Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 11 March 2010 (MN LaTeX style file v2.2)

Gas driven massive black hole binaries: signatures in the nHz gravitational wave background

B. Kocsis[1,2,3,4]⋆ and A. Sesana[5,6]†
[1] *Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*
[2] *Institute for Advanced Study, Einstein Dr, Princeton, NJ, 08540, USA*
[3] *Eötvös University, Pázmány P. st. 1/A, Budapest H-1117, Hungary*
[4] *Einstein Fellow*
[5] *Albert Einstein Institute, Am Muhlenberg 1 D-14476 Golm, Germany*
[6] *Center for Gravitational Wave Physics, The Pennsylvania State University, University Park, PA 16802, USA*

Received —

ABSTRACT

Pulsar timing arrays (PTAs) measure nHz frequency gravitational waves (GWs) generated by orbiting massive black hole binaries (MBHBs) with periods between 0.1 – 10 yr. Previous studies on the nHz GW background assumed that the inspiral is purely driven by GWs. However, torques generated by a gaseous disk can shrink the binary much more efficiently than GW emission, reducing the number of binaries at these separations. We use simple disk models for the circumbinary gas and for the binary-disk interaction to follow the orbital decay of MBHBs through physically distinct regions of the disk, until GWs take over their evolution. We extract MBHB cosmological merger rates from the Millennium simulation, generate Monte Carlo realizations of a population of gas driven binaries, and calculate the corresponding GW amplitudes of the most luminous individual binaries and the stochastic GW background. For stationary α–disks with $\alpha > 0.1$ we find that the nHz GW background can be significantly modified. The number of resolvable binaries is however not changed by the presence of gas; we predict 1–10 individually resolvable sources to stand above the noise for a 1–50 ns timing precision. Gas driven migration reduces predominantly the number of small total mass or unequal mass ratio binaries, which leads to the attenuation of the mean stochastic GW–background, but increases the detection significance of individually resolvable binaries.

Key words: black hole physics, gravitational waves – cosmology: theory – pulsars: general

1 INTRODUCTION

Inspiraling massive black hole binaries (MBHBs) with masses in the range $\sim 10^4 - 10^{10}\,\mathrm{M}_\odot$ are expected to be the dominant source of gravitational waves (GWs) at $\sim$ nHz – mHz frequencies (Haehnelt 1994; Jaffe & Backer 2003; Wyithe & Loeb 2003; Sesana et al. 2004, 2005). The frequency band $\sim 10^{-5}\,\mathrm{Hz} - 1\,\mathrm{Hz}$ will be probed by the *Laser Interferometer Space Antenna* (*LISA*, Bender et al. 1998), a space-borne GW laser interferometer developed by ESA and NASA. The observational window $10^{-9}\,\mathrm{Hz} - 10^{-6}\,\mathrm{Hz}$, corresponding roughly to orbital periods 0.03 – 30 yr, is already accessible using Pulsar Timing Arrays (PTAs; e.g. the Parkes radio-telescope, Manchester 2008). The complete Parkes PTA (PPTA, Manchester 2008), the European Pulsar Timing Array (EPTA, Janssen et al. 2008), and NanoGrav (Jenet et al. 2009) are expected to improve considerably on the capabilities of these surveys, eventually joining their efforts in the international PTA project (IPTA, Hobbs et al. 2009); and the planned Square Kilometer Array (SKA, Lazio 2009) will produce a major leap in sensitivity.

Radio pulses generated by rotating neutron stars travel through the Galactic interstellar medium and are detected by radio telescopes on Earth. The arrival times of pulses are fitted for a model including all the known and measured systematic effects affecting the signal generation, propagation and detection (Edwards, Hobbs, & Manchester 2006). Timing residuals between the observed pulses and the best fit model, carry information on additional unmodelled effects, including the presence of GWs. Indeed, GWs modify the propagation of radio signals from the pulsar to the Earth (Sazhin 1978; Detweiler 1979; Bertotti et al. 1983;

⋆ E-mail: bkocsis@cfa.harvard.edu
† E-mail: alberto.sesana@aei.mpg.de

Hellings & Downs 1983; Jenet et al. 2005), and PTAs measure the direction dependent systematic variations in the arrival times of signals from a sample of nearly stationary pulsars in the Galaxy distributed over the sky.

PTAs provide a direct observational window onto the MBH binary population, and can contribute to address a number of open astrophysical questions, such as the shape of the bright end of the MBH mass function, the nature of the MBH-bulge relation at high masses, and the dynamical evolution at sub-parsec scales of the most massive binaries in the Universe (particularly relevant to the so-called "final parsec problem", Milosavljevic & Merritt 2003) PTAs can detect gravitational radiation of two forms: (i) the stochastic GW background produced by the incoherent superposition of radiation from the whole cosmic population of MBHBs and (ii) individual sources that are sufficiently bright in GWs to outshine the background (typically massive, $M \gtrsim 10^9\,\mathrm{M}_\odot$, and "cosmologically nearby", $d_L \lesssim 3\,\mathrm{Gpc}$). Both classes of signals are of great interest, and PTAs could lead to the discovery of systems difficult to detect with other techniques (for alternatives using active galactic nuclei, see Haiman, Kocsis, & Menou 2009, and references therein).

Popular scenarios of massive black hole (MBH) formation and evolution (e.g. Volonteri, Haardt, & Madau 2003; Koushiappas & Zentner 2006; Malbon et al. 2007; Yoo et al. 2007) predict frequent MBH mergers (up to several hundreds per year), implying the existence of a large number of sub-parsec MBHBs. The prospect for detecting GW signals using PTAs depends on the number and cosmological distribution of MBHBs with orbital periods of 0.03 – 30 yr, or separations typically in the range $0.001 - 0.1$ pc. The three main ingredients for calculating the GW background are

(i) the merger rate of MBHBs as a function of mass and redshift,
(ii) the relative time each binary spends at these separations during a merger episode and
(iii) the amplitude of the GW signal produced by each individual stationary system.

Recently Sesana, Vecchio, & Colacino (2008, SVC08 hereinafter) and Sesana, Vecchio, & Volonteri (2009, SVV09 hereinafter), carried out a detailed study of the expected signals (stochastic and individual), focusing on the uncertainties related to (i). They found that the background is affected by the galaxy merger rate evolution along the cosmic history, the massive black hole mass function, and the accretion history of the MBHB during a galaxy merger, and they predict a factor of ~ 10 uncertainty for the characteristic strain amplitude in the range $2 \times 10^{-15} - 2 \times 10^{-16}$, at $f = 1/\mathrm{yr}$, within the expected detection capabilities of the complete PPTA and of the SKA. They pointed out that the GW signal can be separated into individually resolvable sources and a stochastic background, and found the number of individually resolvable sources for a 1ns timing precision level to be between 5 to 15, depending on the considered model.

In this paper, we examine for the first time how predictions relevant for PTA observations are modified by the presence of ambient gas, affecting the inspiral rate of binaries during a merger episode, ingredient (ii) above. A gaseous envelope is expected to surround the binary because MBHBs are produced in galaxy mergers, which are known to trigger inflows of large quantities of gas into the central region, as shown by hydrodynamic simulations (Springel et al. 2005). This gas, accreted onto the MBHs, is responsible for luminous AGN activity, and is also expected to catalyze the coalescence of the new-formed MBH pair (e.g., Escala et al. 2004; Dotti et al. 2007), as described below. The forming MBHB spirals inward initially as a result of dynamical friction on dark matter, ambient stars, and gas (Begelman, Blandford, & Rees 1980). As the binary separation shrinks to sub-pc scales, the supply rate of stars crossing the orbit decreases, and the interaction with stars becomes less and less efficient to shrink the binary. In gas rich mergers, the dense nuclear gas is expected to cool rapidly and settle into a geometrically-thin circumbinary accretion disk (e.g., Barnes 2002; Escala et al. 2004). Torques from the tidal field of the binary clear a gap in the gas with a radius less than twice the separation of the binary, and generates a spiral density wave in the gaseous disk, which in turn drains angular momentum away from the binary on a relatively short timescale within the last parsec, $\lesssim 10^7$ yr (Escala et al. 2005; Armitage & Natarajan 2002, 2005; Dotti et al. 2007, see however Lodato et al. 2009). Ultimately, at even smaller separations, corresponding to an orbital timescale of $\sim$ years, the emission of GWs becomes the dominant mechanism driving the binary to the final coalescence. The main point of this paper, is to notice that the most sensitive PTA frequency band corresponds to orbital separations near the transition between gas and GW dominated evolution . *As binaries shrink more quickly inwards in the gas-driven phase, the number of binaries emitting at each given separation is decreased compared to the purely GW-driven case.* The subject of this work is to explore how various gas-driven models modify the expectations on the GW signal potentially observable by PTAs.

Recently, Haiman, Kocsis, & Menou (2009, HKM09 hereafter), examined the evolution of MBHBs in the gas–driven regime for simple models of geometrically thin circumbinary disks (see also, Syer & Clarke 1995; Armitage & Natarajan 2005). The interaction between the binary and the gaseous disk is analogous to type-II planetary migration, and evolves through two main phases. First, the inspiral is analogous to the disk–dominated type-II migration of planetary dynamics, where the binary migrates inwards with a radial velocity equal to that of the gas accreting towards the center. Later, as the mass of the gas within a few binary separations becomes less than the reduced mass of the binary, the evolution slows down, and it is analogous to the planet–dominated (or secondary–dominated) type-II migration. In both cases, the radial inspiral rate is still much faster than in the purely GW driven case at orbital separations beyond a few hundred Schwarzschild radii. For standard Shakura-Sunyaev α–disk models, the viscosity is assumed to be proportional to the total pressure, and is consequently very large in the radiation pressure dominated phase at small radii, increasing the migration rate in the radiation pressure dominated phase. On the other hand, for β–disk models, where the viscosity is proportional to the gas pressure only, the increase of radiation pressure does not impact the viscous timescale, and the migration rate is relatively slower in this regime. Finally, we note that the binary-gas interaction is also significantly different for non–

steady models of accretion (Ivanov, Papaloizou, & Polnarev 1999, HKM09). In the typical secondary–dominated phase, gas flows in more quickly then how the binary separation shrinks, and is repelled close to the outer edge of the gap by the torques of the binary. This causes gas to accumulate near the gap, and delays the merger of the binary relative to the steady-state models.

