxies with $M_{12} \gtrsim 4\times 10^8 \mathcal{M}_\odot$ and any q have coalescence times that are longer than a Hubble time; at present they might not have reached the GW-dominated regime. This results in h_c being ~ 3 times lower at high frequencies compared with the "triaxial" case. At the same time, h_c in the "axisymmetric" case is higher at low frequencies because the binaries spend more time radiating at large orbital separations. In the case of spherical galaxies (not shown here), coalescence times are so long that there is essentially no GW emission at PTA-accessible frequencies.

The coalescence timescales in our nonrotating models are $6-7$ times shorter than those found by Vasiliev et al. [44]. This difference is a consequence of different definitions of the influence radius: we define it as GM/σ^2 while Vasiliev et al. use an empirical relation between the observed $r_{\rm infl}$ and black hole mass, obtained via the $M-\sigma$ relation (Eq. 9 in their paper), which implies larger values of $r_{\rm infl}$.

At sufficiently high frequencies, where binary dynamics are dominated by GW emission, the characteristic strain has the power-law dependence of Eq. (1):

$$h_c(f) = A_{\rm yr}\left(\frac{f}{1\,{\rm yr}^{-1}}\right)^{-2/3}.$$

Fig. 7b shows the dependence of $A_{\rm yr}$ on β assuming circular orbits ($e = 0$):

$$A_{\rm yr} \approx 2.7\times 10^{-16}\left(\frac{\beta}{10^{-3}}\right)^{0.85} \quad \text{for triaxial galaxies,} \tag{50a}$$

