The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays

A. Sesana,[1,2]★ A. Vecchio[1,3] and C. N. Colacino[1]

[1]*School of Physics and Astronomy, The University of Birmingham, Edgbaston, Birmingham B15 2TT*
[2]*Centre for Gravitational Wave Physics, Penn State University, University Park, PA 16802, USA*
[3]*Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA*

ABSTRACT

Massive black hole binary systems, with masses in the range $\sim 10^4$–$10^{10}\,\mathrm{M_\odot}$, are among the primary sources of gravitational waves in the frequency window $\sim 10^{-9}$–0.1 Hz. Pulsar Timing Arrays (PTAs) and the Laser Interferometer Space Antenna (*LISA*) are the observational means by which we will be able to observe gravitational radiation from these systems. We carry out a systematic study of the generation of the stochastic gravitational-wave background from the cosmic population of massive black hole binaries. We consider a wide variety of assembly scenarios and we estimate the range of signal strength in the frequency band accessible to PTAs. We show that regardless of the specific model of massive black hole binaries formation and evolution, the characteristic amplitude h_c of the gravitational-wave stochastic background at 10^{-8} Hz varies by less than a factor of 2. However, taking into account the uncertainties surrounding the actual key model parameters, the amplitude lies in the interval $h_c(f = 10^{-8}\,\mathrm{Hz}) \approx 5 \times 10^{-16}$–$8 \times 10^{-15}$. The most optimistic predictions place the signal level at a factor of ≈ 3 below the current sensitivity of PTAs, but within the detection range of the complete Parkes PTA for a wide variety of models, and of the future Square-Kilometer Array PTA for all the models considered here. We also show that at frequencies $\gtrsim 10^{-8}$ Hz, the frequency dependency of the generated background follows a power law significantly steeper than $h_c \propto f^{-2/3}$, which has been considered so far; the value of the spectral index depends on the actual assembly scenario and provides therefore an additional opportunity to extract astrophysical information about the cosmic population of massive black holes. Finally, we show that *LISA* observations of individual resolvable massive black hole binaries are complementary and orthogonal to PTA observations of a stochastic background from the whole population in the Universe. In fact, the detection of gravitational radiation in both frequency windows will enable us to fully characterize the cosmic history of massive black holes.

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

1 INTRODUCTION

Massive black hole (MBH) binary systems with masses in the range $\sim 10^4$–$10^{10}\,\mathrm{M_\odot}$ are amongst the primary candidate sources of gravitational waves (GWs) at approximately nHz – mHz frequencies (see e.g. Haehnelt 1994; Jaffe & Backer 2003; Wyithe & Loeb 2003; Sesana et al. 2004, 2005). MBHs are ubiquitous today in the nuclei of nearby galaxies (see e.g. Magorrian et al. 1998). If MBHs were also common in the past (as implied by the notion that many distant galaxies harbour active nuclei for a short period of their life, e.g. Haehnelt & Rees 1993), and if their host galaxies experi-

enced multiple mergers during their lifetime, as dictated by popular cold dark matter hierarchical cosmologies (e.g. White & Rees 1978; Peebles 1982), then massive black hole binaries (MBHBs) would inevitably form during cosmic history. Many different scenarios for MBH formation and evolution have been proposed (e.g. Volonteri, Haardt & Madau 2003; Koushiappas, Bullock & Dekel 2004; Begelman, Volonteri & Rees 2006; Lodato & Natarajan 2006): the predicted number density of MBHBs and their distribution as a function of mass and redshift vary considerably, especially at $z \gtrsim 2$, where current observational constraints are not tight, but all of them predict the existence of MBHBs whose GW emission could be observed with present day and future experiments.

GWs emitted by MBHBs during their whole coalescence span a frequency range that extends from the nHz to mHz region. The

★E-mail: alberto@gravity.psu.edu

frequency band $\sim 10^{-5}$–1 Hz will be probed by the *Laser Interferometer Space Antenna* (*LISA*, Bender et al. 1998); *LISA* will observe the last months-to-years of the inspiral of two MBHs followed by the catastrophic merger and ringdown phases. In this frequency region, MBHBs are very bright (in the GW band) resolved sources. The frequency window 10^{-9}–10^{-6} Hz is accessible by *ongoing* (e.g. the Parkes radio telescope, Manchester 2008) and planned [e.g. the Square-Kilometer Array (SKA), www.skatelescope.org] Pulsar Timing Array (PTA) experiments. The observable window is set at low frequencies by the span of the observations, typically lasting several years, and at high frequencies by the time interval between successive observations of the same pulsar, typically a week. At these frequencies, the most likely signal to be detected is a stochastic GW background generated by the incoherent superposition of radiation from the whole cosmic population of MBHBs,[1] although individual sources could also be observed (see e.g. Jenet et al. 2004; Jenet, Creighton & Lommen 2005a). Pulsar timing provides therefore a unique tool to study GWs at very low frequencies.

GWs affect the propagation of radio signals from the pulsar to the receiver on Earth (e.g. Sazhin 1978; Detweiler 1979; Bertotti, Carr & Rees 1983), producing a characteristic signature in the time of arrival (ToA) of radio pulses. In general, one computes the difference between the expected arrival time, according to a given model, and the actual arrival time (e.g. Hellings & Downs 1983; Jenet et al. 2005b); the residuals from this fit carry the physical information about the unmodelled effects in the pulse propagation, including possible GWs. Ongoing and future PTAs are reaching sensitivity levels that can lead to the detection of a stochastic background generated by MBHBs. It is therefore essential to understand the uncertainties affecting the predictions from different models and to rigorously characterize the statistical properties of the signal, which in turn affect the data analysis approach.

So far, the stochastic GW background from MBHBs in the frequency window $\sim 10^{-9}$–10^{-7} Hz has been treated as an isotropic, Gaussian and stationary signal described by a characteristic amplitude h_c that follows a power law $h_c \propto f^{-2/3}$ (e.g. Phinney 2001). Consequently, the relevant data analysis technique to search for such a signal consists in looking for statistically significant correlations in the ToA residuals of different pulsars. Jenet et al. (2006) placed the currently tightest constraint on the level of the GW background, and showed (Jenet et al. 2005b) that observations of 20 ms pulsars spanning 5 yr with a timing precision of ~ 100 ns would allow the detection of a GW background at the level predicted by several models of MBH assembly (Wyithe & Loeb 2003; Enoki et al. 2004; Sesana et al. 2004).

In this paper, we carry out a systematic study of the generation of a GW stochastic background from MBHBs in the frequency range accessible to PTAs by considering a wide variety of models of MBH assembly. They qualitatively encompass the whole range of scenarios currently proposed for MBH formation and quantitatively cover a very broad spectrum of predictions. The main results of our analysis can be summarized as follows (the reader who wants to skip all the details and just see the essential results is advised to

---

[1] Stochastic signals generated by the incoherent superposition of deterministic signals from astrophysical sources are usually referred to as *foregrounds* (especially in the *LISA* community) to distinguish them from the background of GWs produced in the very early Universe. Here, we stick to the naming convention of the radio pulsar community and we will refer to the stochastic signal from a cosmic population of MBHBs as *background*.

Table 1. Value of the parameters for the analytical expression of the characteristic amplitude of a GW stochastic background given by equation (14).

Model	h_0 ($\times 10^{-15}$)	f_0 ($\times 10^{-8}$ Hz)	γ
VHM	2.15	1.42	−1.09
VHMhopk	0.69	4.27	−1.08
KBD	0.67	5.24	−1.04
BVRhf	0.89	3.95	−1.11

refer to equation 14 – with parameter values provided in Table 1 and equations 45–47 – and Figs 9 and 13).

(i) We show that contrary to what has been assumed so far, the generated stochastic background does *not* follow a power law $h_c \propto f^{-2/3}$ across the whole frequency range $\sim 10^{-9}$–10^{-7} Hz, but the frequency dependence becomes steeper above $\approx 10^{-8}$ Hz – see equation (14), Table 1 and Fig. 9 – due to the intrinsic *discrete nature* of the sources.

(ii) The frequency at which there is a significant deviation from the behaviour $h_c \propto f^{-2/3}$ and the new spectral index both depend on the actual MBHBs cosmic history, see Table 1.

(iii) We estimate the strength of the stochastic GW background as a function of the adopted astrophysical model and the uncertainties surrounding the model's parameters, see equations (45)–(47) and Fig. 11. Regardless of the specific model of MBHB formation and evolution, the characteristic amplitude h_c of the GW stochastic background at 10^{-8} Hz varies by less than a factor of 2; however, taking into account the uncertainties surrounding the actual key model parameters, the amplitude varies by a factor of ≈ 20.

(iv) The most optimistic predictions place the signal level at a factor of ≈ 3 below the current sensitivity of PTAs, but future surveys are likely to detect the signal predicted by the whole range of astrophysical models considered here; they can potentially measure the spectral shape of the background in the frequency range 3×10^{-9}–3×10^{-8} Hz and therefore provide new information on the MBH population at low redshift, see Fig. 13.

(v) PTA observations of a stochastic background are complementary and 'orthogonal' to *LISA* observations of individual events; in fact, models of MBH assembly that generate large (by over an order of magnitude) fluctuations in the number of *LISA* events, can provide very similar levels of stochastic backgrounds in the PTA frequency range and vice versa; this is due to the fact that PTAs are especially sensitive to the low-redshift population of MBHBs, whereas *LISA* will be able to observe also very high-redshift sources, see Fig. 14.

(vi) As a byproduct of our analysis, we have compared the observed massive (luminous) galaxy merger rates at $z < 1$ derived by a number of authors with the merger rates inferred using the extended Press & Schechter (EPS) approach adopted in this paper, and find good agreement.

The paper is organized as follows. In Section 2, we summarize the theory of the generation of GWs from a population of MBHBs. In Section 3, we introduce four different MBH assembly scenarios that are representative of the formation and evolution of this class of sources, and evaluate the total GW signal using a semi-analytical approach and Monte Carlo methods. In Section 4, we derive a more accurate estimate of the GW stochastic background and introduce a simple mathematical model to account for the discrepancy between semi-analytical predictions and Monte Carlo simulations. Section 5

is devoted to quantifying the uncertainties in the strength of the background and in Section 6 we compare our results with those from previous works. Observational prospects are discussed in Section 7, and the main results and pointers to future work are summarized in Section 8.

2 GRAVITATIONAL WAVE BACKGROUNDS: KEY CONCEPTS

In this section, we review the key concepts and formulae regarding a GW stochastic background signal produced by the superposition of radiation from a large number of individual sources. For more details, we refer the reader to Phinney (2001), and references therein. Throughout the paper, unless differently specified, we adopt geometrical units in which $c = G = 1$.

A stochastic background is described in terms of the present-day GW energy density $\rho_{gw}(f)$ per logarithmic frequency interval normalized to the critical energy density ρ_c according to

$$\Omega_{gw}(f) = \frac{1}{\rho_c} \frac{d\rho_{gw}(f)}{d \ln f}. \tag{1}$$

Here, f is the *observed* frequency at the instrument. Let us now consider a population of GW events characterized by a comoving number density per unit redshift dn/dz; the events generate energy per logarithmic frequency interval $dE_{gw}/d\ln f_r$, where the energy is measured in the source rest frame at redshift z and $f_r = (1 + z) f$ is the *rest-frame* frequency. The energy spectrum $\Omega_{gw}(f)$ of the background produced by the superposition of the radiation from these cosmic events is therefore given by (Phinney 2001)

$$\frac{d\rho_{gw}(f)}{d \ln f} = \frac{\pi}{4} f^2 h_c^2(f) = \int_0^\infty dz \, \frac{dn}{dz} \frac{1}{1 + z} \frac{dE_{gw}}{d \ln f_r}\bigg|_{f_r = f(1+z)}, \tag{2}$$

where the factor of $1/(1 + z)$ accounts for the redshift of gravitons and $h_c(f)$ is the characteristic GW amplitude in a frequency interval of width $\sim f$. The interpretation of equation (2) is straightforward: the energy density per logarithmic frequency interval is equal to the integral over the cosmic history of the comoving number density of the sources, multiplied by the redshifted energy emitted by each source in the corresponding redshifted frequency range.

This general result can be specialized to the scenario considered in this paper: a stochastic background generated by a cosmic population of MBHBs during their slow adiabatic inspiral phase. As we consider radiation generated when the sources are far from the last stable orbit, it is sufficient to model gravitational radiation at the leading Newtonian quadrupole order. Under this assumption, for a source characterized by a chirp mass $\mathcal{M} = \mu^{3/5} M^{2/5}$ – here μ and M are the reduced and total mass, respectively – at orbital frequency $f_{orb} = f_r/2$, the energy emitted per logarithmic frequency interval is (Thorne 1987)

$$\frac{dE_{gw}}{d \ln f_r} = \frac{\pi^{2/3}}{3} \mathcal{M}^{5/3} f_r^{2/3}. \tag{3}$$

Such energy is radiated between a minimum frequency, which we can safely assume to be smaller than the lowest frequency ~ 1 nHz probed by PTAs, and the frequency corresponding to the last stable orbit $f_{LSO} \approx 4 \times 10^{-6}(1 + z)^{-1}(M/10^9 \, M_\odot)^{-1}$ Hz, which is larger than the highest frequency reached in PTA observations. As a consequence, we will ignore such limits in the rest of the paper.