In this paper, we couple the HKM09 models for the migration of MBHBs in the presence of a steady gaseous disk, to the population models derived in SVV09, and we compute the effects on GWs at nHz frequencies. Here, we restrict to nearly circular inspirals for simplicity, using the corresponding GW spectrum (ingredient–iii above). This assumption might be violated in gas driven inspirals (Armitage & Natarajan 2005; Cuadra et al. 2009), and is the subject of a future paper (Sesana & Kocsis 2010, in prep).

The paper is organized as follows. In Section 2 we introduce the theory of the GW signal from a MBHB population, describing its characterization in terms of its *stochastic level* and of the statistics of *individually resolvable sources*. In Section 3 we describe our MBHB population model, coupling models of coalescing binaries derived by cosmological N-body simulations to a scheme for the dynamical evolution of the binaries in massive circumbinary disks. We present in detail our results in Section 4, and in Section 5 we briefly summarize our main findings. Throughout the paper we use geometric units with $G = c = 1$.

2 DESCRIPTION OF THE GRAVITATIONAL WAVE SIGNAL

The theory of the GW signal produced by the superposition of radiation from a large number of individual sources, was extensively presented in SVC08 and SVV09, here we review the basic concepts, deriving the GW signal for gas driven mergers.

The dimensionless characteristic amplitude of the GW background produced by a population of binaries with a range of masses m_1 and m_2 and redshifts z is given by (Phinney 2001)

$$h_c^2(f) = \frac{4}{\pi f^2}\int \mathrm{d}z\mathrm{d}m_1\mathrm{d}m_2\, \frac{\partial^3 n}{\partial z\partial m_1\partial m_2}\frac{1}{1+z}\,\frac{\mathrm{d}E_{\rm gw}}{\mathrm{d}\ln f_r}\,. \quad (1)$$

where $dE_{\rm gw}/d\ln f_r$ is the total emitted GW energy per logarithmic frequency interval in the comoving binary rest-frame, $f_r = (1+z)f$ is the rest-frame frequency, f is the observed frequency, n is the comoving number density of sources, and the $1/(1+z)$ factor accounts for the redshift of the observed GW energy. The characteristic GW amplitude h_c is related to the present day total energy in GWs as $\rho_{\rm GW} = \frac{\pi}{4}\int f h_c^2(f)\mathrm{d}f$.

2.1 GW-driven inspirals

To the leading quadrupole order, for circular purely GW-driven binaries orbiting far outside the innermost stable circular orbit (ISCO), Eq. (1) can be evaluated assuming that $E_{\rm gw} = E_{\rm pot} = -m_1 m_2/a$ is equal to the Newtonian potential energy of the binary, and that f_r is equal to twice the Keplerian orbital frequency (Phinney 2001). In this case

$$\frac{\mathrm{d}E_{\rm gw}}{\mathrm{d}\ln f_r} = \frac{1}{3}\mu(\pi M f_r)^{2/3} = \frac{1}{3}(\pi f_r)^{2/3}\mathcal{M}^{5/3} \quad (2)$$

for $f_r < f_{\rm ISCO} \equiv 1/(6^{3/2}\pi M)$. Here $\mu = m_1 m_2/M$, $M = m_1 + m_2$, and $\mathcal{M}^{5/3} = \mu M^{2/3}$ are the reduced, total, and chirp masses for a binary, and $f_{\rm ISCO}$ is the GW frequency at ISCO. Substituting into Eq. (1),

$$h_c^2(f) = \frac{4f^{-4/3}}{3\pi^{1/3}}\int \mathrm{d}z\mathrm{d}m_1\mathrm{d}m_2\, \frac{\partial^3 n}{\partial z\partial m_1\partial m_2}\frac{\mu M^{2/3}}{(1+z)^{1/3}}. \quad (3)$$

It is also useful to examine the number of MBHBs and their respective contributions to the total signal (Phinney 2001; SVC08). Eq. (3) can be rewritten as

$$h_c^2(f) = \int \mathrm{d}z\mathrm{d}m_1\mathrm{d}m_2\, \frac{\partial^4 N}{\partial m_1\partial m_2\partial z\partial\ln f_r}h^2(\mathcal{M},z,f_r), \quad (4)$$

where N is the number of sources, which we can calculate from a comoving merger rate density as explained below, and

$$h(\mathcal{M},z,f_r) = \frac{8}{10^{1/2}}\frac{\mathcal{M}}{d_L(z)}(\pi\mathcal{M}f_r)^{2/3}, \quad (5)$$

is the sky and polarization averaged characteristic GW strain amplitude of a single binary with chirp mass $\mathcal{M}$, at the particular orbital radius corresponding to f_r.

We generate the distribution $\partial^4 N/(\partial m_1\partial m_2\partial z\partial\ln f_r)$ corresponding to a comoving merger rate density[1] $\partial^4 N/(\partial m_1\partial m_2\partial t_r\partial V_c)$ (see Sec. 3.1), assuming that the number of binaries emitting in the interval $\ln f_r$ is proportional to the time the binary spends at that frequency,

$$\frac{\partial^4 N}{\partial m_1\partial m_2\partial z\partial\ln f_r} = \frac{\partial^4 N}{\partial m_1\partial m_2\partial t_r\partial V_c}\frac{\mathrm{d}V_c}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}t_r}\frac{\mathrm{d}t_r}{\mathrm{d}\ln f_r} \quad (6)$$

where $\mathrm{d}V_c/\mathrm{d}z$ and $\mathrm{d}z/\mathrm{d}t_r$ are given by the standard cosmological relations between comoving volume, redshift, and time, given in e.g. Phinney (2001). The last factor can be expressed using the residence time $t_{\rm res} = a(\mathrm{d}a/\mathrm{d}t_r)^{-1}$ the binary spends at a particular semimajor axis as

$$\left|\frac{\mathrm{d}t_r}{\mathrm{d}\ln f_r}\right| = \left|\frac{\mathrm{d}t_r}{\mathrm{d}\ln a}\frac{\mathrm{d}\ln a}{\mathrm{d}\ln f_r}\right| = \frac{2}{3}t_{\rm res}, \quad (7)$$

where Kepler's law, $a = M(\pi M f_r)^{-2/3}$, was used to obtain $\mathrm{d}\ln a/\mathrm{d}\ln f_r = 2/3$, and the residence time for a purely GW-driven evolution is

$$t_{\rm res} = t_{\rm res}^{\rm gw} \equiv \frac{\mathrm{d}t_r}{\mathrm{d}\ln a} = \frac{5}{64}\mathcal{M}(\pi\mathcal{M}f_r)^{-8/3}. \quad (8)$$

In summary, the distribution of sources in Eq. (4) becomes

$$\frac{\partial^4 N}{\partial m_1\partial m_2\partial z\partial\ln f_r} = \frac{2}{3}\frac{\partial^4 N}{\partial m_1\partial m_2\partial t_r\partial V_c}\frac{\mathrm{d}V_c}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}t_r}t_{\rm res}. \quad (9)$$

Equations (3) and (4–9) are equivalent, but (4–9) are practical to generate discrete Monte Carlo realizations of a given source population. Moreover, Equations (4–9) provide a transparent interpretation of (3). The total RMS background, $h_c \propto \sqrt{N}h \propto f^{-2/3}$, comes about because the mean number of binaries per frequency bin is $N \propto t_{\rm res}^{\rm gw} \propto f_r^{-8/3}$ and each binary generates an RMS strain $h \propto f_r^{2/3}$. The scaling $h_c \propto f^{-2/3}$ is a consequence of averaging over the

[1] In practice $\partial^4 N/(\partial m_1\partial m_2\partial t_r\partial V_c)$ is a function of m_1, m_2, and z.

local inspiral episodes of merging binaries in the GW driven regime, but is completely independent on the overall cosmological merger history or on the involved MBH masses. The latter affects only the overall constant of proportionality. Eq. (3) also shows that this scaling constant is insensitive to the cosmological redshift distribution of mergers, $h_c \propto (1+z)^{1/6}$, as well as the number of minor mergers with $\mu \ll M$, once the total mass in satellites merging with BHs of mass M is fixed (Phinney 2001). However, the background is sensitive to the assembly scenarios of major mergers (SVV09). More importantly, as shown in SVC08, the actual GW signal in any single realization of inspiralling binaries is qualitatively different from $h_c \propto f^{-2/3}$, as a small discrete number of individual massive binaries dominate the nHz GW-background, creating a very spiky GW spectrum. We further discuss the discrete nature of the signal in Section 2.3 below.

2.2 Gas-driven inspirals

Let us now derive the GW background for an arbitrary model of binary evolution. We can derive the background using the residence time $t_{\rm res}$ the binary spends at each semi-major axis a, where $t_{\rm res} < t_{\rm res}^{\rm gw}$ in the gas driven phase (cf. Eq. [8] for the purely GW driven case). We define $t_{\rm res}$ for various accretion-disk models in Section 3.2. Generally, the emitted GW spectrum is

$$\frac{{\rm d}E_{\rm gw}}{{\rm d}\ln f_r} = \frac{{\rm d}E_{\rm gw}}{{\rm d}t_r}\frac{{\rm d}t_r}{{\rm d}\ln f_r} = \frac{2}{3}\frac{{\rm d}E_{\rm gw}}{{\rm d}t_r}t_{\rm res}. \quad (10)$$

The second equality follows the definition of $t_{\rm res}$ and Kepler's law (see Eq. [7]). The emitted power ${\rm d}E_{\rm gw}/{\rm d}t_r$ depends only on the masses and the geometry of the orbit, but is independent of the global migration rate of the binary, and therefore it is the same as in the pure GW driven case. The effects of migration is fully encoded in $t_{\rm res}$. Plugging back in Eq. (1), the mean square signal in the gas driven phase is

$$h_c^2(f) = \frac{4f^{-4/3}}{3\pi^{1/3}}\int {\rm d}z{\rm d}m_1{\rm d}m_2\,\frac{\partial^3 n}{\partial z\partial m_1\partial m_2}\frac{\mu M^{2/3}}{(1+z)^{1/3}} \times\frac{t_{\rm res}(M,\mu,f_r)}{t_{\rm res}^{\rm gw}(\mathcal{M},f_r)} \quad (11)$$

A comparison with (3) shows that the RMS signal is attenuated in the gas driven phase by $\sqrt{t_{\rm res}(M,\mu,f_r)/t_{\rm res}^{\rm gw}(\mathcal{M},f_r)}$. This factor is a complicated function of f_r, that behaves differently for different masses and mass ratios of the binaries; the overall spectrum is no longer a powerlaw.