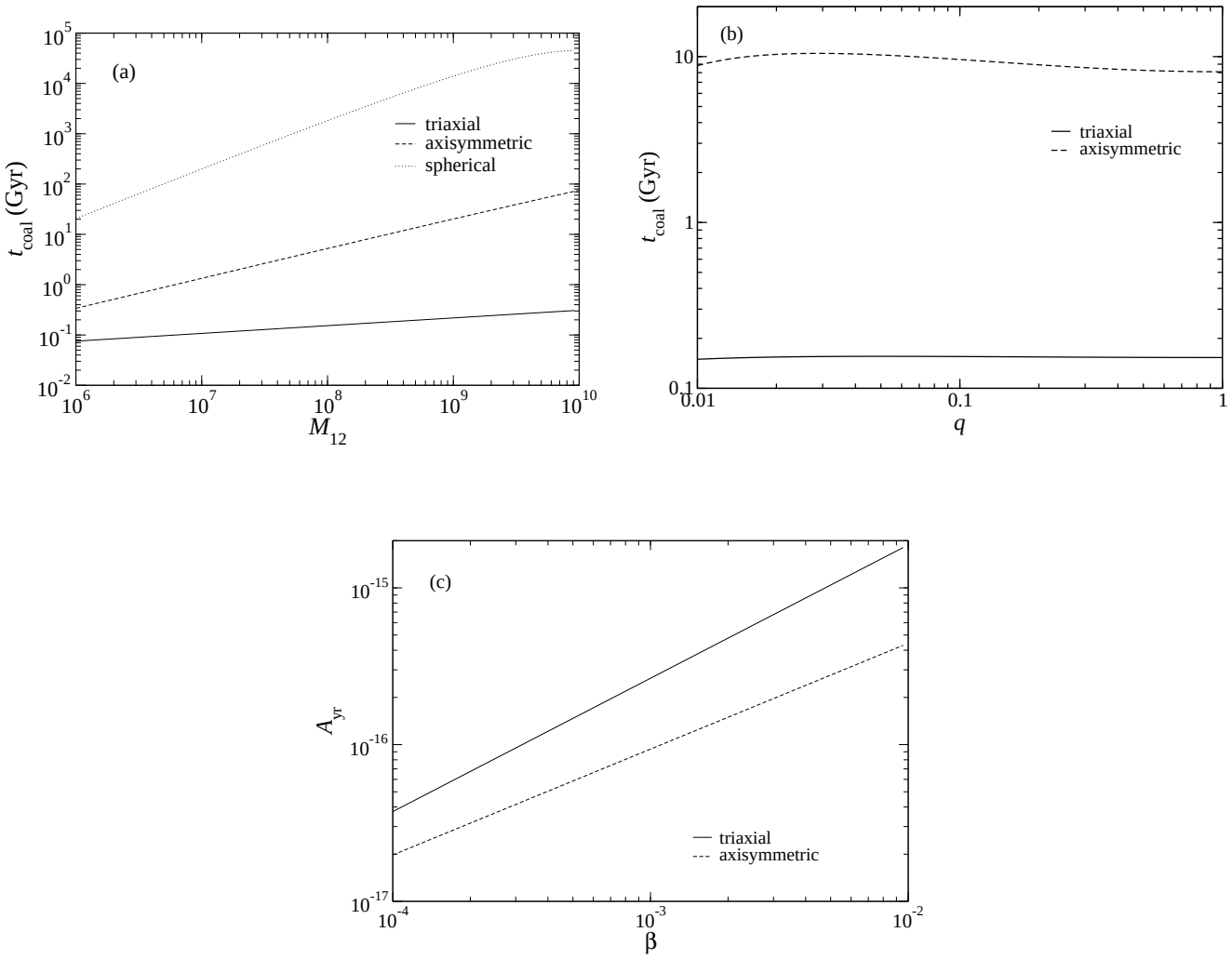

FIG. 7: (a) Coalescence time (from $a = a_h$ to $a \approx 0$) as a function of M_{12} for equal-mass circular binaries in triaxial, axisymmetric and spherical galaxies. (b) Coalescence time as a function of q for $M_{12} = 10^8 \mathcal{M}_\odot$. (c) Strain amplitude (Eq. 50) for circular-orbit binaries in triaxial and axisymmetric galaxies as a function of β (for spherical galaxies $A_{\rm yr} = 0$).

$$A_{\rm yr} \approx 9.3 \times 10^{-17} \left(\frac{\beta}{10^{-3}}\right)^{0.68} \quad \text{for axisymmetric galaxies.} \tag{50b}$$

(Note that the degree of nuclear rotation, η, is unimportant in the circular-orbit case.) Among all the possible parameter combinations, the choice "$\beta = 0.003$ + triaxial galaxies" should yield results most similar to those in the recent studies of Sesana [32], Ravi et al. [27], and Simon & Burke-Spolaor [39]. We indeed find that our estimate $A_{\rm yr} \approx 6.8 \times 10^{-16}$ is consistent, within 1σ, with those in the aforementioned papers.

We have found the following formula to be a good analytical approximation for the "$e = 0$ + triaxial galaxies" case (which is a reasonable assumption, as we'll show in the end of this section):

$$h_c(f) = A \frac{(f/f_{\rm yr})^{-2/3}}{1 + (f_{\rm b}/f)^{53/30}}, \tag{51a}$$

$$A = 2.77 \times 10^{-16} \left(\frac{\beta}{10^{-3}}\right)^{0.83} \sqrt{\frac{\dot{\mathcal{N}}_m}{\dot{\mathcal{N}}_{m,0}}}, \tag{51b}$$