The GW energy and the number density of events (i.e. mergers) throughout cosmic history depend on the MBHB masses; we therefore define

$$\frac{dn}{dz} = \int_0^\infty d\mathcal{M} \frac{d^2 n}{dz d\mathcal{M}}, \tag{4}$$

where $d^2 n / dz d\mathcal{M}$ is the comoving number density per unit redshift and chirp mass. Using equation (4), we can rearrange equation (2) and obtain

$$h_c^2(f) = \frac{4}{\pi f^2} \int_0^\infty dz \int_0^\infty d\mathcal{M} \frac{d^2 n}{dz d\mathcal{M}} \frac{1}{1 + z} \frac{dE_{gw}(\mathcal{M})}{d \ln f_r}. \tag{5}$$

The previous equation provides the characteristic amplitude as a function of the number density of mergers, and it allows us to compute the GW spectrum using equations (1) and (2).

It is also convenient to express the previous result as a function of the number of sources emitting at a given frequency. We first express the comoving number density of coalescences dn/dz as a function of $d^3N/dz d\mathcal{M} d\ln f_r$, the comoving number of binaries emitting in a given logarithmic frequency interval with chirp mass and redshift in the range $[\mathcal{M}, \mathcal{M} + d\mathcal{M}]$ and $[z, z + dz]$, respectively,

$$\frac{d^2 n}{dz d\mathcal{M}} = \frac{d^3 N}{dz d\mathcal{M} d\ln f_r} \frac{d\ln f_r}{dt_r} \frac{dt_r}{dz} \frac{dz}{dV_c}. \tag{6}$$

In the previous expression, dV_c is the comoving volume shell lying between z and $z + dz$, and t_r is time as measured in the source rest frame. A binary slowly inspirals by losing energy and angular momentum through GW emission; assuming circular orbits gravitational radiation is emitted at twice the orbital frequency – this is the dominating quadrupole order contribution to the total GW flux – whose sky- and polarization-averaged strain amplitude is given by (Thorne 1987)

$$h = \frac{8\pi^{2/3}}{10^{1/2}} \frac{\mathcal{M}^{5/3}}{d_L(z)} f_r^{2/3}, \tag{7}$$

where $d_L(z)$ is the luminosity distance to the source. The rate of change of the emission frequency is

$$\frac{df_r}{dt_r} = \frac{96}{5} \pi^{8/3} \mathcal{M}^{5/3} f_r^{11/3}. \tag{8}$$

Combining equations (3), (7) and (8), we obtain

$$\frac{dE_{gw}}{d\ln f_r} = \frac{dt_r}{d\ln f_r} \pi^2 d_L(z)^2 f_r^2 h^2. \tag{9}$$

Substituting equations (6) and (9) into equation (5), it yields

$$h_c^2(f) = \int_0^\infty dz \int_0^\infty d\mathcal{M} \frac{d^3 N}{dz d\mathcal{M} d\ln f_r} h^2(f_r). \tag{10}$$

The above equation is equivalent to equation (5), and has a simple interpretation: the observed characteristic-squared amplitude of the GW background is given by the integral over all the sources emitting in the frequency bin $d \ln f_r$ multiplied by the squared strain of each source.

It is straightforward to see from both equations (5) and (10) that the predicted characteristic amplitude scales as $\propto f^{-2/3}$, with a normalization that depends on the details of the MBHB population, and is usually represented as (e.g. Jenet et al. 2006)

$$h_c(f) = h_{1\,yr} \left(\frac{f}{yr^{-1}} \right)^{-2/3}, \tag{11}$$

where $h_{1\,yr}$ is a model-dependent constant.

3 SIGNAL GENERATION

In order to predict the spectrum of the stochastic background generated by MBHBs, one needs to compute either $d^2 n / dz d\mathcal{M}$ or $d^3 N / dz d\mathcal{M} d \ln f_r$, see equations (4) and (6), according to a given cosmological model of MBH assembly; the characteristic strain amplitude $h_c(f)$ and spectrum $\Omega_{gw}(f)$ can then be evaluated using

equation (5) or (10), and equation (1), respectively. In this section, we consider a wide range of MBHB formation scenarios – and compare them with models considered in earlier work, although this issue will be discussed in much greater detail in Section 6 – and estimate the overall GW signal $h_c(f)$ produced by such a population. This forms the basis for our predictions about the strength and spectral shape of the *stochastic background* signal in the frequency range probed by PTAs that will be discussed in detail in Section 4. We will also highlight some subtleties in the computation of the background that have been ignored so far and have considerable impact on the prediction of the signal at frequencies $\gtrsim 10^{-8}$ Hz.

3.1 Models of formation of massive black hole binary systems

The GW spectrum produced by populations of MBHBs has already been computed in a handful of studies, based however on *specific* cosmic histories of MBH formation. Early works (Rajagopal & Romani 1995, hereafter R&R, and Jaffe & Backer 2003, hereafter J&B) derived the MBHB coalescence rate from available observational constraints on the fraction of galaxies that form kinematically close pairs in the local Universe (Burkey et al. 1994; Colin, Schramann & Peimbert 1994; Carlberg et al. 2000), and then assigned to each galaxy a MBH according to a given MBH mass function. A series of subsequent papers (Wyithe & Loeb 2003; Enoki et al. 2004; Sesana et al. 2004) applied the EPS formalism (Press & Schechter 1974; Lacey & Cole 1993; Sheth & Tormen 1999) to the hierarchical assembly of dark matter haloes, using a range of prescriptions for the evolution of the population of MBHs residing in the halo centres. The halo hierarchy is followed both analytically (e.g. Wyithe & Loeb 2003), or by means of Monte Carlo realizations of the merger hierarchy (e.g. Sesana et al. 2004). The 'observational' and EPS approach yield different estimates for the amplitude of the expected signal that need to be reconciled, and we will return to this point in great detail in Sections 5 and 6. It is worth noting that, except for a qualitative discussion in Wyithe & Loeb (2003), no comparison regarding the different predicted levels of the GW spectrum has been carried out so far, nor the reasons of these differences investigated. Moreover, several results show a significant departure from the simple power-law behaviour $h_c \propto f^{-2/3}$ (R&R and J&B) and this point has been overlooked. Both these issues have a significant impact on the prospects of detection of a stochastic background with PTAs – and implications for analysis strategies – and will be discussed in detail.

One of the main goals of this paper is to provide a systematic exploration of different scenarios of MBH assembly, to quantify the range of signal strength and the consequences for observations with PTAs, and to explore the potential of PTAs to shed new light on the relevant epoch of cosmic history and the demographics of MBHs. We consider four classes of models of MBH assembly that qualitatively encompass the whole range of scenarios proposed so far and quantitatively cover a very broad spectrum of predictions for MBH formation (see Sesana, Volonteri & Haardt 2007a for full details): (i) the VHM model (Volonteri et al. 2003), (ii) the KBD model (Koushiappas et al. 2004), (iii) the BVRhf model (the 'high feedback' model described in Begelman et al. 2006) and (iv) the VHMhopk model. In these scenarios, seed black holes are massive ($M \sim 10^4\,M_\odot$) as in the case of KBD and BVRhf, or light ($M \sim 10^2\,M_\odot$), as prescribed by VHM; seed black holes are abundant (VHM, KBD) or just a few (BVRhf). The VHMhopk model assumes essentially the same assembly history of the VHM model, but with a somewhat different accretion prescription based on the fitting formula by Hopkins et al. (2005) (see Volonteri, Salvaterra

& Haardt 2006 for details). Each model is constructed tracing backwards the merger hierarchy of 220 dark matter haloes in the mass range 10^{11}–$10^{15}\,M_\odot$ up to $z = 20$ (Volonteri et al. 2003), then populating the haloes with seed black holes and following their evolution to the present time. For each of the 220 haloes, all the coalescence events happening during the cosmic history are collected. The outputs are then weighted using the EPS halo mass function and integrated over the observable volume shell at every redshift to obtain numerically the coalescence rate of MBHBs as a function of black hole masses and redshift (see e.g. fig. 1 in Sesana et al. 2004 and fig. 1 in Sesana et al. 2007a). In other words, the outcome of this procedure is the numerical distribution $d^3N/dz\,d\mathcal{M}\,dt$.

It is worth noting that in all the above models, the MBH population fits by construction *at every redshift* the relation, observed in local galaxies, connecting the central black hole mass M_{BH} and the 1D stellar bulge velocity dispersion σ (the so called M_{BH}–σ relation, e.g. Tremaine et al. 2002). Moreover they all include a gravitational recoil prescription appropriate for non-spinning black holes. It is now well established that the recoil magnitude strongly depends on the spins (both magnitude and relative geometry) of the two black holes (Herrmann et al. 2007); the MBH build-up through merger and accretion typically results in highly spinning MBHs, which on average are affected by kicks with higher recoil velocity (e.g. Schnittman & Buonanno 2007). The adopted prescriptions for both the GW recoil and the MBH mass function in the local Universe are likely to influence the expected GW background. In Section 5, we map these effects and other poorly known dynamical processes into a range of uncertainty of the strength of the stochastic background level.

3.2 Semi-analytical approach

The natural outcome of the Monte Carlo EPS models described in the previous section consists of a list of coalescence events that are weighted on the EPS mass function and integrated in volume to obtain the numerical distribution $d^3N/dz\,d\mathcal{M}\,dt$ of MBHB coalescences throughout the Universe. Starting from here, we easily calculate $d^3N/dz\,d\mathcal{M}\,d\ln f_r$ changing variable from t to f according to equation (8) and we finally derive the characteristic amplitude h_c of gravitational radiation generated from the whole population using equation (10). The very same result is equivalently obtained by considering the number density of coalescing binaries using equation (6) and substituting equation (3) into equation (5). This is the approach used by several authors in the past, though only for selected models. The resulting estimated characteristic amplitude h_c for each of the four models VHM, VHMhopk, BVRhf and KBD is shown in Fig. 1. Not surprisingly, the characteristic amplitude scales always as $h_c \propto f^{-2/3}$ – it must be so by construction – and the level of the signal $h_{1\,yr} \approx 10^{-15}$ is similar for all of them; one can expect a much larger spread of values due to the uncertainty on specific model parameters, which we will discuss in Section 5, rather than the specific prescription for MBH assembly.

The fact that in the window 10^{-9}–10^{-7} Hz the level of the signal does not strongly depend on the specific assembly scenario is easily explainable: the major contributors to very low frequency GWs are massive ($\mathcal{M} > 10^8\,M_\odot$) and nearby ($z < 2$) binary systems, whose population is fairly well constrained by present observations. Whether these MBHs are the end-product of light or massive seeds has only modest effects on the signal in the PTA frequency window. We recall, however, that the models considered here assume a MBH mass function based on the M_{BH}–σ relation. We will see in Section 5
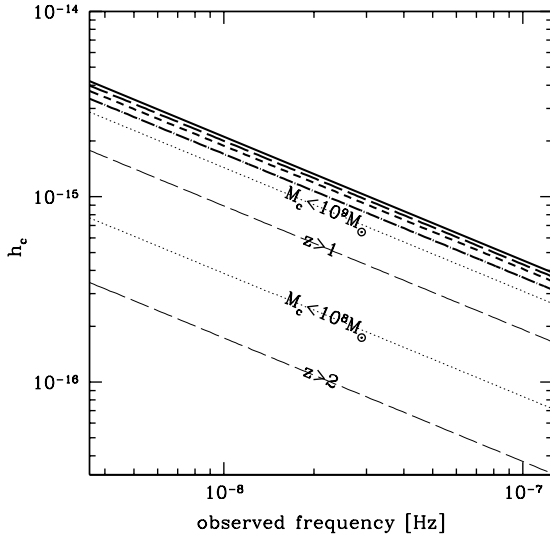
Figure 1. The characteristic amplitude h_c, see equation (10), generated by a population of MBHBs in the frequency range accessible with Pulsar Timing Arrays for the assembly models discussed in the text. The thick lines show the total signal for the VHM (solid line), the VHMhopk (dot–dashed line), the KBD (short-dashed line) and the BVRhf (long-dashed line) models. The thin lines show the contributions from selected portions of the binary population predicted by the VHM model: the contribution to the signal considering different redshift cuts (thin dashed lines) and different cuts in chirp mass M (thin dotted lines).

that different assumptions for the MBH–stellar bulge relation have a significant (a factor of $\gtrsim 2$) impact on the overall level of the signal.