We generate Monte Carlo realizations of the GW signal by sampling the population of the inspiralling systems. To do this, it is sufficient to recognize that the derivation given by Eqs. (4–9) remains valid in the gas drive phase if using the appropriate $t_{\rm res}(M,\mu,f_r)$, since the contribution given by each individual source to the signal is the same. *The net GW spectrum changes because of a reduction in the number of sources in the gas driven case.*

Note that the individual GW signal given by Eq. (5) depends on three parameters only: $\mathcal{M}$, z, and f_r. This implies that signals from sources with the same observed frequency, redshift, and chirp mass, $(f_r, z, \mathcal{M})$, but different mass ratios, q, are totally indistinguishable[2]. We can make use of this property and reduce the number of independent parameters in the distribution by integrating over the mass ratio in Eq. (4)

$$\frac{\partial^3 N}{\partial\mathcal{M}\partial z\partial\ln f_r} = \int_0^1 dq\frac{\partial^4 N}{\partial m_1\partial m_2\partial z\partial\ln f_r}\left|\frac{\partial(m_1,m_2)}{\partial(\mathcal{M},q)}\right|, \quad (12)$$

Note, that this step is different for the GW and gas driven cases, because $\partial^4 N/(\partial m_1\partial m_2\partial z\partial\ln f_r)$ is proportional to $t_{\rm res}$ in Eq. (9). Here $|\partial(m_1,m_2)/\partial(\mathcal{M},q)|$ is the determinant of the Jacobian matrix corresponding to the variable change from (m_1,m_2) to $(\mathcal{M},q)$. With Eq. (9) and (12) we derive $\partial^3 N/(\partial\mathcal{M}\partial z\partial\ln f_r)$ for any gas driven model given by $t_{\rm res}(M,\mu,f_r)$, and draw Monte Carlo samples of inspiralling binaries from this distribution when generating the GW signal.

2.3 Statistical characterization of the signal

In observations with PTAs, radio-pulsars are monitored weekly for total periods of several years. Assuming a repeated observation in uniform Δt time intervals for a total time T, the maximum and minimum resolvable frequencies are $f_{\rm max} = 1/(2\Delta t)$, corresponding to the Nyquist frequency, and $f_{\rm min} = 1/T$. The observed GW spectrum is therefore discretely sampled in bins of $\Delta f = f_{\rm min}$. For circular orbits, the frequency of the GWs is twice the orbital frequency.

Let us examine whether the sources' GW frequency evolves during the observation relative to the size of the frequency bins. Writing the frequency shift during an observation time T as $\Delta f_{\rm evol} \approx \dot{f}T = ({\rm d}\ln f/{\rm d}\ln a)({\rm d}\ln a/{\rm d}t)fT = \frac{3}{2}fT/[(1+z)t_{\rm res}]$ and considering the frequency resolution bin to be $\Delta f_{\rm bin} = 1/T$, then the frequency evolution relative to the frequency resolution bin is

$$\frac{\Delta f_{\rm evol}}{\Delta f_{\rm bin}} \approx \frac{3}{2}\frac{fT^2}{(1+z)t_{\rm res}} = \frac{0.015}{1+z}\mathcal{M}_{8.5}^{5/3}f_{50}^{11/3}T_{10}^2\frac{t_{\rm res}^{\rm gw}}{t_{\rm res}}, \quad (13)$$

where $\mathcal{M}_{8.5}$ is the chirp mass in units of $10^{8.5}M_\odot$, f_{50} is the frequency in units of 50 nHz, T_{10} is the observation time in unit of 10 years, and for the second equality we have used equation (8). Equation (13) shows that typical binaries contribute to a single frequency bin as stationary sources in the GW-driven regime.[3] We shall demonstrate that this is also true in the gas driven case (see Figure 1 below).

We generate N_k different realizations of the signal (usually $N_k = 1000$), i.e. N_k realizations of the MBHB population, consistent with Eq. (9) (see Section 3.1). Each of those consists of $N_b \sim 10^3 - 10^4$ binaries producing a relevant contribution to the signal, which we label by $(\mathcal{M}_i, z_i, f_{r\,i})$, $i = 1, 2, \ldots, N_{\rm b}$. The total signal (Eq. [4]) in each frequency resolution bin Δf is evaluated as the sum of the contributions of each individual source (see Amaro-Seoane et al. (2009) for the detailed numerical procedure)

$$h_c^2(f) = \sum_i h_{c,i}^2(\mathcal{M}_i, z_i, f_{r\,i}),, \quad (14)$$

[2] This is true only in the angular averaged approximation. We neglect the directional sensitivity of PTAs.
[3] The detection of the frequency shift of an $\mathcal{M}_{8.5} = f_{50} = z = 1$ source would require an extended observation with $T \gtrsim 35$ yr.

where for each f the sum is over the inspiralling sources emitting in the corresponding blue–shifted (i.e. restframe) Δf_r frequency resolution bin, and $h_{c,i} = h_i\sqrt{f_i T}$ is the angle and polarization averaged GW strain given by Eq. (5), multiplied by the square root of the number of cycles completed in the observation time. For comparison, we also evaluate the continuous integral Eq. (4), which represents the RMS average of Eq. (14) over different realizations of MBHB populations in the limit $N_k \to \infty$.

Since the mass function of merging binaries is in general quite steep, the relative contributions of the few most massive binaries turn out to dominate the background in each frequency bin. The total GW signal depends very sensitively on these rare binaries, and the inferred spectrum is very spiky. It is useful to separate the total signal h_c into a part generated by a population of GW–bright *individually resolvable sources*, and a *stochastic level* h_s, which includes the contribution of all the unresolvable, dimmer sources. More precisely, in each frequency resolution bin, we find the MBHB with the largest $h^2_{c,i}(\mathcal{M}_i, z_i, f_{r\,i})$, and we define it *individually resolvable* if its signal is stronger than the total contribution of all the other sources in that particular frequency bin. The *stochastic level* is consequently defined by adding up only the unresolvable sources in Eq. (14). Since the signal is dominated by few individual sources in each frequency bin, the $h_s(f)$ distribution obtained over the N_k realizations is far from being Gaussian or just even symmetric. To give an idea of the uncertainty range of $h_s(f)$, we calculate the 10%, 50% and the 90% percentile levels of the $h_s(f)$ distribution of the N_k different realizations.

The separation of individually resolvable sources is useful for several reasons (see SVC; SVV09). First, it is useful from a statistical point of view for understanding the variance of the expected GW spectrum among various realizations of the inspiralling MBHBs. The discrete nature of the resolvable sources allows a different statistical analysis than for the smooth background level corresponding to the stochastic level, $h_s(f)$. The individually resolvable signals could also be important observationally. A sufficiently GW–bright resolvable binary allows to measure the GW polarization using PTAs, and give information on the sky position of the binary (Sesana & Vecchio 2010), which might be used to search for direct electromagnetic signatures like periodically variable AGN activity (HKM09). A coincident detection of GWs and electromagnetic emission of the same binary would have far reaching consequences in fundamental physics, cosmology, and black hole physics (Kocsis, Haiman, & Menou 2008).

2.4 Timing Residuals

In general, the characteristic GW amplitudes (either of a stochastic background or of a resolvable source) can be translated into the pulsar timing language by converting $h_c(f)$ into a "characteristic timing residual" $\delta t_c(f)$ corresponding to the sky position and polarization averaged delay in the time of arrivals of consequent pulses due to GWs,

$$\delta t_c(f) = \frac{h_c(f)}{2\pi f}. \tag{15}$$

The pulsar timing residuals expected from an individual stationary GW source is derived in section 3 of SVV09 in detail. The corresponding measurement can be represented, in the time domain, with a residual:

$$\delta t(t) = r(t) + \delta t_{\rm N}(t) \tag{16}$$

where $r(t)$ is the contribution due to the GW source (which accumulates continuously with observing time t, see below), and $\delta t_{\rm N}(t)$ represents random fluctuations due to noise. The latter is the superposition of the intrinsic noise in the measurements and the GW stochastic level from the whole population of MBHBs, with a root-mean-square (RMS) value

$$\delta t^2_{\rm N,rms}(f) = \langle \delta t^2_{\rm N}(f)\rangle = \delta t^2_{\rm p}(f) + \delta t^2_{\rm s}(f). \tag{17}$$

where $t_{\rm p}(f)$ is the RMS instrumental and astrophysical noise corresponding to the given pulsar, and $\delta t_{\rm s}(f) = h_s(f)/(2\pi f)$ is due to the RMS stochastic GW background of unresolved MBHBs, as defined in the previous section.

The sky angle and orbital orientation-averaged signal-to-noise ratio (SNR) at which one MBHB, radiating at (GW) frequency f, can be detected using a *single* pulsar with matched filtering is

$${\rm SNR}^2 = \frac{\delta t^2_{\rm gw}(f)}{\delta t^2_{\rm N,rms}(f)}\,. \tag{18}$$

Here $\delta t_{\rm gw}(f)$ is the root-mean-squared timing residual signal resulting from GWs emitted by the individual stationary source over the observation time T defined as:

$$\delta t_{\rm gw}(f) = \sqrt{\frac{8}{15}}\frac{h(\mathcal{M},z,f)}{2\pi f}\sqrt{fT} \tag{19}$$

where $h(\mathcal{M},z,f)$ is the angle and polarization averaged GW strain amplitude given by Eq. (5), the prefactor $\sqrt{8/15}$ averages the observed signal over the "antenna beam pattern" of the array (Eq. (21) in SVV09[4]), and the $\sqrt{fT}$ term accounts for the residual build-up with the number of cycles. For $N_{\rm p}$ number of pulsars, the total detection SNR, of an individually resolvable MBHB is the RMS of the contributions of individual pulsars given by (18). For $N_{\rm p}$ identical pulsars, the effective noise level is therefore attenuated by $N_{\rm p}^{-1/2}$.

In the following we will represent the overall GW signal and the stochastic background by using either their characteristic amplitudes, $h_c(f)$ and $h_s(f)$, or the corresponding characteristic timing residuals, $\delta t_c(f)$ and $\delta t_s(f)$, according to equation (15). We study the detection significance of individually resolvable sources and the distribution of their numbers as a function of the induced $\delta t_{\rm gw}$. For each Monte Carlo realization of the emitting MBHB population, we count the cumulative number of all (N_t) and resolvable (N_r) sources above $\delta t_{\rm gw}$ as a function of $\delta t_{\rm gw}$:

$$N_{t/r}(\delta t_{\rm gw}) = \int_{\delta t_{\rm gw}}^{\infty} \frac{\partial N_{t/r}}{\partial \delta t'_{\rm gw}} \delta t'_{\rm gw}\,, \tag{20}$$

where the integral is either over all sources or only the individually resolvable sources (i.e. restricted to those that produce residuals above the RMS stochastic level, see Sec. 2.3).