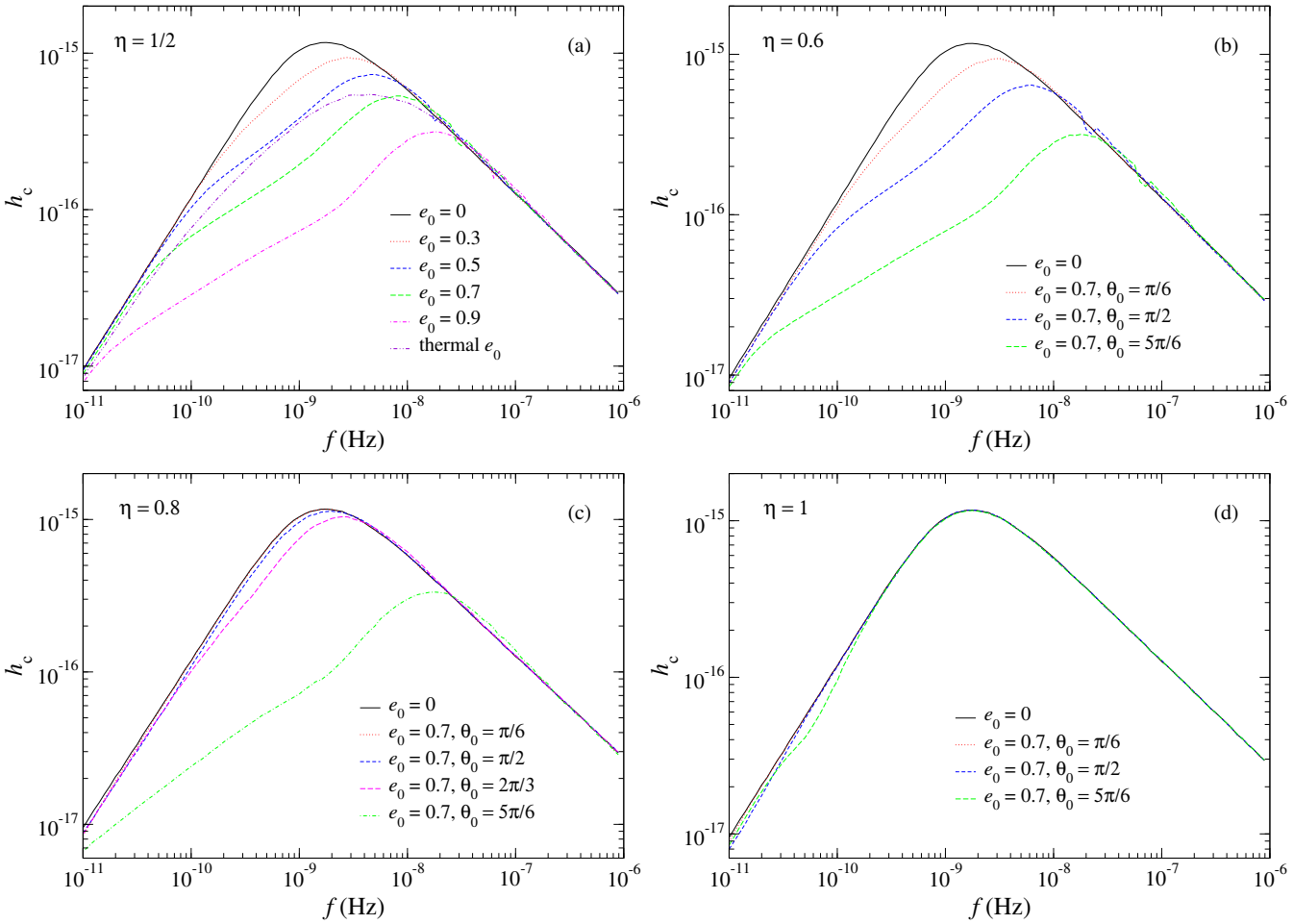

FIG. 8: Predicted GW strain for four different values of the nuclear corotation fraction η. All curves assume $\beta = 0.001$. Initial orbital elements (e_0, θ_0) are assumed to be the same for all binaries.

$$f_{\rm b} \;=\; 1.35\times10^{-9}\,{\rm Hz}\left(\frac{\beta}{10^{-3}}\right)^{-0.68}. \tag{51c}$$

Here we have accounted for the possibility of the actual galaxy merger rate $\dot{\mathcal{N}}_m$ being different than the one given by Eq. (32), $\dot{\mathcal{N}}_{m,0}$; see Section II B for the discussion of the possible reasons for that. We assume that $\dot{\mathcal{N}}_m$ always has the same dependence on galaxy mass and redshift and only the scale factor may vary, which leaves the shape of the spectrum unchanged and only changes its amplitude (Eq. 28).

This approximation is accurate to within 5% at PTA-sensitive frequencies $f > 10^{-9}\,$Hz. It's similar to the one suggested by Sampson et al. (Eq. 4) for single-mass binary population, but has a slightly different low-frequency slope (1.1 vs. 1, because of the difference in stellar hardening rate) and a broader peak (due to different BH masses contributing to the signal). Note that since $f_{\rm b} < f_{\rm yr}$ at all reasonable values of β,

$$A = A_{\rm yr}\left[1+\left(\frac{f_{\rm b}}{f_{\rm yr}}\right)^{53/30}\right] \approx A_{\rm yr} \tag{52}$$

It is also true that $f_{\rm b} < 0.1 f_{\rm yr}$ (minimum frequency currently probed by PTA) unless $\beta \lesssim 3\cdot10^{-4}$, which means it's unlikely that we will be able to see the break in stochastic GW background spectrum in the near future. However, this is true only in the assumption of zero eccentricity or quickly decreasing eccentricity, as we show below.

Fig. 8 illustrates the impact of nonzero orbital eccentricities on the GW spectrum. It confirms the previous results that high eccentricity reduces GW emission at low frequencies [7, 27, 30]. As shown in Fig. 8a, for nonrotating galaxies, the higher the assumed initial eccentricity, the stronger this effect is. At high frequencies ($f \gtrsim 10^{-8}\,$Hz)

the strain is unchanged because by the end of its dynamical evolution the binary orbit is always nearly circular due to GW emission. For rotating galaxies, the parameter θ_0, the initial inclination of the binary's orbit, comes into play. The closer θ_0 is to π (= counterrotation), the greater the maximum eccentricity reached by the binary (cf. Fig. 5). However, as Figs. 8b-d show, the influence of θ_0 is a strong function of a galaxy's degree of rotation. In maximally-rotating galaxies (Fig. 8d) the binary becomes corotating and eccentricities fall to negligible values before the binary starts to emit at PTA frequencies (Fig. 5, right). As a result, the spectrum is almost identical to that produced by circular binaries.

Fig. 10 shows computed spectra assuming a "thermal" distribution of initial eccentricities and different combinations of the θ_0- and η distributions discussed previously (Eqs. 31, 34 and 36). As expected, for triaxial galaxies (Fig. 10a), $h_c(f)$ is mildly (up to ~ 1.5 times) attenuated compared with circular binaries at peak frequencies ($f \sim 2\times10^{-10} \ldots 4\times 10^{-9}$ Hz) with almost no difference at other frequencies. For axisymmetric galaxies (Fig. 10b) the effect of eccentricity is different: high initial eccentricities decrease the coalescence time, allowing more binaries to reach the GW-dominated stage and thus contribute to the GW background at high frequencies (up to ~ 2 times increase in $h_c(f)$).

Fig. 9 shows in more detail the dependence of $A_{\rm yr}$ on (η, e_0, θ_0). Comparison with the plots of $t_{\rm coal}$ illustrates the fact that the increase of $A_{\rm yr}$ at high eccentricities is due to shorter coalescence timescales. In particular, $A_{\rm yr}$ for axisymmetric and triaxial galaxies become comparable when $t_{\rm coal}$ is comparable, which can happen for high ($e_0 \gtrsim 0.9$) initial eccentricities. Also, as the upper-right panel shows, high values of η usually imply smaller $A_{\rm yr}$ values unless θ_0 is high. $A_{\rm yr}$ for triaxial galaxies is not shown on these plots because it is practically independent of eccentricity.