3.3 Monte Carlo approach

A different procedure to estimate the signal produced by a population of MBHBs is to directly build such a population through a Monte Carlo approach and compute the overall signal in the relevant frequency range by adding the contributions from each individual binary, using the same merger trees. In observations with PTAs, radio pulsars are monitored weekly for periods of years. The relevant frequency band is therefore between $1/T$ – where T is the total observation time – and the Nyquist frequency $1/(2\Delta t)$ – where Δt is the time between two adjacent observations, corresponding to 5×10^{-9}–10^{-7} Hz. The frequency resolution bin is $1/T$, and a source can be considered effectively monochromatic if during the observation time the frequency shift $T\mathrm{d}f/\mathrm{d}t$ (due to loss of energy through GW emission) is less than $1/T$. It is simple to verify through equation (8) that this is indeed the case for the overwhelming majority of sources in the relevant mass range for $f \lesssim 10^{-7}$ Hz.

Starting from a model-dependent prescription of the MBH assembly history, one generates a population of MBHBs using a Monte Carlo sampling of the trivariate distribution $\mathrm{d}^3 N/\mathrm{d}z\mathrm{d}\mathcal{M}\mathrm{d}\ln f_r$; the total signal h_c is then computed by adding all the contributions in every frequency interval $1/T$. A key difference with respect to the semi-analytical approach described in Section 3.2 is that here we do not directly perform the integral in equation (10), but, given the Monte Carlo sample of the emitting sources we simply sum the contributions of each individual one. It is also obvious – but this point is essential to explain discrepancies with the semi-analytical approach – that by default one takes into account the *discrete nature* of the GW emitters.
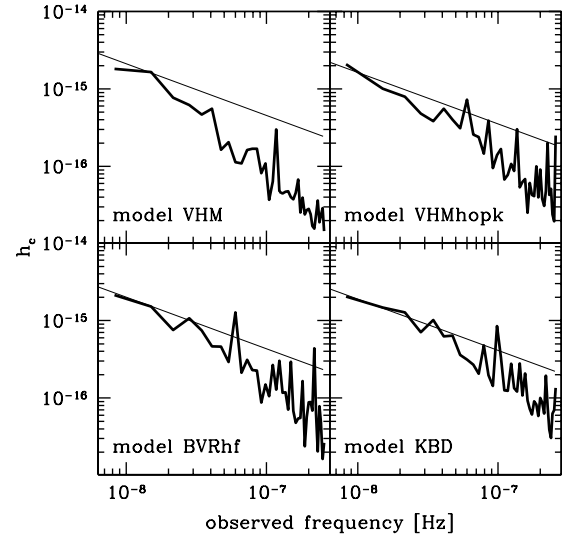
Figure 2. The total contribution to the characteristic amplitude h_c of the GW signal from a population of MBHBs in the frequency range accessible to Pulsar Timing Arrays. In each panel, the thick line shows h_c produced in a specific Monte Carlo realization and compares it to the prediction yielded by the semi-analytical approach (thin line). The observation time is $T = 5$ yr.
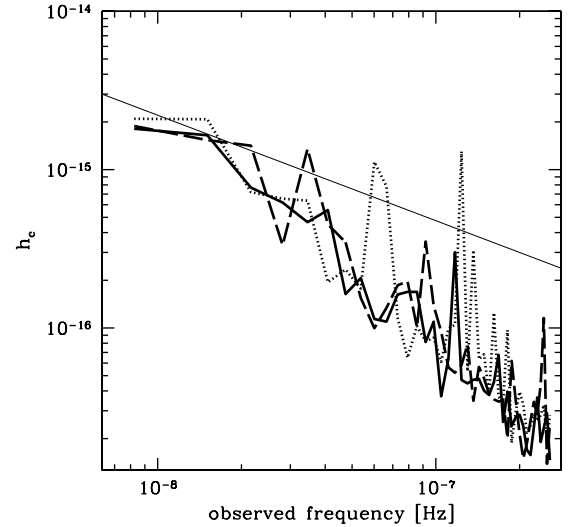
Figure 3. Same as Fig. 2, but now three different Monte Carlo realizations of the GW signal expected from the *same* MBH assembly model (VHM) are shown as a function of frequency (thick solid, dotted and dashed lines). The thin solid line is the prediction obtained using the semi-analytical approach for the same model. The observation time is set to 5 yr.

The results of this approach are reported in Fig. 2 for the different models of MBH assembly discussed in Section 3.1. Each panel shows h_c obtained through both the semi-analytical approach and the Monte Carlo realization in a frequency resolution bin $\Delta f = 1/T$, where T is set to 5 yr. In Fig. 3, we focus on a specific MBH assembly model (VHM) and three different Monte Carlo realizations of the GW signal. Some distinctive features are immediately clear: (i) the Monte Carlo and semi-analytical approach provide estimates of the total GW signal that are very well in agreement at frequencies below $\approx 10^{-8}$ Hz and (ii) at frequencies $\gtrsim 10^{-8}$ Hz, the total contribution to h_c produced through the Monte Carlo procedure falls clearly below the semi-analytical prediction $h_c \propto f^{-2/3}$

suggesting a steepening of the spectrum in this frequency range. We would like to highlight that such behaviour is also present in the results reported in earlier works (e.g. R&R and J&B) that employ an analogous Monte Carlo approach, but consider different assembly scenarios. Such discrepancy has never been discussed so far but is actually symptomatic of an important conceptual difference between the two approaches, with direct observational implications. We will devote the next section to investigate and explain this difference in detail. Before proceeding, we would also like to stress that so far we have referred to h_c as the characteristic amplitude of the *total GW signal* rather than the *GW stochastic background* from a MBHB population; once more this is a crucial difference that will become clear in the next section.

4 STOCHASTIC BACKGROUND

In this section, we investigate the origin of the discrepancy between the semi-analytical and Monte Carlo estimate of the total signal produced by a population of MBHBs and provide a new estimate for the characteristic amplitude, or equivalently the spectral content of the relevant *stochastic background* valid over the whole frequency range of PTA observations. In Section 4.1, we study the details of the source population (MBHB distribution in redshift, mass and frequency) that make up the signal, and their consequences. In Section 4.2, we consider a heuristic model that explains the inconsistency between the two approaches, and provides a simple analytical expression for the stochastic background as a function of three parameters. In Section 4.3, we present an analytical approximation to the Monte Carlo results to further clarify the key conceptual issues; readers not interested in the mathematical details can skip this section without loss of any crucial information for the rest of the paper. We conclude by summarizing the subtleties related to the computation of the overall signal and the pitfalls of the naïve semi-analytical approach in Section 4.4. We concentrate on the VHM model to discuss the fundamental issues; the results are qualitatively identical for all the other astrophysical scenarios, with differences concerning only the numerical values of the relevant quantities that are presented in Table 1.

4.1 Origin of the discrepancy

We start by studying the contributions to $h_c(f)$ at different frequencies as a function of z and $\mathcal{M}$. We consider the contributions to the signal centred at the two fiducial frequencies $f = 8 \times 10^{-9}$ Hz (close to the minimum accessible frequency for a $T = 5$ yr observation) and 10^{-7} Hz, with width $\Delta f = 1/T \simeq 6 \times 10^{-9}$ Hz, from sources within the shell z and $z + \Delta z$, where $\Delta z = 0.05$. Fig. 4 clearly shows that the Monte Carlo and semi-analytical approach provide the same estimate of the total number of sources per redshift interval that contribute to the signal. The estimate of the signal h_c is again consistent between the two approaches at $f = 8 \times 10^{-9}$ Hz; however, at $f = 10^{-7}$ Hz, the estimate of h_c inferred using the Monte Carlo sampling of the population is significantly smaller than that computed semi-analytically. This behaviour can be understood by studying the contribution to the signal by the binaries in the population as a function of $\mathcal{M}$; this is shown in Fig. 5. We first see that the average number $\langle N_{bin} \rangle$ of frequency resolution bins of width $1/T$ spanned by each binary during the observation is always smaller than one; this ensures that we can safely treat all the binaries contributing to the background as stationary sources. Further, note that the main contribution to the characteristic amplitude h_c is generated by the high-mass tail of the $\mathcal{M}$-distribution. At frequencies $\lesssim 10^{-8}$ (upper
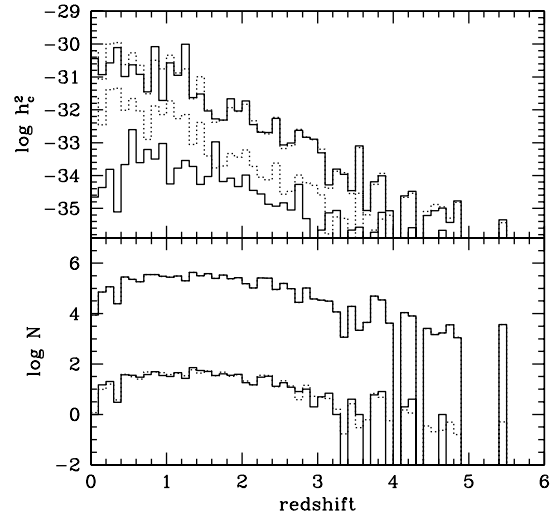
Figure 4. Contribution of different redshift intervals to the build-up of the GW signal at two different frequencies, $f = 8 \times 10^{-9}$ and 10^{-7} Hz, as computed using the two different approaches described in the text: Monte Carlo sampling (solid lines) and semi-analytical approach (dotted lines). In each panel, the upper histograms refer to $f = 8 \times 10^{-9}$ Hz and the lower histograms refer to $f = 10^{-7}$ Hz. These results refer to the VHM model.
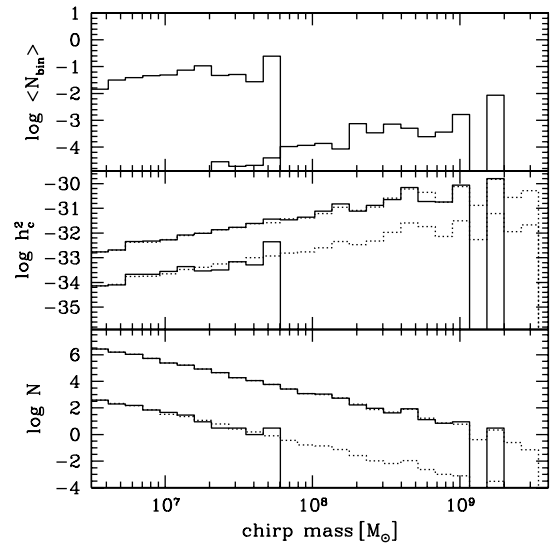
Figure 5. Same as Fig. 4, but now considering the contribution from different intervals in chirp mass. In this case, we also show (top panel) the mean number of frequency bins $\langle N_{bin} \rangle$ spanned by a binary during 5 yr, as a function of $\mathcal{M}$. The line style is the same as in Fig. 4.

curves in each panel of Fig. 5), the number of sources contributing to the signal is $\sim 10^2$–10^6 (depending on the mass range) and the Monte Carlo and semi-analytical approach yield the same estimate for h_c. However, at frequencies $\gtrsim$ a few $\times 10^{-8}$ Hz (bottom curves in each panel of Fig. 5), the overwhelming majority of the radiation inferred using the semi-analytical approach is actually produced by less than one source with $\mathcal{M} \gtrsim 10^8$ M$_\odot$. This is clearly an artefact and the key point to understand the difference of the results: a Monte Carlo sampling of the population returns correctly a discrete number of sources – in this case either one or none – and not fractions of a binary; therefore, it cannot possibly (and it should not!) reproduce the semi-analytical result that contains spurious contributions from
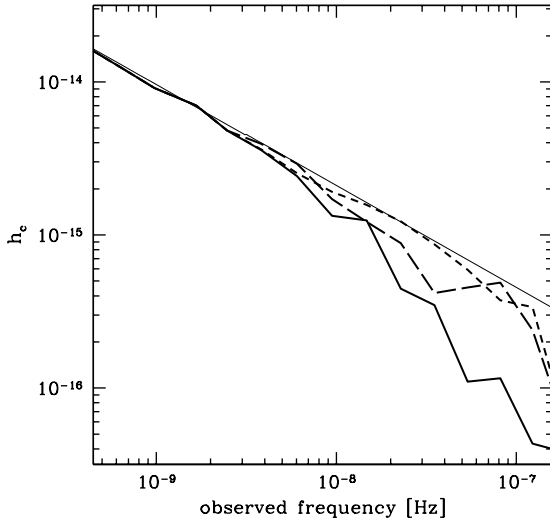
Figure 6. The very low frequency characteristic amplitude h_c of the GW signal predicted by the VHM model. The thin and thick solid lines show h_c computed using the semi-analytical and Monte Carlo approach, respectively. The thick long- and short-dashed lines show the signal from the Monte Carlo realizations obtained by *artificially* multiplying the number of the contributing sources by a factor of 10 and 100, respectively, and dividing the contribution of each sources to h_c by the same factor.

'fractional sources' that are clearly not present; the actual signal is significantly smaller than the one computed semi-analytically.