[4] Note that that the square root in the prefactor $\sqrt{8/15}$ is missing in Eq. (20) of SVV09 because of a typo there.

3 THE EMITTING BINARY POPULATION

We calculate the GW signal by generating a catalogue of binaries consistent with Eq. (9). This requires (i) a model for the comoving merger rate density of coalescing MBHBs, $\partial^4 N/(\partial m_1 \partial m_2 \partial t_r \partial V_c)$, and (ii) a model for the evolution of individual inspiralling binaries $t_{\rm res}(M, \mu, f_r)$. These two items are the subjects of the next two subsections below.

3.1 Population of coalescing massive black hole binaries

We use the population models described in section 2 of SVV09, the reader is deferred to that paper for full details. We extract catalogs of merging binaries from the semi-analytical model of Bertone et al. (2007) applied to the Millennium run (Springel et al. 2005). We then associate a central MBH to each merging galaxy in our catalogue. We explored a total of 12 models, combining four $M_{\rm BH}$-bulge prescriptions found in the literature with three different accretion scenarios during mergers. The twelve models are listed in table 1 of SVV09. In the present study we shall use the Tu-SA population model as our default case. In this model, the MBH masses in the merging galaxies correlate with the masses of the bulges following the relation reported in Tundo et al. (2007), and accretion is efficient onto the more massive black hole, *before* the final coalescence of the binary. Since the comparison among different MBHB population models is not the main purpose of this study, we will present results only for this model. However, we tested our dynamic scenario on other population models presented in SVV09, finding no major differences for alternative models.

Assigning a MBH to each galaxy, we obtain a catalogue of mergers labelled by MBH masses and redshift. From this, we generate the merger rate per comoving volume, $\partial^4 N/(\partial m_1 \partial m_2 \partial t_r \partial V_c)$. In practice, due to the large number of mergers in the simulation, this is a finely resolved continuous function, describing the merger rate density as a function of m_1, m_2, and z. After plugging into Eqs. (9) and (12), we can obtain the continuous distribution $\partial^3 N/(\partial \mathcal{M} \partial z \partial \ln f_r)$ one would observe in an "ideal snapshot" of the whole sky. We then sample this distribution to generate random Monte Carlo realizations of the GW signal. In summary, the chosen MBHB population model fixes the cosmological merger history, i.e. the function $\partial^4 N/(\partial m_1 \partial m_2 \partial t_r \partial V_c)$, and "Monte Carlo sampling" refers to first choosing a realization of the cosmological merger history (i.e. generate N_b number of binary masses and redshifts[5]) and then assigning an orbital frequency (or time to merger) to each binary. In addition to our fiducial merger rate, we also examine for the first time the situation where minor mergers do not contribute to the coalescence rate of the central MBHs. This is motivated by recent numerical simulations indicating that minor mergers with mass ratios $q < 0.1$ lead to tidal stripping of the merging satellite, and the resulting core does not sink efficiently to the centre of the host galaxy (Callegari et al. 2009). We identify the mass ratio of merging galaxies using the mass of the stellar components, to avoid complications due to the tidal stripping of merging dark matter halos. In practice, we find that suppressing all the minor mergers does not affect the resulting GW signal, implying that the contribution of mergers involving dwarf galaxies is negligible.

3.2 Binary evolution in massive circumbinary disks

We adopt the simple analytical models of HKM09 to describe the dynamical evolution of MBHBs in a geometrically thin circumbinary accretion disk.

The inspiral rate is defined using the (comoving) residence time the binary spends in a logarithmic separation bin centred on the semimajor axis a, $t_{\rm res} = a({\rm d}t_r/{\rm d}a)$. The inspiral rate is non-uniform as the binary separation shrinks, as anticipated in Section 1, and it evolves through several phases: (i) initially, migration occurs on a timescale $t_{\rm d}$ dominated by the disk viscosity, and then, when the secondary mass is much larger then the thin disk mass enclosed in the binary orbit, it settles on a timescale $t_{\rm s}$ dictated by the secondary BH dynamics; (ii) the opacity changes from free-free $\kappa_{\rm ff}$ at large binary separations to electron scattering $\kappa_{\rm es}$ at smaller separations; (iii) the pressure is dominated by gas pressure $p_{\rm gas}$ for wide systems and by radiation pressure $p_{\rm rad}$ for close ones; (iv) eventually, the inspiral rate due to GW emission dominates over the gas-driven type-II migration rates. We can define 4 corresponding transition points between the various phases: $a_{\rm s/d}$, $a_{\rm ff/es}$, $a_{\rm gas/rad}$, and $a_{\rm II,GW}$. The transition radii depend on the binary masses and the accretion disk parameters (see HKM09), but typically obey

$$a_{\rm II,GW} < a_{\rm gas/rad} < a_{\rm ff/es} < a_{\rm s/d}. \tag{21}$$

We consider two different models of steady accretion disks, α and β–disks. For the classic Shakura & Sunyaev (1973) α model, the viscosity is proportional to the total (gas+radiation) pressure of the disk. Until very recently, this model, if radiation pressure dominated, has been thought to be thermally and viscously unstable (Lightman & Eardley 1974; Piran 1978). In the alternative β-model, the viscosity is proportional to the gas pressure only[6], and it is stable in both sense. The nature of viscosity is not well understood to predict which of these prescriptions lies closer to reality. Recent numerical magneto-hydrodynamic simulations (Hirose, Krolik, & Blaes 2009) suggests that the thermal instability is avoided in radiation pressure dominated situations because stress fluctuations lead the associated pressure fluctuations, and seem to favor the α prescription over the β–model. We carry out all calculations for both models, but consider the α prescription as our fiducial disk model. In both cases, the model is uniquely determined by three parameters: the central BH mass, the accretion rate $\dot{M}$, and the α viscosity parameter. The exact value of these parameters is not well known. Observations of luminous AGN imply an accretion rate around $\dot{m} = \dot{M}/\dot{M}_{\rm Edd} = (0.1$–$1)$ with a statistical increase towards higher quasar luminosities (Kollmeier et al. 2006; Trump et al. 2009). Here $\dot{M}_{\rm Edd} =$

[5] Here N_b is chosen randomly for each distribution, it can vary for different realizations of the population according to a Gaussian distribution with $\sigma = 1/\sqrt{\langle N_b \rangle}$ around the mean $\langle N_b \rangle$.

[6] The name comes from the definition $\nu \propto \alpha p_{\rm gas}^{\beta} p_{\rm tot}^{1-\beta}$ where $\beta = 1$ for the β–model, while $\beta = 0$ for the α model. In both cases α is a free model parameter.

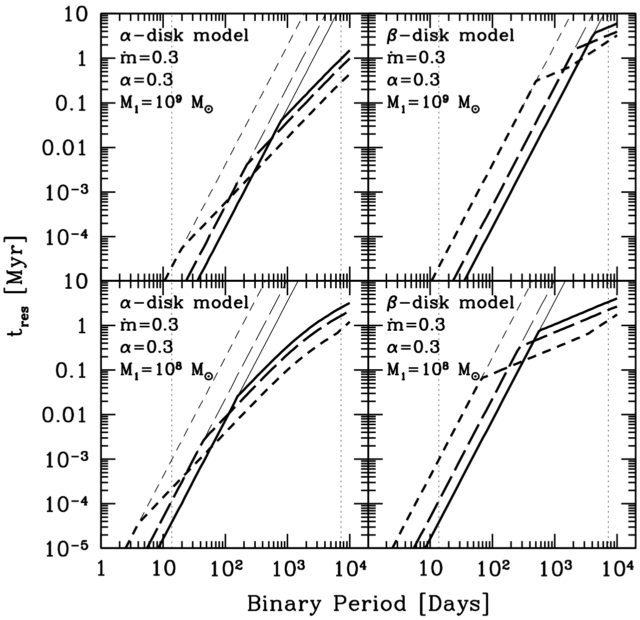

Figure 1. Residence time, $t_{\rm res}$, as a function of the binary period. Different panels correspond to selected model parameters as labelled. In each panel the thick curves represent the residence time for the gas driven dynamics according to HKM09. Solid, long–dashed and short–dashed curves are for $q = 1, 0.1, 0.01$ respectively. The two thin dotted vertical lines approximately bracket the PTA observable window. The extrapolated pure GW-driven evolution ($t^{\rm gw}_{\rm res} \propto t^{8/3}_{\rm orb}$) is shown as thin lines, for comparison. The ratio $t_{\rm res}/t^{\rm gw}_{\rm res}$ gives the relative decrease in the number of binaries in the gas dominated case compared to the GW driven case.

$L_{\rm Edd}/(\eta c^2)$ is the Eddington accretion rate for $\eta = 10\%$ radiative efficiency, where $L_{\rm Edd}$ is the Eddington luminosity. Observations of outbursts in binaries with an accreting white dwarf, neutron star, or stellar black hole imply $\alpha = 0.2$–0.4 (Dubus, Hameury, & Lasota 2001; King, Pringle, & Livio 2007, and references therein). Theoretical limits based on simulations of magneto-hydrodynamic turbulence around black holes are not conclusive, but are consistent with α in the range 0.01–1 (Pessah, Chan, & Psaltis 2007). It is however unclear whether these numbers are directly applicable to circumbinary MBH systems. We explore several choices covering all 6 combinations with $\dot{m} = \{0.1, 0.3\}$ and $\alpha = \{0.01, 0.1, 0.3\}$ for both α and β–disks, respectively. Motivated by the considerations outlined above, we highlight the α-disk with $\dot{m} = 0.3$ and $\alpha = 0.3$ as our default model.