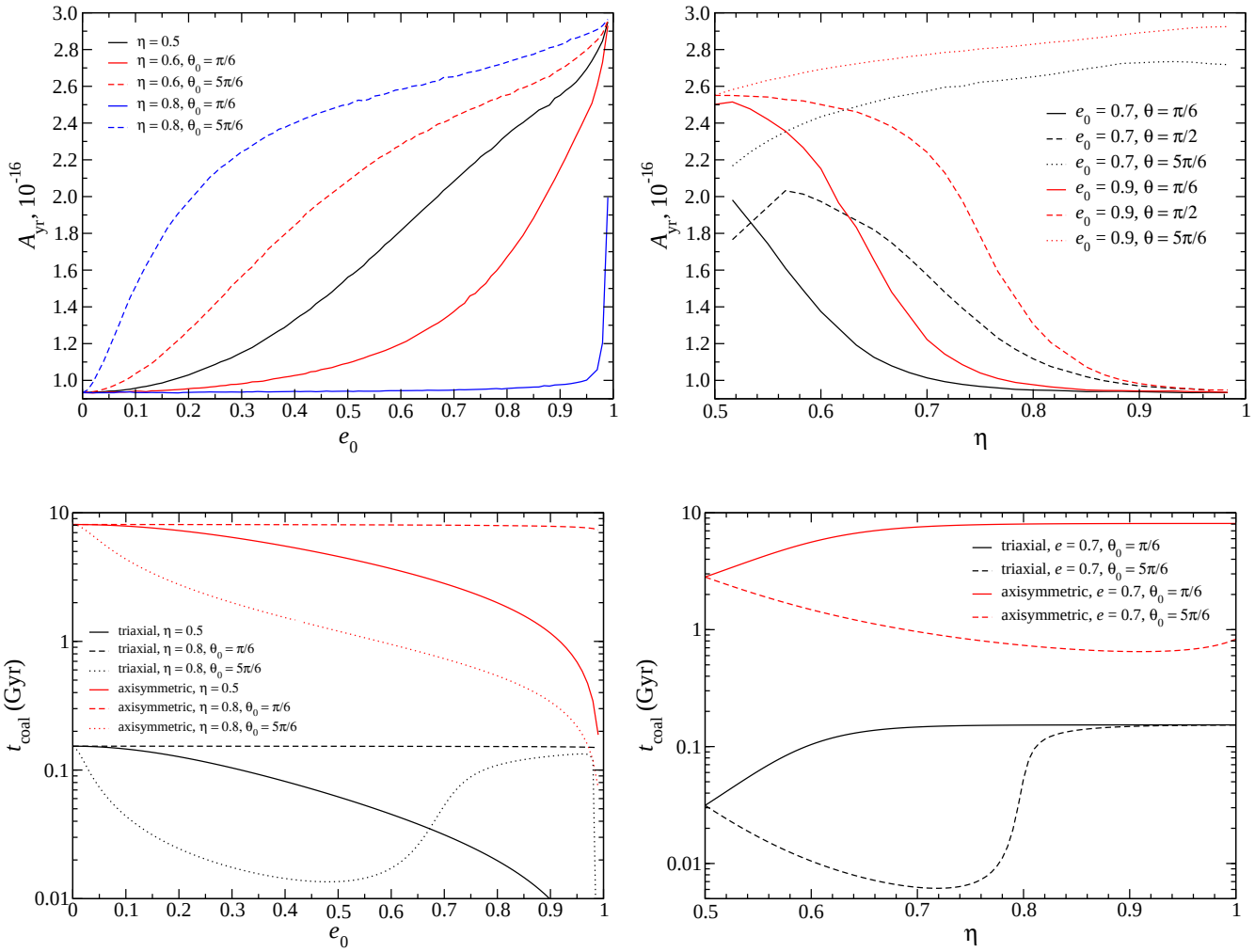

FIG. 9: The dependence of strain amplitude for axisymmetric galaxy model (up) and coalescence time for triaxial and axisymmetric galaxies (down) on the initial conditions.