To cross-check the discrepancy of the two results, we can *artificially* increase the number of sources present in the sample, but divide by the appropriate factor the contribution of each binary to h_c. The semi-analytical estimate is clearly unaffected and produces the familiar $h_c \propto f^{-2/3}$ dependency over the whole frequency band; however, as the number of sources increases, the Monte Carlo result matches progressively better across the entire frequency range the previous result. This is reported in Fig. 6: the number of sources in (a given model of) the Universe is fixed, and the actual total GW signal is far from resembling the simple $f^{-2/3}$ power law over the whole frequency range.

The semi-analytical approach, that treats the distributions of sources as *continuous*, leads to a significant overestimate of the strength of the background in an important frequency range for PTA observations. In order to derive a correct prediction of the level of the GW stochastic backgrounds, it is essential that the *discrete* nature of the sources is consistently taken into account. To reiterate this point, when one evaluates the coalescence rate of MBHBs (e.g. using merger tree models), one performs different realizations of the assembly of individual haloes, and then weights each halo realization over the EPS halo mass function, i.e. $\ll 1$. As a consequence, the rate of massive coalescing binaries at low redshift could be, for example, $\sim 10^{-5}$ yr^{-1}. When this rate is used to compute the number of binaries contributing to a certain frequency interval (see equation 6), one finds a number of sources $\ll 1$ for $f \gtrsim 10^{-8}$. As a matter of fact, one performs an idealized average realization of the Universe. In practice, the same average is done when extrapolating the coalescence rate from the observation of the fraction of close galaxy pairs. Once more, by dividing the number of close pairs by the typical merging time, one obtains rates that are $\ll 1$ yr^{-1}; the associated number of emitting sources is again $\ll 1$ (and for consistency, it could not be otherwise). To summarize, the discrete nature of sources is critical; semi-analytical evaluations of the background

(that treat each distribution function as continuous) lead to spurious predictions in the observational window above $\sim 10^{-8}$ Hz.

4.2 Phenomenological model

It is useful to provide a parametrized expression of h_c to replace the incorrect (in parts of the frequency spectrum) simple dependency $h_c \propto f^{-2/3}$. We start by providing a simple way to quantify the spurious additional contribution to h_c given by semi-analytical approaches.

Let us consider a frequency $\tilde{f}$, and integrate the total number of sources emitting at $f > \tilde{f}$ in a given mass range $[\tilde{\mathcal{M}}, +\infty]$. For any given $\tilde{f}$, we can always find a $\tilde{\mathcal{M}}$ such that

$$\int_{\tilde{\mathcal{M}}}^{\infty} d\mathcal{M} \int_{\tilde{f}}^{f_{LSO}} df \, \frac{d^2N}{d\mathcal{M}df} = 1. \qquad (12)$$

This means that all the signal due to sources with $\mathcal{M} > \tilde{\mathcal{M}}$ at $f > \tilde{f}$ is generated by less than one source, and it is therefore not present in the radiation produced by an actual population of MBHBs. An example is illustrated in Fig. 7. If we consider for example $\tilde{f} = 10^{-7}$ Hz, the implicit solution of equation (12) is $\tilde{\mathcal{M}} \simeq 1.5 \times 10^8 \, M_\odot$; GWs due to MBHBs with heavier chirp mass are generated by less than one source, and so they do not contribute to the signal.

What we have computed so far is the *total contribution* to the characteristic amplitude $h_c(f)$ as a function of frequency from the whole population of MBHBs in the Universe (for different models). However, here we are interested in estimating the *stochastic background* from the population. Whether the superposition of many deterministic signals should be effectively considered a stochastic signal depends on the frequency range and the observation time. Here, we will consider the usual (simplified) criterion that the signal is stochastic if the number of sources whose radiation contributes to the frequency bin of width $1/T$ centred at f are $\gg 1$; the stochastic level of the signal is then the amplitude of the sum of the individual contributions. Note therefore that Figs 2 and 3 do not show the
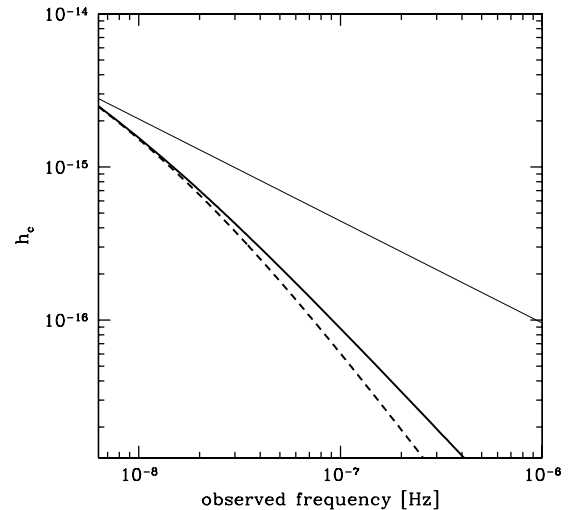

Figure 7. Signal spectrum in the VHM model. The characteristic strain is plotted versus the observed frequency. The thin solid line shows the naïve semi-analytical estimate of the amplitude from equation (5) or (10), the thick solid line is typical level of the 'real' signal once the exceeding spurious contribution is subtracted according to equation (12) and the thick dashed line represents the level at which the signal can be considered a stochastic background (assuming a 5 yr observation) according to the subtraction given by equation (13).
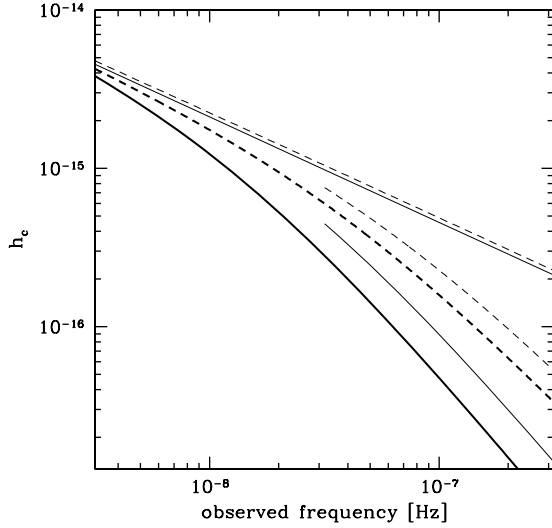
Figure 8. Fits to the stochastic background assuming 1 (thin lines) and 10 (thick lines) yr of observations. The $f^{-2/3}$ straight lines represent the naïve semi-analytical background predictions. The solid lines are for the VHM model, while the dashed lines are for the KBD model.

GW background contribution, but the total signal from the whole population of MBHBs. To evaluate the GW stochastic contribution, we can follow the same approach that has led us to equation (12), by simply replacing the range of integration over frequency with $[\tilde{f}, \tilde{f} + 1/T]$; we can therefore find a value $\tilde{\mathcal{M}}$ such that

$$\int_{\tilde{\mathcal{M}}}^{\infty} \mathrm{d}\mathcal{M} \int_{\tilde{f}}^{\tilde{f}+1/T} \mathrm{d}f \, \frac{d^2N}{\mathrm{d}\mathcal{M}\mathrm{d}f} = 1. \tag{13}$$

The integral above identifies the sources in the population that do not contribute to the stochastic background. The result depends on the observational time, since the longer the observation, the narrower the frequency bin and, accordingly, the lower the level of signal contribution that can be considered stochastic. We note, however, that the $\mathcal{M}$-distribution of sources is typically a steep power law (cf. Fig. 5). If we halve the frequency bin by doubling T, $\tilde{\mathcal{M}}$ in equation (13) would be only slightly reduced, and the level of stochasticity would change by less than 30 per cent. We find that the stochastic background does change by a factor of $\lesssim 2$ for observational times $1 \lesssim T \lesssim 10$ yrs, as shown in Fig. 8. Moreover, for typical observation times of several years, the number of sources radiating in the frequency interval $[\tilde{f}, \tilde{f} + 1/T]$ is of the same order of those radiating in the interval $[\tilde{f}, f_{\mathrm{LSO}}]$ [this is a consequence of equation (8) that implies that the number of sources radiating at a given frequency f is proportional to $f^{-11/3}$]; as a consequence, the amplitude of the stochastic signal is only slightly lower than the one estimated following the condition given in equation (12). Fig. 7 shows this effect; this result is consistent with the outcome of the study of the total number of sources that contribute to the total radiation from a population of MBHBs (at selected frequencies and in a range of either redshift or chirp mass) that is shown in the bottom panel of Figs 4 and 5.

Using the $\mathcal{M}$-distributions predicted by the merger trees (as those shown in the lower panel of Fig. 5), we have calculated $\tilde{\mathcal{M}}$ as a function of $\tilde{f}$ according to both equations (12) and (13). Then, we have subtracted at every frequency the portion of the signal due to the sources with $\mathcal{M} > \tilde{\mathcal{M}}$. Best fits to the residual signals obtained in this way are plotted in Fig. 7 for the VHM model; all the other models show a similar behaviour. The signal deviates significantly

from a simple $f^{-2/3}$ power law for $f \gtrsim 10^{-8}$ Hz. The GW stochastic background is well fitted by a function of the form

$$h_{\mathrm{c}}(f) = h_0 \left(\frac{f}{f_0}\right)^{-2/3} \left(1 + \frac{f}{f_0}\right)^{\gamma}. \tag{14}$$

The three parameters h_0, f_0 and γ are given in Table 1 for the four fiducial MBHB assembly scenarios considered in this paper. In general, the slope of the stochastic contribution to h_{c} starts to deviate from $-2/3$ at around 10^{-8} Hz, and becomes as steep as ≈ -1.5. In the remainder of the paper, we will use equation (14) and Table 1 as estimates of the characteristic amplitude of the GW stochastic background produced by MBHB populations. The astrophysical uncertainties affecting the fit parameters will be discussed in Section 5, and estimated ranges for h_0, f_0 and γ are given in equations (45)–(47). In the next section, we provide a different derivation of the main results that we have just presented.

4.3 The discrepancy revisited

The main result reported in equation (14) has been obtained by carrying out Monte Carlo sampling of the assembly history of MBHBs and then by fitting the numerical result. It is, however, instructive to derive the same result using a simple analytical model that relies only on general assumptions: the mass function of black holes, the GW quadrupole formula and the discreteness of sources.

Let us consider the probability density function of sources per unit chirp mass $p(\mathcal{M})$, so that

$$\int_0^{\infty} p(\mathcal{M}) \, \mathrm{d}\mathcal{M} = 1. \tag{15}$$

We sample now $p(\mathcal{M})$ with N objects (sources). The total number of objects N is model dependent, but we keep it general for the time being; we can define $\tilde{\mathcal{M}}(N)$ such that

$$\int_{\tilde{\mathcal{M}}}^{\infty} p(\mathcal{M}) \, \mathrm{d}\mathcal{M} = \frac{1}{N}. \tag{16}$$

For any function h – this will be the GW strain – we can always define its value weighted by the probability distribution $p(\mathcal{M})$ as

$$\bar{h}^2 = \int_0^{\infty} p(\mathcal{M}) h^2(\mathcal{M}) \, d\mathcal{M}. \tag{17}$$

For a fixed value of N, we define the following quantity:

$$Z \equiv \frac{\int_{\tilde{\mathcal{M}}}^{\infty} p(\mathcal{M}) h^2(\mathcal{M}) \, \mathrm{d}\mathcal{M}}{\int_0^{\infty} p(\mathcal{M}) h^2(\mathcal{M}) \, \mathrm{d}\mathcal{M}}, \tag{18}$$

Z is the fraction of h^2 whose contribution comes from less than one source, and is therefore not actually present. The overall value of h sampled with N objects is then simply

$$h_{\mathrm{eff}} = \bar{h}\sqrt{1 - Z}. \tag{19}$$

We can apply this general framework, and, in particular, equation (19) to the problem of estimating the effective stochastic background from a population of sources. To do this, we need to evaluate, at any given frequency f, the probability density function per unit chirp mass and unit frequency $p(\mathcal{M}, f)$ and the number of sources $N(f)$ given by the distribution of sources $d^2N/\mathrm{d}\mathcal{M}\mathrm{d}f$ emitting in a frequency interval of width $\Delta f = 1/T$ centred on f. First, we note that $p(\mathcal{M}, f) = p(\mathcal{M})$, i.e. the probability distribution does not depend on f. Ignoring subtleties related to the redshift, we can indeed write

$$\frac{d^2N}{\mathrm{d}\mathcal{M}\mathrm{d}f} = \frac{d^2N}{\mathrm{d}\mathcal{M}\mathrm{d}t} \frac{\mathrm{d}t}{\mathrm{d}f}$$
$$= \frac{5\mathcal{M}^{-5/3}}{96\pi^{8/3}} \frac{d^2N}{\mathrm{d}\mathcal{M}\mathrm{d}t} f^{-11/3}, \tag{20}$$

where the second equality comes from equation (8); the merger rate $d^2N/d\mathcal{M}dt$ is independent of f. The frequency dependence factorizes and we obtain

$$
\begin{aligned}
p(\mathcal{M}, f) &= \frac{\int_{f-\Delta f/2}^{f+\Delta f/2} df' \frac{d^2N}{d\mathcal{M}df'}}{\int_0^\infty d\mathcal{M} \int_{f-\Delta f/2}^{f+\Delta f/2} df' \frac{d^2N}{d\mathcal{M}df'}} \\
&= \frac{\mathcal{M}^{-5/3} \frac{d^2N}{d\mathcal{M}dt}}{\int_0^\infty \mathcal{M}^{-5/3} \frac{d^2N}{d\mathcal{M}dt} d\mathcal{M}} = p(\mathcal{M}).
\end{aligned}
\tag{21}
$$