These disk–models were developed for a single accreting BH. In the case of a binary, we assume that the structure of the disk is given by that of a single steadily accreting BH with a mass equal to the total binary mass. The effect of the binary is to generate a perturbation, a spiral density wave, which in turn slowly drains angular momentum from the system. Figure 1 shows the evolution of the residence time $t_{\rm res}$ as a function of the binary period for selected MBH masses and disk models (for more illustrations and detailed discussion see Figures 1–5 in HKM09). The disk evolution in the various regimes summarized by Eq. (21) is as follows. At the smallest orbital periods $t_{\rm orb}$, the evolution is driven by GWs, and $t_{\rm res} = t^{\rm gw}_{\rm res} \propto t^{8/3}_{\rm orb}$ (see Eq. 8). At larger separation, the gas can shrink the binary at a much faster rate and determines the evolution. In the secondary-dominated Type-II migration regime for sufficiently close orbits the radiation pressure dominates, and $t_{\rm res} \propto t^{35/24}_{\rm orb}$–$t^{7/12}_{\rm orb}$, for α and β disks respectively. Further out in the gas pressure dominated regime $t_{\rm res} \propto t^{7/12}_{\rm orb}$–$t^{25/51}_{\rm orb}$ (with electron scattering versus free-free opacity). At the largest separations in the disk-dominated Type-II migration regime $t_{\rm res} \propto t^{5/6}_{\rm orb}$.

Note that for quadrupole radiation the frequency of the GW is simply $f_r = 2/t_{\rm orb}$. The number of binaries at any given $t_{\rm orb}$ is proportional to the residence time $t_{\rm res}$. Therefore, the decrease in $t_{\rm res}$ in the gas driven regime compared to the GW driven case implies a decrease in the population of MBHs, which ultimately leads to the attenuation of the low frequency end of the observable total GW spectrum. The RMS GW spectrum averaged over the whole population of binary inspiral episodes is no longer a powerlaw. It is interesting to examine the contributions of various evolutionary phases according to Eq. (11), $h_c \propto f^{-2/3}\sqrt{t_{\rm res}/t^{\rm gw}_{\rm res}}$. If all binaries were in the GW driven phase then $h_c \propto f^{-2/3}$. If all were in the secondary-dominated Type-II migration regime with a radiation pressure dominated disk[7] $h_c \propto f^{-1/16}$–$f^{3/8}$ for α and β disks, respectively, further out $h_c \propto f^{3/8}$–$f^{43/102}$ in the gas pressure dominated regime (with electron scattering versus free-free opacity), and finally $h_c \propto f^{1/4}$ in the disk-dominated Type-II migration regime. In any case, the GW spectrum $h_c(f)$ is much shallower in the gas driven phase. Note the gas driven phase contributes a nearly flat or an *increasing* spectrum $h_c(f)$, very different from the nominal $f^{-2/3}$ GW driven case. In general, the orbital separation of the more massive objects at a given f_r is smaller in terms of their Schwarzschild radii. Since the transition to the gas driven region, when expressed in Schwarzschild radii, is roughly independent of mass, it follows then at any given frequency bin the more massive objects are typically GW driven and lighter objects are gas driven. The total average spectrum[8] assuming only wet [9] mergers, is between $f^{-2/3}$ and $f^{0.4}$ depending on the ratio of GW driven to gas driven binaries. Note however, that the individually resolvable sources are typically very massive and are in the GW-driven phase for the relevant range of binary separations, and therefore their properties should not be modified by gas effects.

4 RESULTS

4.1 Description of the signal

As stated in Sec. 2.3, the relevant frequency band for pulsar timing observations, assuming a temporal observation baseline T and a time interval between subsequent observations

[7] Most of the gas driven binaries contributing to the background at large frequencies are in this regime.

[8] if averaging over each binary episode but not over the cosmological merger tree

[9] Here “wet” refers to gas rich mergers where the dynamics of the binary is driven by both GW emission and by torques exerted by the circumbinary disk, while “dry” refers to the case where there is a little gas to be funneled in the galactic center, and the binary dynamics is driven by GW emission only. Note that in both cases we neglect interaction with stars.

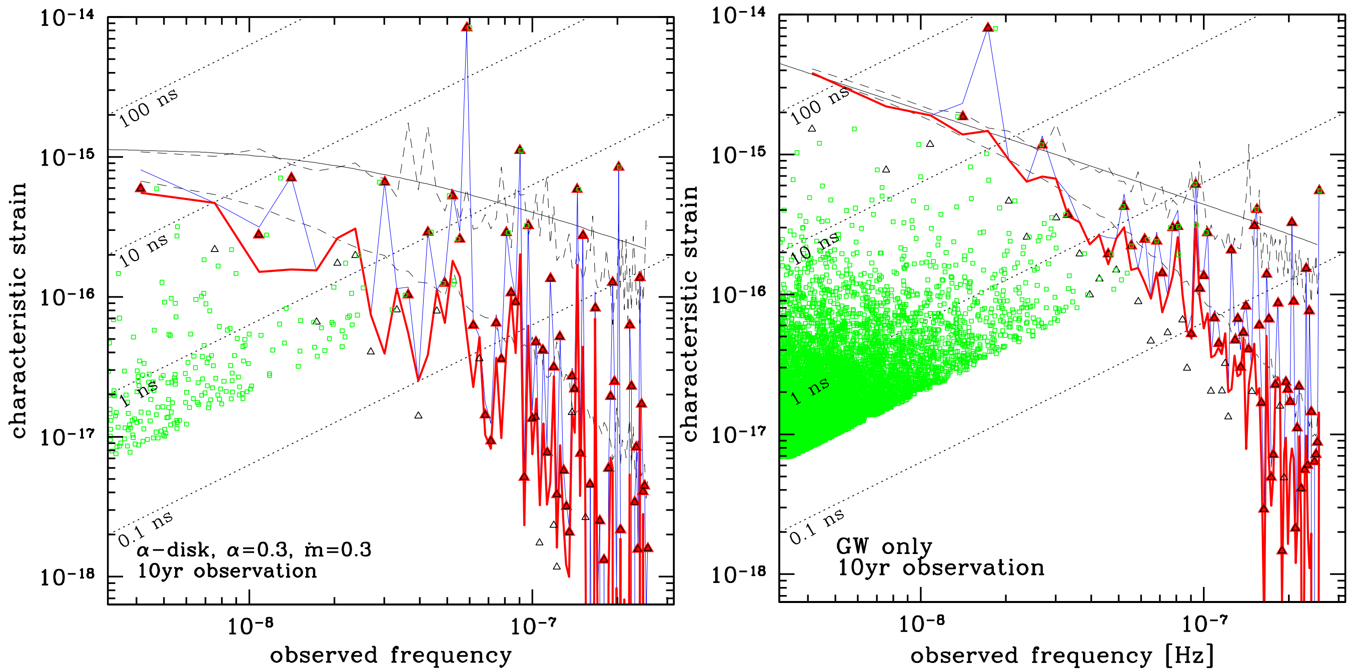

Figure 2. Components of the GW signal from a population of inspiralling MBHBs. In the left panel we consider all binaries embedded in a gaseous α-disk for our default model (all mergers wet), while in the right panel all binaries are purely GW driven (all mergers dry). In each panel, the smooth solid line is the RMS total characteristic GW strain h_c using the integral expression (4–9) (which correspond to an average over $N_k \to \infty$ Monte Carlo realizations), the two dashed black lines represent the RMS total signal (upper) and the RMS stochastic background level (lower) averaged over $N_k = 1000$ Monte Carlo realizations respectively. The jagged blue line displays a random Monte Carlo realization of the GW signal. The small black and red triangles show the contributions of the brightest and resolvable sources in each frequency bin respectively. The jagged red line is the stochastic GW background for this realization, i.e., once the resolvable sources in each frequency bin are subtracted. The green dots label all the systems producing an RMS residual $t_{\rm gw} > 0.3\,$ns over $T = 10\,$years. The dotted diagonal lines shows constant $t_{\rm gw}$ levels as a function of frequency. An observation time of 10 years is assumed.

Δt, is between $f_{\rm min} \approx 1/T$ and $f_{\rm max} \approx 1/(2\Delta t)$ with a resolution $\Delta f \approx 1/T$. In our calculations we assume a default duration of $T = 10\,$yr for the PTA campaign with $\Delta t \approx 1$ week. This gives $f_{\rm min} \approx 3\times10^{-9}\,$Hz, $f_{\rm max} \approx 10^{-6}$Hz and $\Delta f \approx 1/T \approx 3\times10^{-9}\,$Hz. The simulated signal is computed by doing a Monte Carlo sampling of the distribution $\partial^3 N/(\partial z \partial \mathcal{M} \partial \ln f_r)$, and adding the GW contribution of each individual source. In each frequency bin, we identify the individually resolvable sources and the stochastic background. We repeat this exercise $N_k = 1000$ times for the 12 steady disk models defined in Sec. 3.2 and for the purely GW–driven case. In addition, we also calculate the GW signal using the integral expression (4–9) using the continuous distribution function, which corresponds to the RMS average of the GW signal in the $N_k \to \infty$ limit.

The GW signal and its most important ingredients are plotted in Figure 2. The left panel shows results for gas–driven inspirals using our default α-disk (i.e. "gas on" – all mergers wet), the right panel shows results for purely GW–driven inspirals for comparison (i.e. "gas off" – all mergers dry) for the same underlying cosmological MBHB coalescence rate. A randomly selected Monte Carlo realization of the signal is depicted as a dotted blue jagged line. The green dots represent the contributions of individual binaries; systems producing $\delta t_{\rm gw}(f) > 0.3\,$ns timing residual are shown. In each frequency bin, the brightest source is marked by a black triangle. The individually resolvable sources are marked by superposed red triangles, and the stochastic level from all unresolvable sources is shown with a solid red jagged line. Clearly, the GW signal for any single realization is far from being smooth. The noisy nature of the signal is due to rare massive binaries rising well above the stochastic level. Solid black curves in Figure 2 show $h_c(f)$, the RMS value of the GW signal averaged over the merger episodes, using the integral expression (4–9). The upper black dashed curve is the GW signal, averaged over $N_k = 1000$ Monte Carlo realizations of the merging systems. This is still noisy, due to the finite number of realizations used, but is consistent with the integrated average shown by the solid black curve. The lower black dashed curve marks the RMS stochastic component, $h_{s,\rm rms}(f)$, averaged over the same $N_k = 1000$ realizations. Figure 2 shows that the gaseous disk greatly reduces the number of binaries at small frequencies compared to the GW–only case, as gas drives the binaries in quickly towards the final coalescence. Consequently, the signal is more spiky in the presence of gas. There is a clear flattening of the RMS spectrum at low frequencies ($f < 1/$yr) in the gas-driven case compared to the purely GW-driven case. The stochastic level is more suppressed in the wet-only case and is much steeper than $f^{-2/3}$ at high frequencies. Despite the spiky signal in a given Monte Carlo realization, the overall shape of the background is well recognizable in the GW driven model, while the characterization of the global shape of the signal in the gas driven case appears to be less viable.