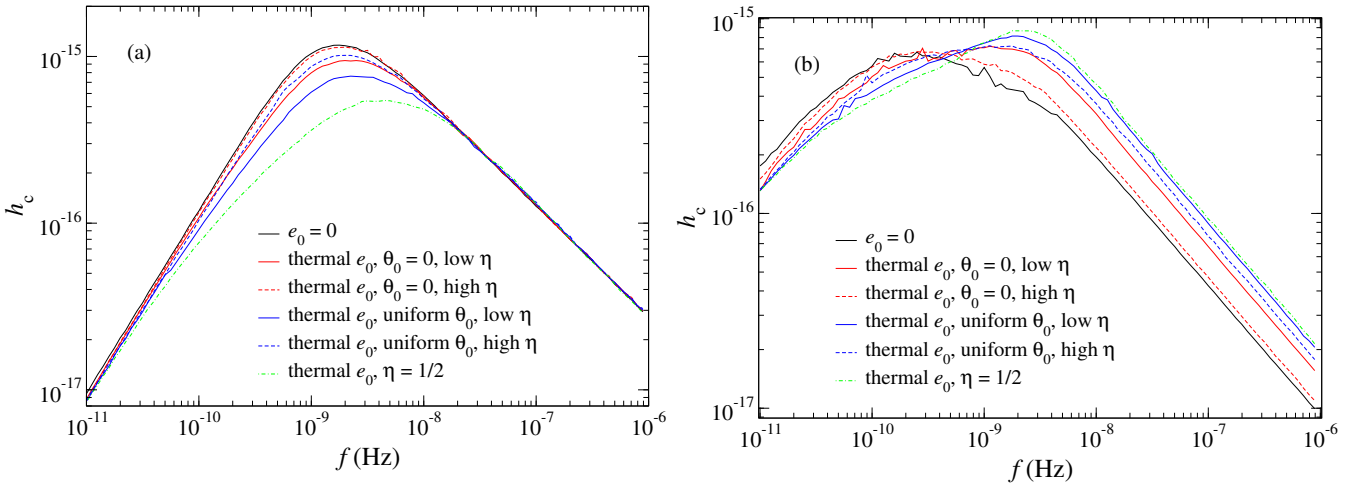

FIG. 10: Predicted GW strain assuming $\beta = 10^{-3}$, "thermal" distributions of initial eccentricity, two different distributions of θ_0 (Eq. 31) and two different distributions of η ("low η", Eq. (34a); "high η", Eq. (34b)) for (a) triaxial and (b) axisymmetric galaxies. Also shown for comparison is the curve computed assuming a thermal eccentricity distribution and no rotation ($\eta = 1/2$).

Finally, we construct a model which assumes that all spiral galaxy bulges are axisymmetric and fast-rotating, and that elliptical galaxies have different shapes depending on their mass: galaxies smaller than $10^{11.25}\mathcal{M}_\odot$ ("fast rotators") are axisymmetric while those heavier than $10^{11.25}\mathcal{M}_\odot$ ("slow rotators") are triaxial. This dichotomy is motivated observationally by the different morphologies and shape distributions of galaxies in the two mass ranges [12, 57]. For spirals and fast rotators we assume the η distribution from Eq. (34a), for slow rotators – the one from Eq. (34b).

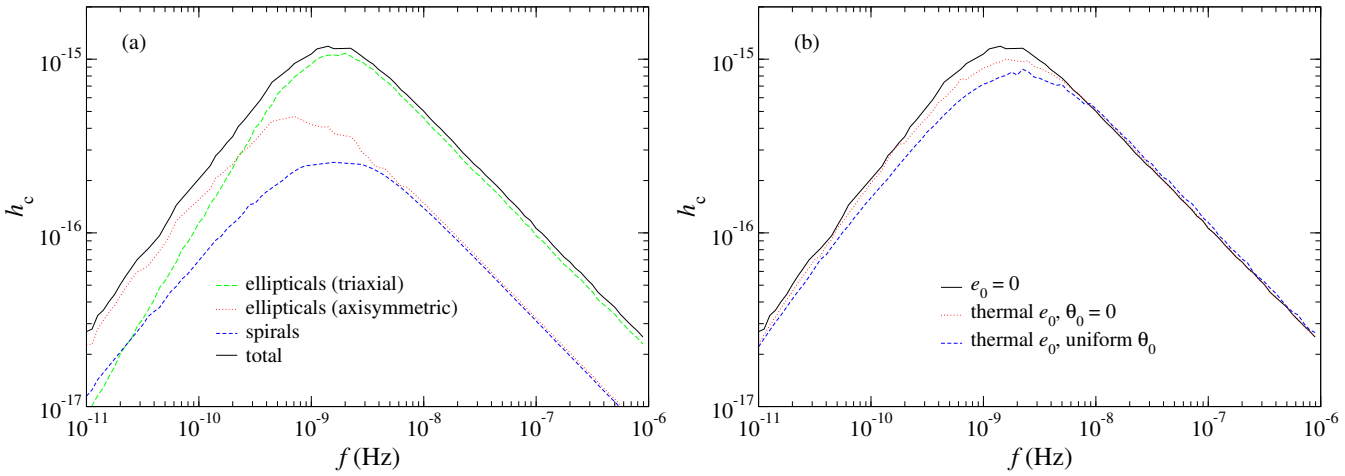

FIG. 11: (a) Contribution of different galaxy types to the predicted GW strain assuming $\beta = 10^{-3}$ and zero eccentricity for all binaries. (b) Strain spectra including the mixture of galaxy types described in the text, for different assumed distributions of e_0 and θ_0.