In most of the models of MBH assembly (e.g. Volonteri et al. 2003), $d^2N/d\mathcal{M}dt$ is a function well approximated by a single power law $\propto \mathcal{M}^{-\beta}$ with $1.5 \lesssim \beta \lesssim 2$; the chirp mass probability density function, equation (21), is therefore well described by the following steep power law

$$
p(\mathcal{M}) = \begin{cases} A\mathcal{M}^{-\alpha} & \mathcal{M}_{\min} \le \mathcal{M} \le \mathcal{M}_{\max} \\ 0 & \text{elsewhere} \end{cases},
\tag{22}
$$

where $3 \lesssim \alpha \lesssim 3.5$, $\mathcal{M}_{\min}$ is of the order of the first MBH seed mass ($\sim 10^2 \, M_\odot$ for the VHM and the VHMhopk models; $\sim 10^4 \, M_\odot$ for the BVRhf and the KBD models) and $\mathcal{M}_{\max} \simeq 1–3 \times 10^9 \, M_\odot$. The normalization constant A is set by equation (15). The quantity $N(f)$ on the right-hand side of equation (16), however, depends on frequency. Using the quadrupole formula, see equation (8), the number of sources per frequency bin $\Delta f = 1/T$ is given by

$$
N(f, \Delta f) = N_0 \left(\frac{f}{f_0}\right)^{-11/3},
\tag{23}
$$

where f_0 is a reference frequency, and $N_0 = N(f_0, \Delta f)$ is the number of sources emitting in the frequency bin Δf centred on f_0, namely

$$
N_0 = \int_0^\infty d\mathcal{M} \int_{f_0-\Delta f/2}^{f_0+\Delta f/2} df \frac{d^2N}{d\mathcal{M}df}
\tag{24}
$$

that depends on the specific MBH assembly scenario. Substituting equations (22) and (23) into equation (16), we find

$$
A \int_{\tilde{\mathcal{M}}(f)}^{\mathcal{M}_{\max}} \mathcal{M}^{-\alpha} d\mathcal{M} = \frac{1}{N_0} \left(\frac{f}{f_0}\right)^{11/3}.
\tag{25}
$$

Solving the previous equation for $\tilde{\mathcal{M}}$ yields

$$
\tilde{\mathcal{M}}(f) = \mathcal{M}_{\max} \left[1 + \frac{\alpha-1}{A \, N_0 \mathcal{M}_{\max}^{1-\alpha}} \left(\frac{f}{f_0}\right)^{11/3}\right]^{1/(1-\alpha)}.
\tag{26}
$$

The previous equation is clearly valid only if $\alpha \ne 1$, which is indeed the case considered here (cf. the discussion following equation 22). Using the fact that $h_c^2 \propto \mathcal{M}^{10/3}$ and substituting $p(\mathcal{M})$ from equation (22) into equation (18), the fraction of signal that is not present due to the discrete nature of sources is

$$
\begin{aligned}
Z(f) &= \frac{\int_{\tilde{\mathcal{M}}(f)}^{\mathcal{M}_{\max}} \mathcal{M}^{10/3-\alpha} d\mathcal{M}}{\int_{\mathcal{M}_{\min}}^{\mathcal{M}_{\max}} \mathcal{M}^{10/3-\alpha} d\mathcal{M}} \\
&\simeq 1 - \left[\frac{\tilde{\mathcal{M}}(f)}{\mathcal{M}_{\max}}\right]^{13/3-\alpha}.
\end{aligned}
\tag{27}
$$

Here, we have used the fact that for GWs from MBHBs $\tilde{\mathcal{M}}(f) \gg \mathcal{M}_{\min}$. The value of $\tilde{\mathcal{M}}(f)$ depends on the frequency. Looking at equation (26), one can identify a critical value of the frequency $\bar{f}$ – for the merger trees used in this paper $\bar{f} \sim 10^{-8}$ Hz – that separates two different regimes. In the frequency region $f \lesssim \bar{f}$, the signal is characterized by $\tilde{\mathcal{M}} \approx \mathcal{M}_{\max}$ and therefore $Z(f) \ll 1$: effectively, the overwhelming majority of the sources equally contribute to the stochastic background. On the other hand, for $f > \bar{f}$ we obtain

$$
\tilde{\mathcal{M}}(f) \sim \mathcal{M}_{\max} \left(\frac{f}{\bar{f}}\right)^{11/(3-3\alpha)} \ll \mathcal{M}_{\max};
\tag{28}
$$

this implies that at higher frequencies $Z(f) \approx 1$ and the vast majority of the signal are due to less than one source, which clearly has no physical meaning; indeed Monte Carlo realizations of the background do not contain such a contribution. If we now insert equation (28) into equation (27) and use equation (19), we find that the level of stochastic background is

$$
h_{\text{eff}}(f) = \bar{h}(f)\sqrt{1 - Z(f)} \propto f^{-2/3-11/[6(1-\alpha)]};
\tag{29}
$$

for $3 \lesssim \alpha \lesssim 3.5$, the previous expression yields $h_{\text{eff}}(f) \propto f^{-1.6} - f^{-1.4}$, which is in good agreement with equation (14) for $h_c(f)$, that was obtained by fitting the Monte Carlo realizations.

The above result depends on the width of the frequency bin because it describes the level at which the signal can be considered stochastic. However, the discrepancy between the overall signal from a MBHB population computed using the semi-analytical and the Monte Carlo approach that we have discussed in the previous sections and is summarized in equation (12) depends only on the details of the specific population and not on the observational time. We can easily derive this result using the mathematical framework developed in this section; we therefore compute the effective strength of the total signal by simply considering, at any given frequency, all the sources emitting at that frequency or above. The probability distribution given by equation (22) does not change, but equation (23) needs to be replaced by

$$
N(f) = N_0' \left(\frac{f}{f_0}\right)^{-8/3},
\tag{30}
$$

where

$$
N_0' = \int_0^{\mathcal{M}_{\max}} d\mathcal{M} \int_{f_0}^{f_{\text{LSO}}} df \frac{d^2N}{d\mathcal{M}df}.
\tag{31}
$$

Using equation (30), we are considering as sampling sources all the binaries contributing to the signal above the frequency f, and the formalism becomes independent of the frequency bin. Replacing equation (23) with equation (30) and re-evaluating equations (25)–(27), the quantity $Z(f)$ is now interpreted as the spurious total signal excess at the frequency f, according to equation (12). In this case $\tilde{\mathcal{M}} \propto f^{8/(3-3\alpha)}$ for $f > \bar{f}$. Inserting this dependence into equation (27) and substituting into equation (19), we find the actual total signal level to be $h_{\text{eff}}(f) \propto f^{-1.35} - f^{-1.2}$, for $3 \lesssim \alpha \lesssim 3.5$.

4.4 Summary of the results

A summary of the results is shown in Fig. 9 for all the four reference models. As expected, the two curves representing the level of the actual total signal and the stochastic background are quite similar in the frequency range of interest; this is due to the fact that $dN/df \propto f^{-11/3}$. In fact, as we have discussed in Section 4.2, the number of sources in a frequency bin around f is of the order of the number of sources emitting in the interval $[f, +\infty]$. The slopes of the curves lie in the expected range: $f^{-1.6}$–$f^{-1.4}$ for the stochastic background, and $f^{-1.4}$–$f^{-1.2}$ for the total signal. The agreement among the Monte Carlo simulated background, the fitting formula given by equation (14) with parameter values shown in Table 1, and the outcome of the model presented in Section 4.3 is good. In conclusion, the expected signal consists of a stochastic background well described by a superposition of two power laws (equation 14), on top of which individual outliers are present due to rare, nearby bright sources. The slopes of the power laws are $-2/3$ at low frequencies, and ≈ -1.5 at high frequencies, with a knee around $\sim 10^{-8}$ Hz. Both the knee frequency and the index of the slope at high frequency depend on the details of the $\mathcal{M}$-distribution predicted by the MBH
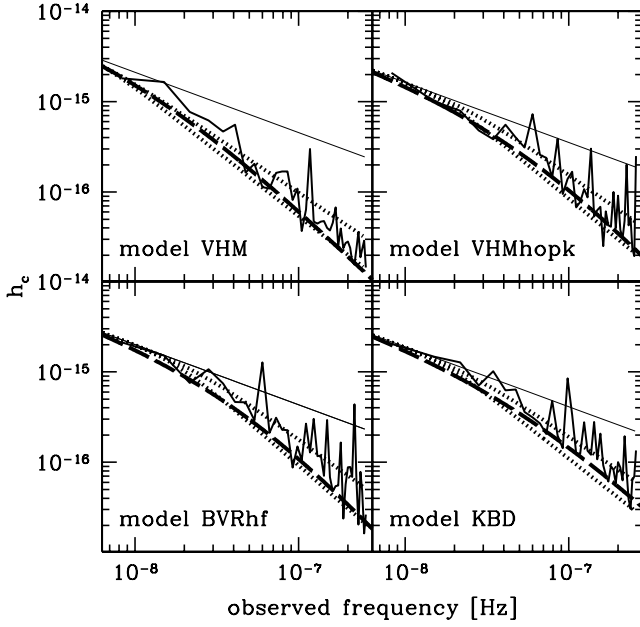
Figure 9. The very low frequency GW signal for the four assembly scenarios explored in the paper. In each panel, the thin line represents the naïve standard semi-analytical amplitude $h_c \propto f^{-2/3}$ while the spiky solid line shows the output of a specific Monte Carlo realization of the radiation from the whole MBHB population assuming a 5 yr observation. The lower dotted line marks the level at which there is at least one source per frequency bin ($\Delta f \sim 6 \times 10^{-9}$) according to the model developed in Section 4.3, and the upper dotted line shows the signal amplitude from the whole population, again evaluated following the model discussed in Section 4.3. The dashed line is the best fit to the Monte Carlo realization of the stochastic background given by equation (14) and Table 1.

assembly model. For $f > 10^{-8}$ Hz, most of the signal predicted by semi-analytical approaches is actually not present because of the *discrete nature of sources*.

5 SOURCES OF UNCERTAINTY

We have shown that the stochastic background is well described by equation (14), that depends on the three parameters h_0, f_0 and γ, see Table 1. The amplitude h_0 sets the overall strength of the signal; equation (10) shows that $h_c^2(f)$ is proportional to the number N of sources emitting at frequency f multiplied by the contribution h^2 of each individual source. Since $h^2 \propto \mathcal{M}^{10/3}$, we have

$$h_0 \propto \sqrt{N} \, \mathcal{M}^{5/3}. \tag{32}$$

The overall level of the GW stochastic background is therefore set by N, that depends on the merger rate of galactic haloes, the MBH occupation fraction and the efficiency of the MBHB coalescence, and by $\mathcal{M}$, that depends on the hosted MBH mass function and on the typical mass ratio of the coalescing binaries. There are considerable uncertainties surrounding the values of these parameters and others that enter the relevant physical processes that ultimately affect the amplitude of the GW stochastic background. In this section, we *quantify* such uncertainties according to our current theoretical understanding and observational data, and, as a consequence the prospects of detection with present and future PTA surveys.

Assuming that N and $\mathcal{M}$ are uncorrelated, uncertainties in N and $\mathcal{M}$ will be reflected into a range of possible values for h_0; we indicate with σ_N and $\sigma_{\mathcal{M}}$ the root-mean-square (rms) spread

of values in N and $\mathcal{M}$, respectively, computed from sampling a number of possible models and scenarios; this translates into a rms range of possible levels of the GW stochastic background, denoted by σ_{h_0} around a mean value $\langle h_0 \rangle$. The model uncertainties σ_N, $\sigma_{\mathcal{M}}$ and σ_{h_0} are related through equation (32) by

$$\left(\frac{\sigma_{h_0}}{h_0}\right)^2 = \left(\frac{1}{2}\frac{\sigma_N}{N}\right)^2 + \left(\frac{5}{3}\frac{\sigma_{\mathcal{M}}}{\mathcal{M}}\right)^2. \tag{33}$$

A rigorous quantitative estimate of the impact of uncertainties in the merger rate and the MBH mass function on f_0 and γ is not straightforward. We note, however, that for the four models considered in this paper γ varies only by $\approx$10 per cent, and f_0 changes by no more than a factor of $\approx$3. We then proceed as follow. We calculate from the four fits given in Table 1 the mean values $\langle f_0 \rangle$ and $\langle \gamma \rangle$. We do not try to model the possible errors on f_0 and γ but we just consider as their appropriate uncertainty range the bracketing values found in our four models. We find $f_0 = 3.72(+1.52, -1.30) \times 10^{-8}$ Hz and $\gamma = -1.08(+0.03, -0.04)$. In the following, we will compute $\langle h_0 \rangle$ and σ_{h_0}.