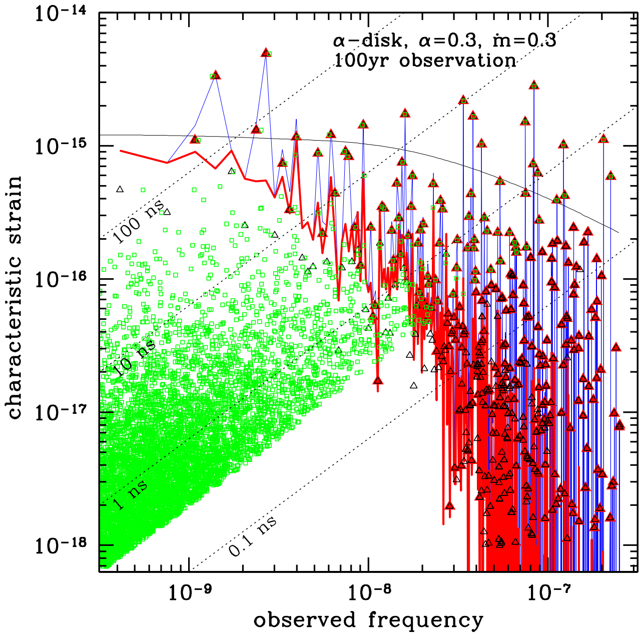

Figure 3. Same as left panel of Figure 2, but for an observation lasting 100 years (Averages over the 1000 realizations are not shown in this case for clarity).

Gas driven migration becomes more and more prominent at large binary separations, corresponding to large orbital times or small GW frequencies. Therefore, to check the ultimate maximum impact of gas effects on future PTA detections, we simulated the spectrum for a hypothetical $T = 100\,\mathrm{yr}$ observation baseline. Figure 3 shows a realization of the spectrum for such an extended observation, assuming our default gas model (i.e. all mergers wet). Given the extended temporal baseline, the minimum observable frequency is pushed down to $\sim 3\times10^{-10}$ Hz, where the spectrum of gas driven mergers is considerably flatter. Moreover, the frequency resolution bin is then narrower, the number of sources per bin is much smaller, and more sources become resolvable. At the smallest frequencies, the induced timing residual can be higher than 1μs and more than 50 sources will be individually resolvable at 1ns precision level, some of them with an SNR as high as 100. These numbers are not severely modified if considering purely GW–driven dry mergers only (see Fig. 5 below). Interestingly, due to the high frequency resolution of such an extended observation, the GW frequency of some of the resolvable binaries may evolve significantly during the observation (e.g. a binary with masses $m_1 = m_2 = 5 \times 10^8\,\mathrm{M}_\odot$ or higher are nonstationary relative to the bin size at frequencies above $f \gtrsim 40\,\mathrm{nHz}$, see equation [13]). Therefore it may be possible to detect the frequency evolution of individually resolvable binaries during such an extended monitoring campaign. However, this computation is idealized: it is questionable if millisecond pulsars can maintain a ns timing stability over such a long timescale, nonetheless, it points out the enormous capabilities of long term PTA campaigns.

4.2 Individually resolvable sources

Let us now examine the prospects for detecting individual sources using PTAs. How many sources are expected to be individually resolvable? How significant is their detection?

As explained in Sec. 2.4, the signal from a specific binary can be detected using a PTA if the corresponding timing residual $\delta t^2_{\rm gw}$ is above the RMS noise level $\delta t^2_{\rm N,rms}$ characterizing the PTA given by Eq. (18). We should notice, however, that this estimate of the number of individually resolvable sources is conservative and is likely to provide only a *lower limit* for the following reasons. Firstly, we average over the sky position of the binaries and pulsars, while in reality, it may be possible to take advantage of the different GW polarization and amplitude generated by sources in different sky positions to deconvolve their signal (even if they have similar strength and frequency). Secondly, the brightest source identification algorithm can be implemented recursively, after an accurate subtraction of the identified sources. However, we find that, especially at low frequencies, the distribution of GW source amplitudes for various binaries in a single frequency bin is not strongly hierarchical, so that a recursive brightest source finding algorithm shouldn't increase significantly the number of resolvable systems. We present here results both in terms of the total number (N_t) and resolvable systems (N_r, see Eq. 20).

In Figure 4 we plot the distribution of the number of sources (total and resolvable) as a function of timing residual, detection frequency, redshift, and chirp mass, found in two particular realizations of our default α-disk (all mergers wet) and in the purely GW-driven models (all mergers dry). The figure shows that even though there are much more sources in the purely GW-driven case, the number of sources rising above the stochastic level is almost the same in the purely dry and wet cases. Figure 4 also shows the chirp mass, redshift, and frequency distribution for sources above 1 ns timing level. As was previously shown in SVV09, the bulk of the sources are cosmologically nearby ($z \lesssim 1$) with masses peaking around $\mathcal{M} \sim 2 \times 10^8\,\mathrm{M}_\odot$. Figure 4 shows that gas dynamics does not introduce a major systematic change in the shape of the redshift and the chirp mass distributions. The lower left panel shows that gas removes a systematically larger fraction of sources at small frequencies.

Figure 5 shows the cumulative number of binaries (total and resolvable) as a function of timing residual. The upper panel shows results for the α–disk models with $\dot{m} = 0.3$ and different α. The statistics of resolvable sources is almost unaffected by the large suppression of the total number of sources at a fixed timing residual. For example, in all of our models, we expect ~ 2 resolvable sources at a timing level of $\delta t_{\rm gw} = 10\,\mathrm{ns}$, even though the total number of sources contributing to the signal at that level spans about an order of magnitude among the different models (~ 5 for $\alpha = 0.3$ to ~ 50 for binaries driven by GW only). The same is true for β–disk model (lower panel), even though in this case the total number of sources at a particular $\delta t_{\rm gw}$ is not reduced dramatically by gas effects. This result can be understood with a closer inspection of Figure 1. Let us focus on the α–disk model. As explained in Sec. 3.2, the impact of gas driven dynamics in the PTA window is more significant for lower binary masses and unequal mass ratios. These are the bina-

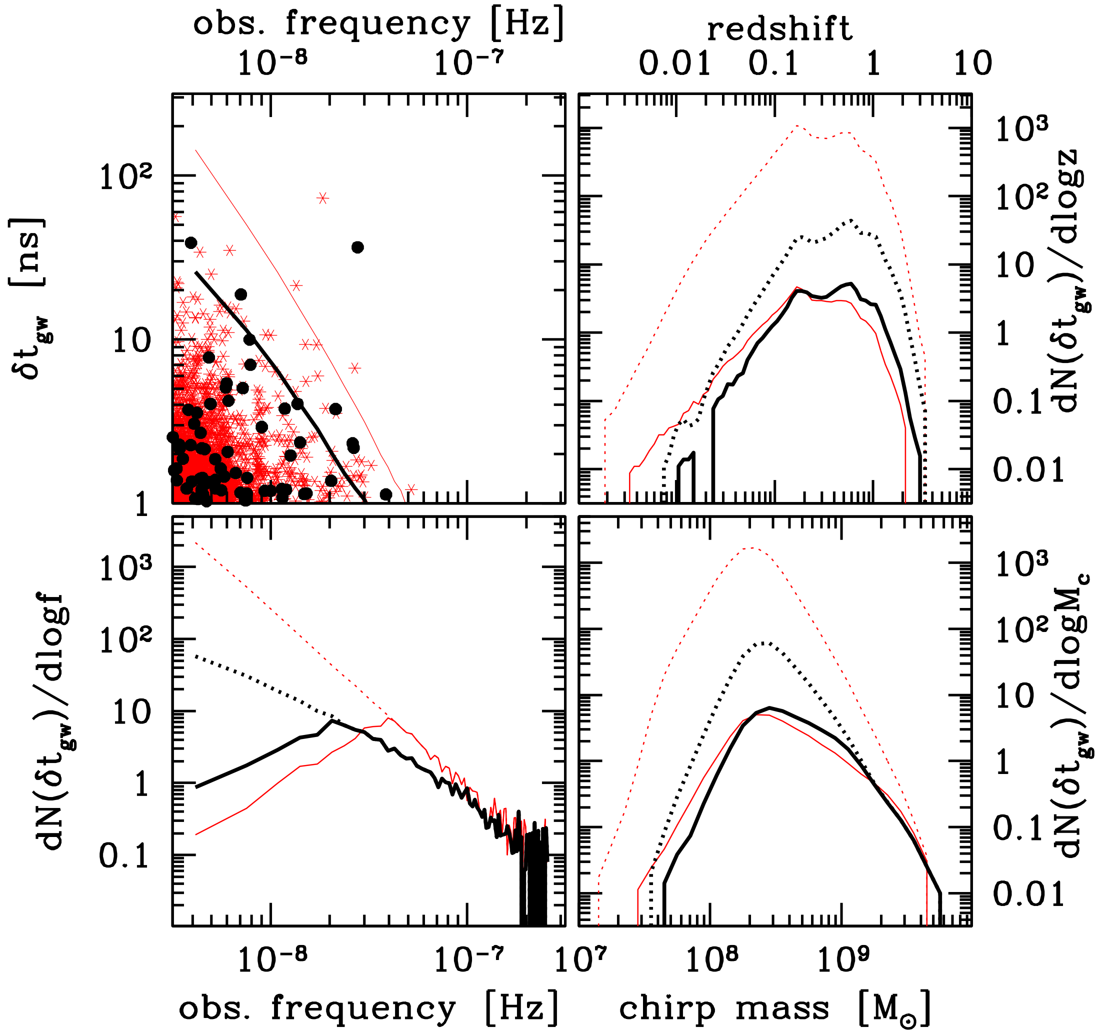

Figure 4. *Top-left panel*: characteristic amplitude of the timing residuals $\delta t_{\rm gw}$ (equation (19)) as a function of frequency; the dots are the residuals generated by individual sources and the solid line is the estimated *stochastic level* of the GW signal. *Top-right panel*: distribution of the number of total (dotted lines) and resolvable (solid lines) sources per logarithmic redshift interval as a function of redshift, generating a $\delta t_{\rm gw} > 1$ns. *Bottom-left panel*: distribution of the number of total (dotted lines) and resolvable (solid lines) sources per logarithmic frequency interval as a function of the GW frequency, generating a $\delta t_{\rm gw} > 1$ns. *Bottom-right panel*: distribution of the total (dotted lines) and individually resolvable (solid lines) number of sources per logarithmic chirp mass interval as a function of chirp mass, generating a $\delta t_{\rm gw} > 1$ns. All the black elements refer to our default disk model, the red elements are for a GW driven MBHB population. Distributions are averaged over $N_k = 1000$ realizations of the MBHB population.

ries that build-up the bulk of the signal, and its stochastic level is consequently greatly reduced in the gas driven case. On the other hand, the population of high equal mass binaries, which constitute most of the individually resolvable sources, is almost unaffected by the presence of the circumbinary disk, as they are already in the GW-driven regime in the relevant range of binary periods.