Fig. 11 shows the results. Fig. 11a plots the contributions from different galaxy types assuming $e = 0$. The signal at PTA frequencies is heavily dominated by triaxial (elliptical) galaxies – not surprising considering that they are the most massive galaxies. Because of that, the dependence of the total signal on the distributions of e_0 and θ_0 (Fig. 11b) is almost identical to that for triaxial galaxies (Fig. 10a), and its amplitude at high frequencies is practically independent of e_0, θ_0:

$$A_{\rm yr} \approx 2.31 \times 10^{-16} \left(\frac{\beta}{10^{-3}}\right)^{0.85} \tag{53}$$

Setting $\beta = 10^{-3}$, our preferred value, would reduce the GW strain amplitude by a factor ~ 3 compared to previous estimates, enough to account for the discrepancies between the models and the current PTA upper limits.

IV. DISCUSSION

We have presented calculations of the isotropic gravitational wave (GW) background spectrum that would be produced by a population of binary supermassive black holes (SBHs) in galactic nuclei. In our model, massive binaries evolve at large separations due to interaction with their stellar environment and at small separations due to emission of GWs. New features of our calculation, and the major results, are summarized here.

1. We model the time dependence of the binary hardening rate, $S = (d/dt)(1/a)$, using the results of Vasiliev et al. [43, 44] who derived expressions for $s(a)$ that are valid in the large-N (collisionless) limit appropriate to giant galaxies. These expressions imply efficient coalescence for binaries at the centers of triaxial (i.e. non-axisymmetric) galaxies, like those that are expected to form in galaxy mergers. In the case of axisymmetric geometries – which may be a better representation of low-luminosity galaxies – binary hardening rates are predicted to be lower (Fig. 7a), implying a dependence of coalescence timescale $t_{\rm coal}$ on galaxy luminosity, hence on $M_{\rm BH}$.

2. Rapid evolution of binary SBHs in triaxial galaxies significantly (by a few orders of magnitude) decreases GW emission at low frequencies, $f \lesssim 10^{-9}\,{\rm Hz}$, compared with a fiducial model in which evolution is driven entirely by GW emission itself. Evolution timescales are short enough in this geometry ($t_{\rm coal} \lesssim 300\,{\rm Myr}$) that essentially all binaries would reach coalescence, hence $h_c(f)$ at high frequencies includes contributions from essentially every binary that forms, and it obeys the standard $h_c \propto f^{-2/3}$ dependence for $f \gtrsim 3 \cdot 10^{-9}\,{\rm Hz} \approx 0.1\,{\rm yr}$, which is approximately the current PTA sensitivity range [1, 8, 37]). In axisymmetric galaxies, binary evolution at large separations is slower. The frequency below which the $h_c \sim f^{-2/3}$ spectrum is truncated is $f \lesssim 10^{-10}\,{\rm Hz}$ in this case; furthermore, since $t_{\rm coal}$ can be longer than a Hubble time for $M_{12} \gtrsim 4 \times 10^8 \mathcal{M}_\odot$ (the "final-parsec problem"), there is a reduction in the contribution of these binaries to $h_c(f)$ at high frequencies as well, lowering the predicted amplitude of $h_c(f)$ (however, the situation is different for highly eccentric binaries which have much lower $t_{\rm coal}$; see Fig. 9).

3. Galactic nuclei are generically rotational in the sense that there is a preferred axis about which stars orbit. Eccentricity evolution of a massive binary in a rotating nucleus depends strongly on its initial angular momentum direction compared with that of the nucleus [26]. Initially corotating binaries as well as some of the counterrotating ones circularize very quickly due to stellar encounters and enter the PTA band while almost circular. Counterrotating binaries can attain very high eccentricities ($e > 0.9$) if their angular momenta are initially strongly inclined with respect to the nucleus and in some cases will retain this high eccentricity even when entering the GW-dominated regime (Fig. 8). High eccentricities imply a reduction in $h_c(f)$ at low frequencies; in the case of axisymmetric galaxies it also increases $h_c(f)$ at high frequencies (by a factor as great as ~ 2) because it shortens the coalescence time, allowing more binaries to enter GW emission regime and contribute to the PTA signal.

4. We argue (§ II C) that previous calculations of $h_c(f)$ have been based on an incorrect (over-estimated) value of $M_{\rm BH}/M_{\rm bulge}$, the mean ratio of SBH mass to bulge mass. We adopt a fiducial value of 0.001 for this ratio, compared with ~ 0.003 in most other studies. This lower value results in a reduction in the predicted $h_c(f)$ at all frequencies (since $L_{\rm GW} \propto M_{12}^{10/3}$), and a shift in the peak of $h_c(f)$ to higher frequencies (lower-mass SBHs enter the GW-dominated regime at higher orbital frequencies). We show that in the frequency regime currently accessible to PTAs, $h_c(f) \propto (M_{12}/M_{\rm bulge})^{0.85}$ (Eq. 50 and Fig. 7b), so that our choice for this ratio implies a factor ~ 2 reduction in the characteristic strain. One consequence is that the existing "tension" between theoretical predictions of $h_c(f)$ and observational non-detections by PTAs [1, 37] is removed.