5.1 Black hole mass function

Our poor knowledge of the shape and the normalization of the black hole mass function (BHMF) at $z = 0$ is the major uncertainty factor in the determination of $\mathcal{M}$. The hierarchical models employed here are bound to reproduce at any redshift the M_{BH}–σ relation observed in the local Universe (Tremaine et al. 2002). The MBH population predicted at $z = 0$ is then found to be consistent with the BHMF derived observationally by inferring MBH masses from measurements of the bulge stellar velocity dispersion (Sheth et al. 2003). The key point is that we can infer the BHMF either relying on the M_{BH}–σ relation, or on the M_{BH}–M_{bulge} relation (e.g. Haring & Rix 2004), and the two approaches give inconsistent results especially for heavier MBHs (e.g. Lauer et al. 2007). This is particularly relevant here because the more MBHBs are the main contributors to the GW background at very low frequencies (see Figs 1 and 5). To quantify the uncertainties related to the BHMF, we refer to the analysis performed by Tundo et al. (2007). In Fig. 10, the local BHMFs (with the relative uncertainties) derived from the two relations are shown (see Tundo et al. 2007 for details). As it is clear from Fig. 5, almost all the signal comes from binaries with $\mathcal{M} > 10^7$. The chirp mass is related to the primary (i.e. heavier) black hole mass M_1 and the binary mass ratio $q = M_2/M_1 \leq 1$ by $\mathcal{M} = M_1 q^{3/5}/(1+q)^{1/5}$. As a consequence, for a given q-distribution, the mean chirp mass of the binaries is proportional to the mean mass of the MBH population. Our analysis here concentrates on MBHs with masses $>10^{7.5} \, M_{\odot}$, that dominate by far the GW stochastic background, see Figs 1 and 5 . For each mass function, we can calculate the average MBH mass of the population as

$$M_{av} = \frac{\int_{7.5}^{\infty} \frac{dn}{d\log M_{BH}} M_{BH} \, d\log M_{BH}}{\int_{7.5}^{\infty} \frac{dn}{d\log M_{BH}} \, d\log M_{BH}}. \tag{34}$$

Since in all our models, the q-distributions have a similar shape peaked at $q \approx 0.1$–0.2, we can assume that, approximately, $\mathcal{M} \propto M_{av}$, so that the error on M_{av} directly quantifies the error on $\mathcal{M}$. We consider the four curves from Tundo et al. (2007) as they were the outcome of four different MBH formation models; we calculate M_{av} for each model, and then we evaluate the mean M_{av} and the relative variance. We find $\Delta M_{av}/\langle M_{av} \rangle \sim 0.31$. Note that the uncertainty on M_{av} is dominated by the discrepancy between the BHMF evaluated via the two different methods. The four merger tree realizations that
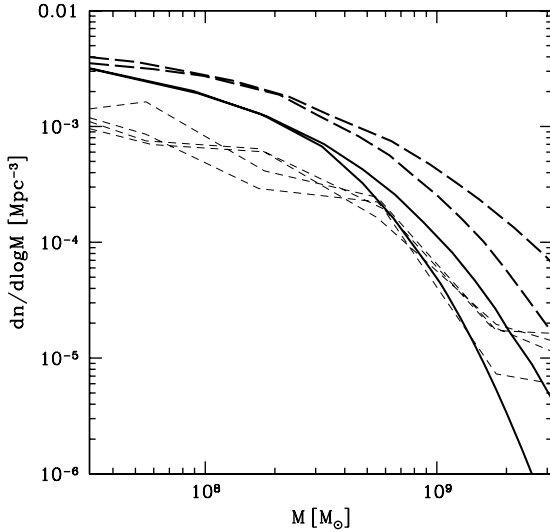
Figure 10. The local black hole mass function computed under different assumptions. The thick lines bracket the uncertainties in the black hole mass function computed by Tundo et al. (2007) either on the basis of the M_{BH}–σ relation (solid lines), or assuming the M_{BH}–M_{bulge} relation (dashed lines). The thin dashed lines are the outputs of the four merger tree models employed in this work.

we have employed in this paper give results broadly consistent with the BHMF determination using the M_{BH}–σ relation. The MBH mass density is in the range 2–$3 \times 10^5\,\mathrm{M}_\odot\,\mathrm{Mpc}^{-3}$, consistent with the values 2.4–$3 \times 10^5\,\mathrm{M}_\odot\,\mathrm{Mpc}^{-3}$ obtained integrating the BHMFs given by Tundo et al. (2007). The associated uncertainties in the BHMF at $z = 0$ are much smaller than those due to the dichotomy M_{BH}–σ/M_{BH}–M_{bulge}, and we will neglect them here.

We note that the poor knowledge of the BHMF causes also an uncertainty in the number density n_{BH} of MBHs at $z = 0$; focusing on the mass range relevant for this discussion, we define

$$n_{BH} = \int_{7.5}^{\infty} \frac{\mathrm{d}n}{\mathrm{dlog}M_{BH}}\,\mathrm{dlog}M_{BH}. \qquad (35)$$

This fact is likely to produce an uncertainty in N. However, we will see in Section 5.2 that we can directly evaluate the uncertainty in N from the halo merger history, and we will also find that the associated uncertainties are much larger than those produced by the uncertainties in n_{BH}; the latter will therefore be ignored.

So far, we have concentrated on the implications of the BHMF uncertainties at $z = 0$. However, Fig. 4 shows clearly that all the MBHs coalescing at $z < 2$ contribute significantly to the GW stochastic background. As our models need to reproduce the local M_{BH}–σ at any redshift, the possibility of a redshift evolution of this relation is a further factor of uncertainty in the background level. Whether the M_{BH}–σ and M_{BH}–M_{bulge} relations evolve with redshift is still a matter of debate. The redshift evolution of the M_{BH}–σ is quite controversial. Shields et al. (2003) found that the quasars between $z = 0$ and 3.5 are consistent with the local M_{BH}–σ relation. Robertson et al. (2006) investigated the galaxy merger driven M_{BH}–σ evolution with redshift by means of detailed hydrodynamical N-body simulations, finding almost no evolution for $0 < z < 2$. On the contrary, Treu, Malkan & Blandford (2004) found an offset of $\Delta \log \sigma = -0.16$ in quasars observed at $z = 0.37$ with respect to the local relation (Tremaine et al. 2002). If confirmed, this would imply, for a given σ, an hosted MBH 4.4 times more massive at $z = 0.37$ than in the local Universe. We can estimate the impact of the results by

Treu et al. (2004) on the strength of the GW stochastic background. Woo et al. (2006) provide the following best fit for the offset with respect to the local relation:

$$\Delta \log M_{BH} = 1.66z + 0.04. \qquad (36)$$

If we correct the masses of the coalescing binaries according to equation (36) in the range $0 < z < 2$ and we decrease accordingly the number of coalescing binaries in order to satisfy the constraints on the MBH mass function given by the Soltan (1982) argument, h_0 would increase from $\simeq 10^{-15}$ to $\simeq 10^{-14}$, i.e. on the border of the current PTA limits (see Jenet et al. 2006 and the discussion in Section 7). This is because the characteristic amplitude of the background is proportional to $\mathcal{M}^{5/3}N^{1/2}$, hence if we increase the mass and lower the number of sources by the same factor, the amplitude of the background increases. This is just an example of how PTAs could provide information on the evolution of the MBH population at low redshift.

There is growing evidence of a pronounced redshift evolution of the M_{BH}–M_{bulge} relation. The best fit is given by McLure et al. (2006)

$$\frac{M_{BH}}{M_{bulge}} = 10^{-3.09}(1 + z)^{2.07}. \qquad (37)$$

Given M_{bulge}, this implies that, e.g., the host MBH at $z = 2$ should be eight time more massive than the host MBH at $z = 0$. However at this stage we cannot rigorously quantify the implications of such possible evolution on the GW background, since our models are based on the M_{BH}–σ. Robertson et al. (2006) found a significant redshift dependence in the M_{bulge}–σ relation. For a given σ, the corresponding M_{bulge} is four times smaller at $z = 2$ than now. This could be a clue that a mild redshift evolution in the M_{BH}–σ relation could be consistent with the evolution shown by equation (37) in the M_{BH}–M_{bulge} relation.

The redshift dependence of the M_{BH}–σ and M_{BH}–M_{bulge} relations is currently debated and there is much freedom at present for theoretical speculations not constrained by observations. We have broadly quantified the possible implications on the strength of the background, depending on the actual scenario and we will not dwell on this any further. We just point out that PTAs – even upper limits on and not just detection of the stochastic background generated by populations of MBHBs – will be able to provide direct new constraints on MBH populations and their assembly history that are likely to be difficult to infer using a number of indirect observations.

5.2 Galaxy merger rate

The major factor of uncertainty about the total number of sources is related to the determination of the merger rate of massive (i.e. with total stellar mass $M_\star > 10^{10}\,\mathrm{M}_\odot$) galaxies at low redshift (say, $z \lesssim 2$): once a relation between the MBH and the spheroid is assumed (either the M_{BH}–M_{bulge} or the M_{BH}–σ relation, as we have discussed in the last section), the number of MBH pairs is proportional to the spheroid merger rate. EPS-based models tend to overpredict the bright end of the quasar luminosity function (LF) at $z < 1$ (e.g. Marulli et al. 2006); however, although the quasar activity is related to merger events, this discrepancy in the predicted and observed LF does not automatically imply an overestimate of the galaxy merger rate. In fact, little is known about the details of the quasar powering mechanism, the accretion efficiency and quasar duty cycle, and the LF overprediction could be ascribed to these factors, without necessarily implying an overestimate of the galaxy merger rate.

A more robust test is to compare the low-redshift merger rates derived with EPS trees, with the merger rates inferred by observations of the fraction of close galaxy pairs. Our analysis is based on the most recent and complete study to date (Bell et al. 2006). Starting from the COMBO survey data (Borch et al. 2006), Bell et al. give the mean fraction of pairs ϕ – defined as galaxies with 3D separation less than 30 kpc – and the comoving number density of galaxies n_G in the redshift range $0.4 < z < 0.8$ assuming two different galaxy total stellar mass cuts $M_{*,\text{th}}$: (i) $M_{*,1,2} > 2.5 \times 10^{10} \, \text{M}_\odot$, and (ii) $M_{*,1} > 3 \times 10^{10} \, \text{M}_\odot$ and $M_{*,2} > M_{*,1}/3$. Here, $M_{*,1}$ and $M_{*,2}$ are the total masses of the primary and secondary galaxy in the pair, respectively.

The merger rate of galaxies, assuming a typical merger time-scale $\mathcal{T}_M$, is given by

$$\dot{N}_M = \frac{\phi}{2\mathcal{T}_M} \int n_G \, dV_c, \tag{38}$$

where the integral is calculated between the survey redshift limits ($0.4 < z < 0.8$) and the factor of $1/2$ takes into account that two objects participate in a single merger event. Applying this calculation to the two samples considered by Bell et al. (2006) and relying on their estimate of $\mathcal{T}_M \sim 4 \times 10^8$ yr (see also Barnes & Hernquist 1992; Patton et al. 2002; Lin et al. 2004; Naab, Khochfar & Burkert 2006), we find $\dot{N}_M \approx 5 \times 10^{-3} \, \text{yr}^{-1}$ for case (i) and $\dot{N}_M \sim 2 \times 10^{-3} \, \text{yr}^{-1}$ for case (ii). Bell at al. estimate an error of ~ 2 in their results. If we assume for massive galaxies a MBH occupation fraction of 1 (e.g. Richstone et al. 1998), the MBHB coalescence rate inferred by observations is the same as the galaxy merger rate. To compare these values to the MBHB coalescence rate obtained from EPS-based merger trees, we need to convert the galaxy stellar mass threshold to a MBH mass threshold $M_{\text{BH,th}}$. As we are dealing with massive galaxies, to a first approximation we can use the M_{BH}–M_{bulge} relation (Haring & Rix 2004)

$$M_{\text{BH}} = 0.0014 M_{\text{bulge}} \tag{39}$$

that has been shown to approximately hold at least up to $z = 1$. McLure et al. (2006) found $M_{\text{BH}}/M_{\text{bulge}} \propto (1 + z)^2$, but the normalization of their best fit to this relation at $z = 0$ is lower than 0.0014; the relation given by equation (39) is within the errorbars given by McLure et al. (2006) for $0.4 < z < 0.8$. Note that by adopting the relation (39) we are implicitly assuming that all the observed galaxies are bulge-dominated. We can then consider only MBH more massive than $M_{\text{BH,th}}$ according to equation (39); if we integrate in the redshift range $0.4 < z < 0.8$ the MBHB merger rate obtained by EPS models, we obtain the values listed in the left-hand columns of Table 2. The mean rates are a factor of 3 lower than those inferred by applying equation (38) to observations, and the dispersion

around the mean is larger for the case (ii). However, the M_{BH}–M_{bulge} conversion is too simplistic. At the considered mass cuts, the luminosity function is dominated by disc type galaxies (Borch et al. 2006; Ilbert et al. 2006). Since typically only ≈ 10 per cent of the stellar content of disc galaxies is contained in the bulge, that galaxies corresponding to the survey stellar mass threshold actually contain a MBH with $M_{\text{BH,th}} = 0.000\,14 \, M_{*,\text{th}}$, if we require them to fit the relation (39). For example, a stellar mass cut of $M_{*,1} = 3 \times 10^{10} \, \text{M}_\odot$ corresponds to a host MBH of mass $\approx 4 \times 10^6 \, \text{M}_\odot$ and not $\approx 4 \times 10^7 \, \text{M}_\odot$. We then evaluate the MBH coalescence rate given by EPS trees lowering the MBH threshold by an order of magnitude. By doing so, we obtain the merger rates listed in the right-hand columns of Table 2; the mean rates are consistent – they differ by a factor of $\lesssim 2$ – with those obtained from observational estimates. The EPS halo mass function is quite sensitive at high halo masses; we checked that changes in the mass function within reasonable limits affect the merger rate by $\lesssim 40$ per cent. Since Bell et al. (2006) quote a factor of 2 uncertainty in their rate estimate, we can safely conclude that their MBHB coalescence rates are consistent with our EPS-based theoretical models.