Figure 6 shows expectations on the detection significance of resolvable sources. The SNR distribution of resolvable sources (equation [18]) are shown as thin lines, accounting for the astrophysical GW noise from the unresolved binaries, but neglecting the intrinsic noise of the array (i.e. assuming an ideal detector with infinite sensitivity, $\delta t_{\rm p}^2(f) = 0$, in equation 17). This calculation represents *an upper limit* of the SNR. Thick lines, in Figure 6 plot the SNR considering a total detector noise of 1ns (appropriate for SKA). The figure shows that the expected detection significance of resolvable sources is systematical higher for gas driven models with larger α. In general, in an array with 1ns sensitivity, we may expect a couple of individually resolvable sources with SNR > 5. For near future instruments with much worse sensitivity, the identification of resolvable sources might be more challenging.

4.3 Stochastic background

Figure 7 shows the RMS stochastic level (i.e. after subtracting off the individually resolvable sources, see Sec. 2.3) for selected steady state gas disk models, averaged over $N_k = 1000$ realizations. The top left panel highlights the difference between wet and dry models, the top right and bottom left panels collect different α–disk models and the lower right panel is for selected β–disk models. The different line styles show the effect of changing the α parameter

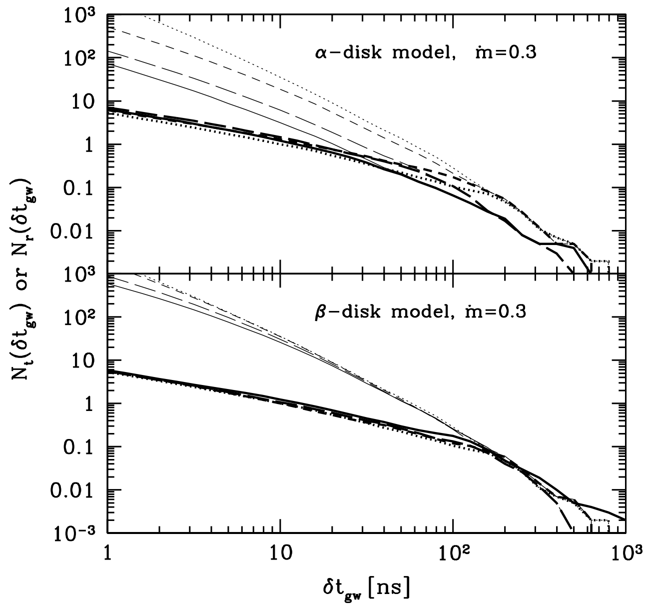

Figure 5. Cumulative number of total ($N_t(\delta t_{\rm gw})$, thin lines) and individually resolvable ($N_r(\delta t_{\rm gw})$, thick lines) sources emitting above a given $\delta t_{\rm gw}$ threshold as a function of $\delta t_{\rm gw}$. *Upper panel*: α–disk model with $\dot{m} = 0.3$. Lines are for α =0.3 (solid), 0.1 (long–dashed) and 0.01 (short–dashed). The dotted lines refer, for comparison, to the GW driven model. *Lower panel*: same as upper panel but for a β-disk model. All the distributions refer to the ensemble mean computed over all $N_k = 1000$ realizations of the MBHB population.

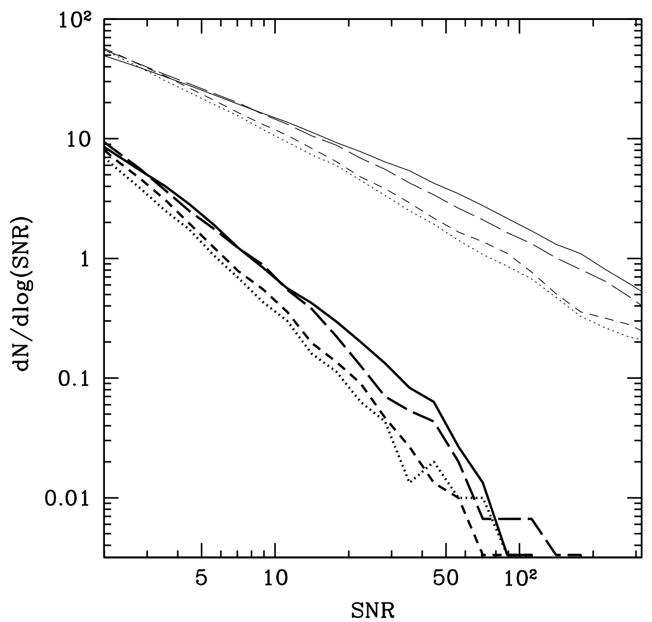

Figure 6. SNR distribution of individually resolvable sources. Thin and thick curves refer to neglecting the instrumental noise or using a $\delta t_{\rm gw} = 1$ ns timing precision, respectively. Linestyle as in the upper panel of Figure 5.

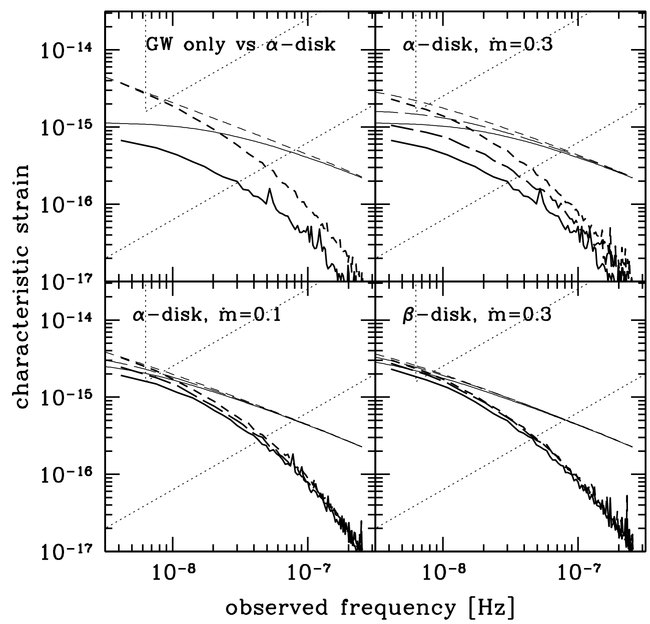

Figure 7. Influence of the gas driven dynamics on h_c (thin lines) and on h_s (thick lines). *Upper left panel*: GW driven dynamics versus gas driven dynamics for our default disk model. *Upper right panel*: $\alpha = 0.3$ (solid), 0.1 (long–dashed), 0.01 (short–dashed), for an α–disk with $\dot{m} = 0.3$. *Lower left panel*: same as the upper right panel, considering an α–disk with $\dot{m} = 0.1$. *Lower right panel*: same as the upper right panel, considering a β–disk with $\dot{m} = 0.3$. The two dotted lines in each panel represent the sensitivity of the complete PPTA survey and an indicative sensitivity of 1ns for the SKA.

of the disk. We also show the RMS total signal level, which is exactly proportional to $f^{-2/3}$ in the dry case. In general, the stochastic level matches the total signal level at low frequencies, but is increasingly suppressed for frequencies above $f \gtrsim 10^{-8}$ Hz for all of our models. At sufficiently large frequencies GW emission dominates even for wet mergers, and both the RMS total signal and the stochastic level approaches the purely GW–driven case. However, at small frequencies, a significant fraction of binaries is driven by gas, and the signal is attenuated and the spectrum is less steep compared to the dry case. Figure 7 shows that gas–driven migration suppresses the stochastic background significantly, by a factor of 5 for our standard disk model below 10^{-8} Hz. The suppression of the total and stochastic levels is a strong function of the model parameters. Interestingly, there is almost no suppression for β–disks, or for α–disks with a small accretion rate and/or a small α value. In these cases, the local disk mass is smaller, resulting in longer viscous timescales, and hence the population of widely separated binaries is not reduced significantly.

Figure 8 quantifies how the stochastic background changes among different Monte Carlo realizations, showing the range attained by 10–90% of all $N_k = 1000$ realizations. The variance is not the product of uncertainties related to the parameters of the adopted cosmological and dynamical model, but is purely determined by the small number statistics of sources per frequency bin, intrinsic to the source distribution. At $f = 10^{-8}$Hz, the variance of the signal produced by our default disk model is ~ 0.5dex. Decreasing α to 0.1 and 0.01, the variance drops to ~ 0.35dex and ~ 0.25dex

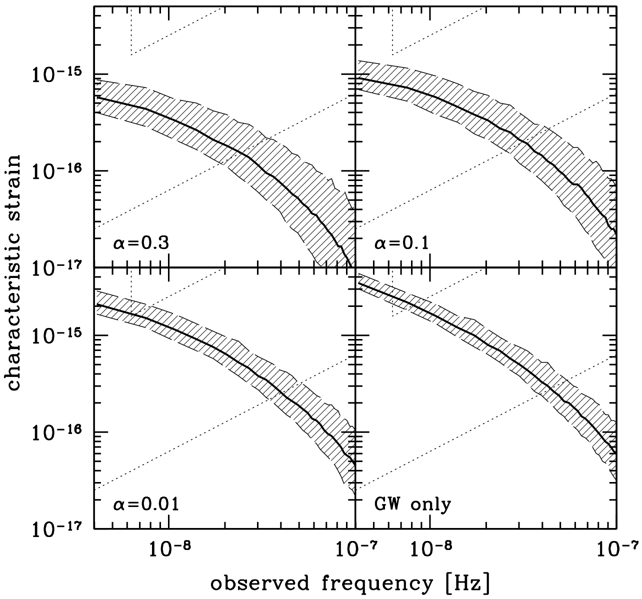

Figure 8. Variance of the expected stochastic level of the signal as a function of α. In all the panels we assume the α–disk model with $\dot{m} = 0.3$ (except for the lower right panel, referring to the GW driven model). Solid lines represent the median h_s over 1000 Monte Carlo realizations, while the shaded area enclosed within the two dashed lines is the 10%–90% confidence region . The two dotted lines in each panel represent the sensitivity of the complete PPTA survey and an indicative sensitivity of 1ns for the SKA.

respectively. In the case of all mergers driven by GWs, the variance at the same frequency is only $\sim$ 0.15dex. This means that, contrary to the GW driven model (SVC08), it is impossible to predict the stochastic level of the signal accurately for our default gas driven model. For a linear fit, any power law in the range $f^{-0.3} - f^{-1.1}$ is acceptable within the level of variance of the signal in the frequency range $3 - 30\,$nHz .