Shankar et al. [36] argued that a selection bias exists such that almost all galaxies in which the SBH influence sphere has been resolved have velocity dispersions that are higher than average for a fixed galaxy mass. The impact of this bias on detection of GWs by PTAs was analyzed in [35]. Both of these studies accept at face value claims that SBH influence radii have been resolved. We discuss, in § II C, why this assumption is likely to be incorrect.

For the values of $M_{\rm BH}/M_{\rm gal}$ that we favor, almost all of our models predict $A_{\rm yr} < 5 \times 10^{-16}$. This value of $A_{\rm yr}$ was identified as a "conservative *lower* limit" (emphasis added) by Siemens et al. [38] in their study of time-to-detection of the stochastic GW background by PTAs. Those authors considered two ways in which PTA detection limits might improve over time: due to lengthened data streams for individual pulsars, and due to the addition of new pulsars.

Assuming an average of three new pulsars per year (which they considered conservative), they estimated a probable date of first detection of GWs of $\sim$ 2021 for $A_{\rm yr} = 5.6 \times 10^{-16}$, and somewhat later if red noise is present in the auto-correlations.

In our assumed SBH mass – bulge mass relation, and our expressions for the bulge fractions in different galaxy types, we ignored scatter, assuming the relations to be exact. As Figure 9 of Simon & Burke-Spolaor [39] demonstrates, the presence of scatter in the SBH mass – bulge mass relation does not simply increase scatter in the computed $h_c(f)$ or $A_{\rm yr}$; it also increases their mean values. In this sense, our results could be viewed as a lower limit on $h_c(f)$.

Even if we accept the pessimistic view that detection of the isotropic GW background by PTAs may lie many years in the future, sufficiently massive or nearby systems may rise above the stochastic background signal and be individually detectable. Identification of the binary's host galaxy, and detection of electromagnetic radiative processes associated with the late evolution of the binary, can assist in the extraction of the binary parameters from PTA data. Photometric or spectroscopic measurement of the host galaxy's cosmological redshift would provide a distance estimate, allowing a chirp mass to be derived even for a binary SBH that does not evolve in frequency over PTA observing timescales [2]. These "multi-messenger" studies may hold the greatest hope for finally establishing the SBH mass-to-host-mass relations.

Acknowledgments

We thank J. Ellis, A. Graham, L. Lentati, J. Schnittman, A. Sesana, R. Shannon, E. Vasiliev and M. Zemcov for helpful comments and advice. This work was supported by the National Science Foundation under grant no. AST 1211602 and by the National Aeronautics and Space Administration under grant no. NNX13AG92G.

[1] Arzoumanian, Z., et al. 2016, ApJ, 821, 13
[2] Burke-Spolaor, S. 2013, Classical and Quantum Gravity, 30, 224013
[3] Cordes, J. M. 2013, Classical and Quantum Gravity, 30, 224002
[4] Dehnen, W. 1993, MNRAS, 265, 250
[5] Ferrarese, L., & Merritt, D. 2000, ApJL, 539, L9
[6] Foster, R. S., & Backer, D. C. 1990, Astrophys. J. , 361, 300
[7] Huerta, E. A., McWilliams, S. T., Gair, J. R., Taylor, S. R. 2015, Physical Review D, 92, 063010
[8] Lentati, L., Taylor, S. R., Mingarelli, C. M. F., et al. 2015, MNRAS, 453, 2576
[9] Milosavljević, M., & Merritt, D. 2003, The Astrophysics of Gravitational Wave Sources, 686, 201
[10] McConnell, N. J., & Ma, C.-P. 2013, Astrophys. J. , 764, 184
[11] Sampson, L., Cornish, N. J., & McWilliams, S. T. 2015, Phys. Rev. D , 91, 084055
[12] Emsellem, E. 2011, MNRAS, 414, 888
[13] Gebhardt, K., Adams, J., Richstone, D., et al. 2011, Astrophys. J. , 729, 119
[14] Ilbert, O., et al. 2013, A&A, 556, A55
[15] Kitzbichler, M. G., White, S. D. M. 2008, MNRAS, 391, 1489
[16] Lotz, J. M., Jonsson, P., Cox, T. J., Primack, J. R. 2008, MNRAS, 391, 1137
[17] Lotz, J. M., Jonsson, P., Cox, T. J., Primack, J. R. 2010, MNRAS, 404, 575
[18] Lotz, J. M., Jonsson, P., Cox, T. J., Primack, J. R. 2010, MNRAS, 404, 590
[19] Lotz, J. M. et al. 2011, Astrophys. J. , 742, 103
[20] McWilliams, Sean T., Ostriker, J. P., Pretorius, F. 2014, Astrophys. J. , 789, 156
[21] Merritt, D. 2013, Dynamics and Evolution of Galactic Nuclei. Princeton: Princeton University Press
[22] Mikkola, S., Valtonen, M. J. 1992, MNRAS, 259, 115
[23] Peters, P. C., Mathews, J. 1963, Phys. Rev. 131, 435
[24] Phinney, E. S. 2001, arXiv:astro-ph/0108028
[25] Quinlan, G. D. 1996, New Astronomy, 1, 35
[26] Rasskazov, A., Merritt, D. 2016, arXiv:1610.08555v2
[27] Ravi, V., Wyithe, J. S. B., Shannon, R. M., Hobbs, G., Manchester, R. N. 2014, MNRAS, 442, 56
[28] Robotham, A. S. G. et al. 2014, MNRAS, 444, 3986
[29] Savorgnan, G., Graham, A., Marconi, A., Sani, E. 2016, Astrophys. J. , 817, 21
[30] Schnittman, J. D., Krolik, J. H. 2015, Astrophys. J. , 806, 88
[31] Sesana, A., Haardt, F., Madau, P. 2006, Astrophys. J. , 651, 392
[32] Sesana, A. 2013, MNRAS, 433, L1
[33] Sesana, A. 2013, Class. Quantum Grav., 30, 224014
[34] Sesana, A., Barausse, E., Dotti, M., & Rossi, E. M. 2014, Astrophys. J. , 794, 104
[35] Sesana, A., Shankar, F., Bernardi, M., Sheth, R. K. 2016, arXiv:1603.09348