Though Bell et al. (2006) give directly the merger rate of galaxies above a chosen mass threshold, there are several other works that evaluate the merger rate of galaxies in a given magnitude range. The comparison with the EPS merger tree estimates used in this paper is more difficult in this case, since we have to convert the magnitude limit into a mass limit which introduces a whole new range of uncertainties. However, it is still valuable to compare the merger rates obtained in this way with those derived with EPS merger trees as a sanity check. In order to do so, we combine the relations

$$\log L_H = 0.8 - 0.5 M_B \tag{40}$$

and

$$\log M_* = \frac{\log L_H - 0.12}{0.99}, \tag{41}$$

where L_H is the H-band luminosity and M_B in the B band absolute magnitude (Zibetti et al. 2002). Lin et al. (2004) derive an average merger rate of $4 \times 10^{-4} \, h_{100}^3 \, \text{Mpc}^{-3} \, \text{Gyr}^{-1}$ at $0.5 < z < 1.2$ for galaxies with $-21 < M_B < -19$ (where h_{100} is $H_0/100 \, \text{km} \, \text{s}^{-1} \, \text{Mpc}^{-1}$ and we adopt $h_{100} = 0.7$). If we multiply this number by the comoving volume enclosed in the redshift range $0.5 < z < 1.2$, we obtain a merger rate of $2.5 \times 10^{-3} \, \text{yr}^{-1}$. Converting the magnitude limits in stellar mass thresholds using equations (40) and (41) and estimating the MBH mass using equation (39), we obtain, from our merger trees, rates of the order of $\sim 5 \times 10^{-4} \, \text{yr}^{-1}$. If we use instead $M_{\text{BH,th}} = 0.000\,14 \, M_{*,\text{th}}$, the rates reach values $\sim 10^{-2} \, \text{yr}^{-1}$. Also in this case the observational rate is consistent (at least within a factor of ≈ 5) with the merger tree estimations. Repeating the same analysis using the rate found by De Propris et al. (2007) for galaxies with $-21 < M_B < -18$ at $0.01 < z < 0.123$, we obtain once again consistent results. Merger tree rates are found to be slightly higher (by a factor of $\lesssim 2$) than the observed rates at very low redshift.

Since the differences between model predictions and observations are of the same order of the spread of the merger rate values that we find in our four models, we directly estimate the mean square range of MBHB merger rate $\sigma_{\dot{N}_M}$ using the rate shown in Table 2; we find $\sigma_{\dot{N}_M}/\langle \dot{N}_M \rangle \simeq 0.76$. Since $\mathcal{M} \propto M_{\text{av}}$ and $N \propto \dot{N}_M$, we can write

$$\left(\frac{\sigma_{h_0}}{\langle h_0 \rangle} \right)^2 = \left(\frac{1}{2} \frac{\sigma_{M_{\text{av}}}}{\langle M_{\text{av}} \rangle} \right)^2 + \left(\frac{5}{3} \frac{\sigma_{\dot{N}_M}}{\langle \dot{N}_M \rangle} \right)^2. \tag{42}$$

Table 2. Merger rates inferred from our EPS merger trees in the redshift range $0.4 < z < 0.8$. Columns (i) assume a stellar mass cut $M_{*,1,2} > 2.5 \times 10^{10} \, \text{M}_\odot$, while columns (ii) consider $M_{*,1} > 3 \times 10^{10} \, \text{M}_\odot$, $M_{*,2} > 1/3 M_{*,1}$. Results are shown by assuming that all the observed galaxies are elliptical ($M_{\text{BH,th}} = 0.0014 M_{*,\text{th}}$), and by taking into account that the galaxy sample at the threshold mass is dominated by spirals ($M_{\text{BH,th}} = 0.000\,14 \, M_{*,\text{th}}$).

| | $\dot{N}_M$ | | | |
| | $M_{\text{BH,th}} = 0.0014 M_{*,\text{th}}$ | | $M_{\text{BH,th}} = 0.000\,14 \, M_{*,\text{th}}$ | |
Model	(i)	(ii)	(i)	(ii)
VHM	1.4×10^{-3}	1.1×10^{-4}	4.9×10^{-3}	5.3×10^{-4}
VHMhopk	2.4×10^{-3}	2.0×10^{-3}	9.4×10^{-3}	5.1×10^{-3}
KBD	2.1×10^{-3}	2.2×10^{-4}	9.6×10^{-3}	1.1×10^{-3}
BVRhf	1.8×10^{-3}	3.4×10^{-4}	1.2×10^{-2}	1.5×10^{-3}

Substituting $\sigma_{\dot{N}_M}/\langle\dot{N}_M\rangle = 0.76$ and $\sigma_{M_{av}}/\langle M_{av}\rangle = 0.31$ into equation (42), we find

$$\frac{\sigma_{h_0}}{\langle h_0\rangle} = 0.65. \tag{43}$$

We need now to determine $\langle h_0\rangle$ and we proceed as follows. The merger trees are bound to reproduce the M_{BH}–σ relation, so they provide a value of $\langle h_0\rangle$ that is not representative of the whole range of possibilities for the BHMF shown in Fig. 10. To provide a more representative estimate of $\langle h_0\rangle$, we calculate $\langle h_{0(M_{BH}\sigma)}\rangle$ from our merger tree models, and then we correct it for the fact that the M_{av} inferred from the M_{BH}–σ relation [$M_{av,(M_{BH}-\sigma)}$] is lower than M_{av} if also the M_{BH}–M_{bulge} relation is taken into account:

$$\langle h_0\rangle = \langle h_{0(M_{BH}-\sigma)}\rangle\left[\frac{M_{av}}{M_{av,(M_{BH}-\sigma)}}\right]^{5/3}. \tag{44}$$

(Here, we used the fact that, for a given mass ratio distribution, the mean $\mathcal{M}$ is proportional to M_{av}.) Since $M_{av}/M_{av,(M_{BH}-\sigma)} \simeq 1.4$ (from Tundo et al. 2007) and $\langle h_{0(M_{BH}-\sigma)}\rangle \simeq 1.1 \times 10^{-15}$, we finally find $h_0 = (1.93 \pm 1.25) \times 10^{-15}$.

Summarizing, the GW stochastic background is well approximated by the relation given by equation (14) that depends on three parameters whose range of values is

$$h_0 = (1.93 \pm 1.25) \times 10^{-15}, \tag{45}$$

$$f_0 = 3.72^{+1.52}_{-1.30} \times 10^{-8}\,\mathrm{Hz}, \tag{46}$$

$$\gamma = -1.08^{+0.03}_{-0.04}, \tag{47}$$

their specific value for the four models considered in this paper are shown in Table 1. The results of our analysis are shown in Fig. 11. As expected, the outcomes of our merger trees, based on the M_{BH}–σ relation, lie in the lower region of the estimated errorbar, below the best fit to the background.
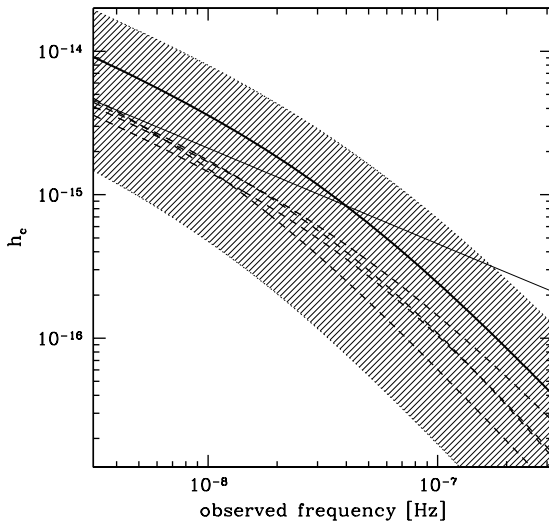
Figure 11. Summary of the uncertainties of the predicted characteristic amplitude of the GW stochastic background. The thick solid line shows the mean level of the signal according to the analysis discussed in the text, while the thin dashed lines represent the predictions of the four MBHB assembly scenarios considered in this paper. The shaded area delimits the 1σ region around the mean value. For comparison, the na*i*ve power law $f^{-2/3}$ is plotted as a thin solid line for the VHM model.

5.3 Other uncertainty factors

The spread of values of the estimate of GW stochastic background that we have discussed above is likely to be affected by other factors, in particular the MBHB coalescence efficiency, the gravitational recoil and the MBH accretion prescription adopted during mergers. We consider them now in some detail.

Our knowledge of the MBHB coalescence efficiency is far from robust. The major contribution to the GW background signal comes from massive binaries at low redshift, that are expected to be hosted in early type, gas poor spheroidal galaxies. Stellar dynamical processes are therefore likely to be the main driver in MBHB coalescences. It is, however, known that three-body interactions (that extract energy from the MBHBs through the so-called slingshot mechanism) are not sufficient to lead to a MBHB merger within an Hubble time in spherical systems (e.g. Sesana, Haardt & Madau 2007b), unless the binary is very eccentric. The scenario is different in axisymmetric and triaxial systems (more realistic if we consider the end-product of a merger), and there are indications (Berczik et al. 2006) that in the latter case the stellar supply to the slingshot process could be sufficient to drive binaries to coalescence within an Hubble time. If only a fraction ϵ_c of binaries coalesce, then the stochastic background characteristic amplitude h_c drops by a factor of $\sqrt{\epsilon_c}$ at all frequencies.

A MBH formed via the merger of a binary system recoils due to the net linear momentum imparted on the object through emission of GWs (Redmount & Rees 1989). The recoil velocity has been shown to possibly reach high values $\sim 10^3\,\mathrm{km\,s^{-1}}$ (e.g. Campanelli et al. 2007; Tichy & Marronetti 2007) and may therefore lead to a depletion of MBHs in the core of some galaxies. These galaxies involved in mergers at later cosmic times would therefore not form a MBH binary, with a consequent decrease in the number of MBHBs and in the strength of the associated GW stochastic signal. In order to address the impact of GW recoil on the estimates of the level of the stochastic background, we have run a series of tests using the VHMhopk and BVRhf models where we compare the signal strength in the case where the recoil is neglected and the case where the recoil is included using an *extreme* prescription following Campanelli et al. (2007). We find negligible differences in the BVRhf case, whereas differences as large as a factor of 2 in the GW background amplitude occur in the VHMhopk scenario. However, such a difference is between the two prescriptions of no recoil at all and of extreme recoil. We know that the GW recoil exists, and the differences in the computed background between the non-spinning MBH recoil recipe (i.e. our default models) and the extreme recipe for maximally spinning MBHs are indeed tiny.

The modelling of accretion on to a MBH could also influence the background level. Although we have tested that the GW background is basically unaffected by the details of the accretion on to a single MBH (e.g. the dependence on z and spin), an important issue is whether both MBHs typically accrete mass during mergers. We have considered models in which both MBHs are subject to mass accretion (in our original models only the more MBH in the pair undergoes such a stage), finding that the predicted background amplitude could even double, due to the increase in the MBHB chirp mass.

All these factors add further uncertainties to the estimates reported in the previous section, but our knowledge of the MBH accretion and the coalescence efficiency is too poor at present to allow us to provide stringent quantitative constraints on the level of the GW background. In general, the uncertainties due to the accretion prescription should change by at most a factor of ≈ 2 the

amplitude of the signal, and the coalescence efficiency could just reduce the strength of the background with respect to the values reported here, since in all our models we have set $\epsilon_c = 1$.