5 DISCUSSION AND CONCLUSIONS

The presence of a strong nHz gravitational wave signal from a cosmological population of massive black hole binaries, is a clear prediction of hierarchical models of structure formation, where galaxy evolution proceeds through a sequence of merger events. The detailed nature of the signal depends however on a number of uncertain factors: the MBHB mass function, the cosmological merger rate, the detailed evolution of binaries, and so on. In particular, MBHB dynamics determines the number of sources emitting at any given frequency, and thus the overall shape and strength of the signal. Previous works on the subject (e.g. SVC08 and SVV09) considered the case of GW driven binaries only. In this paper we studied the impact of gas driven massive black hole binary dynamics on the nHz gravitational wave signal detectable with pulsar timing arrays. This is relevant because in any merger event, cold gas is efficiently funnelled toward the centre of the merger remnant, providing a large supply of gas to the MBHB formed following the galaxy interaction. Even a percent of the galaxy mass in cold gas funnelled toward the centre, is much larger than the masses of the putative MBHs involved in the merger so that the newly–formed binary evolution may be driven by gas until few thousand years before final coalescence.

To conduct our study, we coupled models for gas driven inspirals of HKM09 to the MBHB population models presented in SVV09. Our simulations cover a large variety of quasistationary one-zone disk models for the MBHB-disk dynamical interactions (with an extensive exploration of the α viscosity parameter and $\dot{m}$ for α and β-disks), along with several different prescriptions for the merging MBHB population (four different black hole mass–galaxy bulge relations coupled with three different accretion recipes). The differences with respect to the purely GW driven models (presented in SVV09) are qualitatively similar for all the considered MBHB populations, we thus presented the results for the Tu-SA population model only, focusing on the impact of the different disk models. Our main findings can be summarized as follows:

- The effect of gas driven dynamics may or may not be important depending on the properties of the circumbinary disk. A robust result is that if the viscosity is proportional to the gas pressure only (β-disk models), there is basically no effect on the GW signal, independently of the other disk parameters. However, if the viscosity is proportional to the total (gas+radiation) pressure (α-disk models), then the GW signal can be significantly affected, especially for $\alpha \gtrsim 0.1$ and $\dot{m} \gtrsim 0.1$.
- With respect to the GW driven case, the presence of massive circumbinary disks affects the population of low–unequal mass binaries predominantly ($M < 10^8\,\mathrm{M}_\odot$, $q < 0.1$), causing a significant suppression of the *stochastic level* of the signal, but leaving the number and strength of massive *individually resolvable sources* basically unaffected. In our default model ($\alpha = 0.3$, $\dot{m} = 0.3$), the stochastic background is suppressed by a factor of ~ 5 at $f < 10^{-8}$Hz. This suppression factor decreases by increasing $\dot{m}$ and/or decreasing α. *About 10 individual sources are resolvable at 1 ns timing level, independently of the adopted disk model.*
- All the results shown here for the Tu-SA model, hold for every other MBHB population model we tested. There is a certain level of degeneracy between disk dynamics and MBHB mass function: the stochastic level given by a population of heavy binaries evolving by gas dynamics, can mimic that of a population of lighter binaries that are driven by GWs only. However, the variance of the signal would be much bigger in the former case, because the signal is produced by fewer massive sources.
- The detection of GWs emitted by MBHBs embedded in gaseous disks with high viscosity and accretion rate, may be very challenging for relatively short term PTA campaigns like the PPTA. In fact, we find that most of the 12 MBHB population models tested in SVV09 would *not* produce a stochastic signal detectable by the PPTA (i.e. the signal is a factor of three below the PPTA capabilities for the Tu-SA model). However, long term projects like the SKA, which aim to nanosecond sensitivities, are expected to be able to detect the GW signal, resolving a handful of individual sources with high significance.

A word of caution should be spent to stress the fact that our models are idealized in many ways. Firstly, we considered circular binaries only. If all the systems were significantly eccentric, then the overall signal would be modified by the multi-harmonic emission of each individual source. Moreover, we only considered radiatively efficient, geometrically thin, one-zone, stationary accretion disks. Further studies should examine the accuracy of these approximations for gas driven migration in circumbinary disks around MBHBs. Finally, it is likely that not all the binaries are gas driven on their way to the coalescence, and the efficiency of the disk-binary coupling may vary from merger to merger depending on the environmental conditions, which may modify the properties of the expected signal as well. Nonetheless, our calculations provide clear predictions for the possible attenuation of the stochastic GW background, which may be confirmed or discarded by ongoing and forthcoming pulsar timing arrays.

ACKNOWLEDGMENTS

A.S. is grateful to M. 'penguin' Giustini and V. Cappa for enlightening discussions. B.K. acknowledges support by NASA through Einstein Postdoctoral Fellowship grant number PF9-00063 awarded by the Chandra X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for NASA under contract NAS8-03060, and partial support by OTKA grant 68228.

REFERENCES

Amaro-Seoane P., Sesana A., Hoffman L., Benacquista M., Eichhorn C., Makino J., Spurzem R., 2009, arXiv:0910:1587
Armitage P. J., Natarajan, P., 2002, ApJ, 567, L9
Armitage P. J., Natarajan P., 2005, ApJ, 634, 921
Barnes J. E., 2002, MNRAS, 333, 481
Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307
Bender P. et al., 1998, *LISA*, Laser Interferometer Space Antenna for gravitational wave measurements: ESA Assessment Study Report
Bertone S., De Lucia G., Thomas P. A., 2007, MNRAS, 379, 1143
Bertotti B., Carr B. J., Rees M. J., 1983, MNRAS, 203, 945
Callegari S., Mayer L., Kazantzidis S., Colpi M., Governato F., Quinn T., Wadsley J., 2009, ApJ, 696, 89
Cuadra J., Armitage P. J., Alexander R. D., Begelman M. C., 2009, MNRAS, 393, 1423
Detweiler S., 1979, ApJ, 234, 1100
Dotti M., Colpi M., Haardt F., Mayer, L., 2007, MNRAS, 379, 956
Dubus G., Hameury J.-M., Lasota J.-P., 2001, A&A, 373, 251
Edwards R. T., Hobbs G. B., Manchester R. N., 2006, MNRAS, 372, 1549
Escala A., Larson R. B., Coppi P. S., Mardones D., 2004, ApJ, 607, 765
Escala A., Larson R. B., Coppi P. S., Mardones D., 2005, ApJ, 630, 152
Haehnelt M. G., 1994, MNRAS, 269, 199
Haiman Z., Kocsis B., Menou K., 2009, ApJ, 700, 1952 (HKM09)
Hellings R. W., Downs G. S., 1983, ApJ, 265, 39
Hirose S., Krolik J. H., Blaes O., 2009, ApJ, 691, 16
Hobbs G. et al., 2009, arXiv:0911.5206
Ivanov P. B., Papaloizou J. C. B., Polnarev A. G., 1999, MNRAS, 307, 79
Janssen G. H., Stappers B. W., Kramer M., Purver M., Jessner A., Cognard I., 2008, in "40 YEARS OF PULSARS: Millisecond Pulsars, Magnetars and More", AIP Conference Proceedings, 983, 633
Jaffe A. H., Backer D. C., 2003, ApJ, 583, 616
Jenet F. A., Hobbs G. B., Lee K. J., Manchester R. N., 2005, ApJ, 625, 123
Jenet R. et al., 2009, arXiv:0909.1058
King A. R., Pringle J. E., Livio M., 2007, MNRAS, 376, 1740
Kollmeier J. A. et al., 2006, ApJ, 648, 128
Koushiappas S. M., Zentner A. R., 2006, ApJ, 639, 7
Kocsis B., Haiman Z., Menou K., 2008, ApJ, 684, 870
Lazio J., 2009, arXiv:0910.0632
Lightman A. P., Eardley D. M., 1974, ApJ, 187, L1
Lodato G., Nayakshin S., King A. R., Pringle J. E., 2009, MNRAS, tmp.1156, arXiv:0906.0737
Malbon R. K., Baugh C. M., Frenk C. S., Lacey, C. G., 2007, MNRAS, 382, 1394
Manchester R. N., 2008, in "40 YEARS OF PULSARS: Millisecond Pulsars, Magnetars and More", AIP Conference Proceedings, 983, 584
Milosavljevic M., Merritt D., 2003, ApJ, 596, 860
Piran T., 1978, ApJ, 221, 652
Pessah M. E., Chan C-K., Psaltis D., 2007, ApJ, 668, 51L
Phinney E. S., 2001, arXiv:astro-ph/0108028
Sazhin M. V., 1978, Soviet Astron., 22, 36
Sesana A., Haardt F., Madau P., Volonteri, M., 2004, ApJ, 611, 623
Sesana A., Haardt F., Madau P., Volonteri, M., 2005, ApJ, 623, 23
Sesana A., Vecchio A., Colacino C. N., 2008, MNRAS, 390, 192 (SVC08)
Sesana A., Vecchio A., Volonteri M., 2009, MNRAS, 384, 2255 (SVV09)
Sesana A., Vecchio A., 2010, submitted to Phys. Rev. D.
Shakura N. I., Syunyaev R. A., 1973, A&A, 24, 337
Springel V. et al., 2005, Nature, 435, 629
Syer D., Clarke C. J., 1995, MNRAS, 277, 758
Thorne K. S., 1987, in 300 Years of Gravitation, ed. S. Hawking & W. Israel (Cambridge: Cambridge Univ. Press), 330
Trump J. R. et al., 2009, ApJ, 700, 49
Tundo E., Bernardi M., Hyde J. B., Sheth R. K., Pizzella A., 2007, ApJ, 663, 57
Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 599
Wyithe J. S. B., Loeb A., 2003, ApJ, 590, 691
Yoo J., Miralda-Escudé J., Weinberg D. H., Zheng Z., Morgan C. W., 2007, ApJ, 667, 813