[36] Shankar, F. et al. 2016, arXiv:1603.01276
[37] Shannon, R. M. et al. 2015, Science, 349, 1522
[38] Siemens, X., Ellis, J., Jenet, F., & Romano, J. D. 2013, Classical and Quantum Gravity, 30, 224015
[39] Simon, J., Burke-Spolaor, S. 2016, arXiv:1603.06577
[40] Thanjavur, K., Simard, L., Bluck, A. F. L., Mendel, T. 2016, arXiv:1602.08674
[41] Tomczak, A. R. et al. 2014, Astrophys. J. , 783, 85
[42] Valluri, M., Merritt, D., & Emsellem, E. 2004, Astrophys. J. , 602, 66
[43] Vasiliev, E., Antonini, F., Merritt, D. 2014, Astrophys. J. , 785, 163
[44] Vasiliev, E., Antonini, F., Merritt, D. 2015, Astrophys. J. , 810, 49
[45] Xu, C. K. et al. 2012, Astrophys. J. , 747, 85
[46] Emsellem, E. 2013, MNRAS, 433, 1862
[47] Kormendy, J. & Ho, L. 2013, ARAA, 51, 511
[48] Lopez-Sanjuan, C. et al. 2012, A&A, 548, A7
[49] Magorrian, J. et al. 1998, AJ, 115, 2285
[50] Merritt, D. & Ferrarese, L. 2001, MNRAS, 320, L30
[51] Reines, A. E. & Volonteri, M. 2015, ApJ, 813, 82
[52] Richstone, D. O. et al. 1998, Nature, 385, A14
[53] van den Bosch, R. et al. 2012, Nature, 491, 729
[54] Volonteri, M. & Ciotti, L. 2013, ApJ, 768, 29
[55] Walsh, J. L. et al. 2013, ApJ, 770, 86
[56] Kormendy, J., Drory, N., Bender, R., & Cornell, M. E. 2010, Astrophys. J. , 723, 54
[57] Tremblay, B., & Merritt, D. 1996, AJ, 111, 2243
[58] McGaugh, S. S. 2015, Canadian Journal of Physics, 93, 250
[59] Naab, T., & Ostriker, J. P. 2009, Astrophys. J. , 690, 1452
[60] Weinzirl, T., Jogee, S., Khochfar, S., Burkert, A., & Kormendy, J. 2009, Astrophys. J. , 696, 411
[61] Toomre, A. 1977, Evolution of Galaxies and Stellar Populations, 401
[62] Gebhardt, K., & Thomas, J. 2009, Astrophys. J. , 700, 1690
[63] Macchetto, F., Marconi, A., Axon, D. J., et al. 1997, Astrophys. J. , 489, 579
[64] Marconi, A., & Hunt, L. K. 2003, ApJL, 589, L21
[65] Scott, N., Graham, A. W., & Schombert, J. 2013, Astrophys. J. , 768, 76
[66] Walsh, J. L., van den Bosch, R. C. E., Gebhardt, K., et al. 2016, Astrophys. J. , 817, 2
[67] Ferrarese, L., Ford, H. 2005, Space Science Reviews, 116, 523
[68] Huertas-Company, M. et al. 2013, MNRAS, 428, 1715
[69] Rawat, A. et al. 2008, Astrophys. J. , 681, 1089
[70] Dosopoulou, F., Antonini, F. 2016, arXiv:1611.06573
[71] Merritt, D. (2000), in Dynamics of Galaxies: from the Early Universe to the Present, Eds. Francoise Combes, Gary A. Mamon, and Vassilis Charmandaris. ASP Conference Series, Vol. 197, p. 221.
[72] Heggie, D. 1975, MNRAS, 173, 729
[73] Ivanov, P. B., Papaloizou, J. C. B., Polnarev, A. G. 1999, MNRAS, 307, 79