6 COMPARISON WITH PREVIOUS WORK

The very low frequency GW stochastic background from MBHBs has been computed in a handful of papers over the last 10 yr. Previous works have only concentrated on a (in turn different) *specific* MBHB assembly scenario, none of which matches exactly any of those considered in this paper; they have modelled the background as a simple power law as given by equation (11), that we have shown does not correctly describe the signal across the whole frequency band accessible to PTA observations. In this section, we briefly summarize how previous estimates compare among each other and with those presented in this paper.

Studies based on the EPS formalism (Wyithe & Loeb 2003; Enoki et al. 2004; Sesana et al. 2004), though relying on different prescriptions for the seed BH mass, the MBH accretion and the evolution of the stellar spheroids, yield all $h_{1\,\mathrm{yr}} \sim 10^{-15}$, see equation (11). This is not surprising, since the very low frequency background is strongly dominated by massive nearby sources that are fairly constrained by current observations (see the discussion in the previous section): there are well-established relations between the BH mass, the spheroids mass (Magorrian et al. 1998) and the velocity dispersion (Ferrarese & Merritt 2000; Gebhardt et al. 2000; Tremaine et al. 2002) that have been shown to hold at least up to $z \lesssim 1$ (McLure & Dunlop 2002; McLure et al. 2006; Peng et al. 2006), and the stellar and nuclear MBH mass density are today known within an uncertainty of a factor of ≈ 2 (e.g. Marconi et al. 2004; Borch et al. 2006). Since the goal of every assembly model is to reproduce these low-redshift observational constraints, all the scenarios based on the same halo merger rate function (extracted in this case using the EPS formalism, either with analytical or numerical techniques), lead to comparable (i.e. within a factor of ≈ 2) low-frequency GW stochastic background levels.

Earlier work (R&R, J&B), that inferred the MBHB coalescence rate from available observations of the fractions of galaxies forming kinematically close pairs in the local Universe, give somewhat lower (by a factor of ~ 5) normalizations of the spectrum. It is useful to recall here that these models focus specifically on the GW background, and do not try to match other observational constraints in a self-consistent MBH evolution model. J&B used the spheroid merger rate found by Carlberg et al. (2000) for $0.1 < z < 1.1$, and then populate the spheroids with MBHs according to a lognormal MBH mass function given by

$$\left\langle \log_{10} \frac{M_{\mathrm{BH}}}{\mathrm{M_\odot}} \right\rangle = \log_{10}(1.2 \times 10^7) \pm 0.6. \tag{48}$$

However, the merger rates of Carlberg et al. (2000) are evaluated for galaxies with magnitude $M_B < -20.4$, corresponding to massive galaxies with stellar mass $M_* \gtrsim 2 \times 10^{10}\,\mathrm{M_\odot}$. In practice, J&B use the merger rate of massive galaxies as the *total* galaxy merger rate. In Fig. 12, the galaxy merger rates used by J&B and R&R are compared with the MBHB coalescence rates obtained with our EPS-based halo merger tree models placing a lower cut at $\mathcal{M} = 10^{7.5}\,\mathrm{M_\odot}$. EPS rates for $\mathcal{M} > 10^{7.5}\,\mathrm{M_\odot}$ are similar to those used by both J&B and R&R as *total* rates at $z < 1$, reflecting the fact that those 'observational' rates refer to massive galaxies. Using the MBH mass function distribution given by equation (48), we find that only ≈ 13 per cent of the J&B mergers involve binaries with $\mathcal{M} > 10^{7.5}\,\mathrm{M_\odot}$, so the discrepancy of a factor of ≈ 5
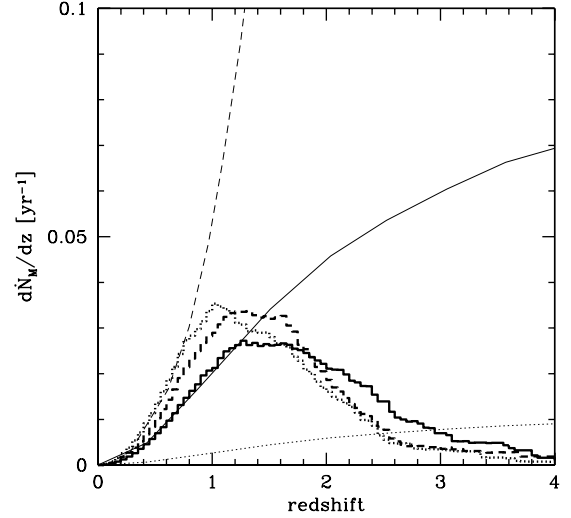

Figure 12. Spheroid merger rates and MBHB coalescence rates (per unit redshift) as a function of redshift. The histograms represent the coalescence rates for binaries with $\mathcal{M} > 10^{7.5}\,\mathrm{M_\odot}$ predicted by the VHM (thick solid line), KBD (thick dotted line) and BVRhf (thick dashed line) models. The thin solid line shows the spheroid merger rate adopted by J&B, while the thin dotted line is the coalescence rate for MBHBs with $\mathcal{M} > 10^{7.5}\,\mathrm{M_\odot}$ that we have derived from the same model considered by J&B. The thin dashed line corresponds to the galaxy merger rate adopted by R&R.

in the expected signal is reasonable. Similar considerations hold for the R&R model, which moreover employs very conservative assumptions on the MBH number density evolution with redshift.

7 OBSERVATIONAL CONSEQUENCES

7.1 Pulsar Timing Arrays

We have considered a wide range of models of formation and evolution of MBHBs and estimated the GW stochastic background that one can expect as a function of the model parameters. We have further shown that the stochastic signal does not follow a simple power law $h_c(f) \propto f^{-2/3}$ for $f \gtrsim 10^{-8}$ Hz, as previously claimed. These results are summarized in equation (14), Table 1 and Fig. 9. We can now explore, for a much more realistic range of models than those considered in the past, the observational consequences; in particular, we can investigate whether current and/or future PTAs would be able to detect such a signal, opening therefore a new window for the study of MBHBs and structure formation, or at least setting sufficiently stringent upper limits to rule out selected scenarios.

As we have discussed in Introduction, the technique to search for a stochastic background using PTAs consists in correlating the timing residuals from N_{P} pulsars.[2] The timing residual from a given pulsar is given by

$$\delta t = \delta t_n + \delta t_h, \tag{49}$$

where $\delta t_h \sim h/f$ and δt_n are the GW signal and noise contribution, respectively. If one considers for simplicity the case of only two pulsars, $N_{\mathrm{P}} = 2$, then the minimum detectable signal (assuming that

---

[2] We note that an upper bound to any stochastic GW background can be placed by monitoring a *single* pulsar and assuming that the observed value of the timing residuals is entirely due to GWs.

its spectrum is dominated by the low-frequency contribution, as it is the case here) is characterized by

$$h_{100}^2 \Omega_{gw}(f) \propto \frac{\delta t_{rms}^2 f^4}{\sqrt{T \Delta f}}, \qquad (50)$$

where $\delta t_{rms} = \sqrt{\langle \delta t^2 \rangle}$ is the rms value of the timing residuals and Δf the bandwidth of the search. Note that this is the PTA equivalent of the result that one obtains by considering *direct* searches for GW stochastic backgrounds using the cross-correlation of the data from two GW laser interferometers (see Allen & Romano 1999 for a detailed discussion, and in particular section 3 for material relevant to this section). If instead of two, one has many pulsars, then the optimal signal-to-noise ratio (SNR) is given by the combination of all the statistically independent correlations that can be formed:

$$SNR_T^2 = \sum_{i<j}^{N_P} SNR_{ij}^2 = \frac{N_P(N_P - 1)}{2} SNR^2, \qquad (51)$$

where SNR_{ij} is the SNR of the pair composed by the *i*th and the *j*th pulsars; and in the second equality, we have assumed for simplicity $SNR_{ij} = SNR \ \forall i, j$. For N_P sufficiently large, the SNR scales as $SNR_T^2 \propto N_P^2$ and therefore

$$h_{100}^2 \Omega_{gw}(f) \propto \frac{\delta t_{rms}^2 f^4}{N_P \sqrt{T \Delta f}}. \qquad (52)$$

Using equations (1) and (2), we finally obtain

$$h_c(f) \propto \frac{\delta t_{rms} f}{N_P^{1/2} (T \Delta f)^{1/4}}. \qquad (53)$$

The PTA sensitivity scales as $h_c(f) \propto f$ and reaches a minimum detectable frequency $f \sim 1/T$. This produces the characteristic wedge-like sensitivity curves as those shown in Fig. 13. The sensitivity is proportional to the timing precision (δt_{rms}) and improves as the square root of the number of observed pulsars; the improvement with the observational time *at a given frequency* scales as $h_c(f) \propto T^{-1/4}$, but note that a longer T implies that lower frequencies can be reached, which provides an additional sensitivity gain. The curves shown in Fig. 13 are derived by normalizing equation (53) to the sensitivity quoted by Jenet et al. (2006) for the complete Parkes PTA (labelled as 'complete PPTA' in Fig. 13), defined as the monitoring of 20 radio pulsars for 5 yr with $\delta t_{rms} = 100$ ns: this array would provide a peak sensitivity of $h_c \approx 2 \times 10^{-15}$ at $f = 7 \times 10^{-9}$ Hz. Jenet et al. (2006) quote for the current Parkes PTA (labelled as 'current limit' in Fig. 13) an optimal sensitivity equivalent to $h_c \approx 2 \times 10^{-14}$ at $f = 8 \times 10^{-9}$ Hz. Fig. 13 clearly shows that such a sensitivity is a factor of ~ 3 above the most optimistic estimate of the background that we have considered. No meaningful astrophysical information about MBHBs can therefore be derived from the current upper limit. However, the complete Parkes PTA should provide a sensitivity improvement of about an order of magnitude in amplitude; several scenarios and/or parameter values within a given model could yield a detection at this level; if detection is not achieved, upper limits could start ruling out, for example, scenarios in which the M_{BH}–M_{bulge} relation holds for BHs of all masses at all redshifts. However, the sensitivity of the PTA decreases in the frequency range where the stochastic signal is expected to deviate from the $f^{-2/3}$ behaviour, and positive detection is therefore not guaranteed. The planned SKA will provide a major increase in PTA capabilities. Assuming that the SKA will be able to monitor 20 ms pulsars for 10 yr at a precision level of $\delta t_{rms} \sim 10$ ns, it will allow us not only to detect the GW stochastic background, but also to study its spectrum in the frequency range $3 \times 10^{-9} \lesssim f \lesssim$ few
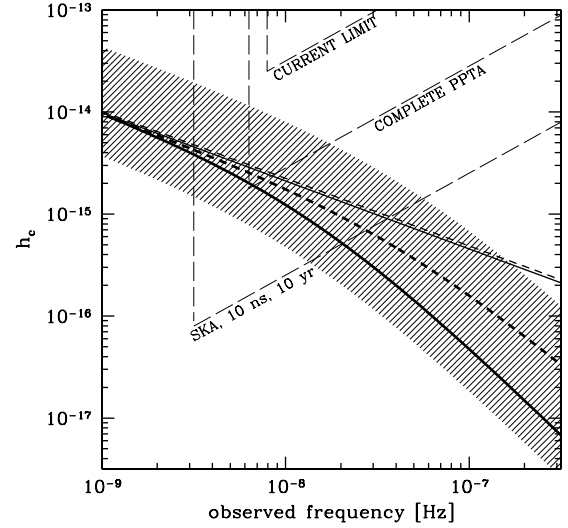

Figure 13. The sensitivity of Pulsar Timing Array in observations of a GW stochastic background. The long-dashed thin lines represent the sensitivities of current and future timing experiments, as labelled in the figure (see the text for more details). Three cases are considered: the current limit and the estimated sensitivity achievable by monitoring 20 pulsars for 5 yr with $\delta t_{rms} = 100$ ns – complete Parks PTA (PPTA) – are taken from Jenet et al. (2006). The sensitivity achievable with the SKA – assuming the monitoring of 20 pulsars for 10 yr at a precision level of $\delta t_{rms} \sim 10$ ns – is also shown. The solid lines depict the stochastic signal predicted by the naïve semi-analytical approach (thin), see equation (11) and by our Monte Carlo approach (thick), see equation (14) and Table 1, assuming the VHM model for the MBHs cosmic evolution; the short-dashed lines refer to the KBD model. The shaded area marks the possible range of GW background level, according to the model uncertainties, as discussed in Section 5.

$\times 10^{-8}$, enabling the complete characterization of the GW signal. In turn, this would allow us to derive detailed information about the MBH formation and evolution history. Even considering a less optimistic timing precision of $\delta t_{rms} \sim 50$, a comparable sensitivity can be obtained by monitoring ≈ 100 ms pulsars.

7.2 Complementarity of *LISA* and PTAs

In Section 3.2, see Fig. 1, we have shown that very different models of MBH assembly (VHM, KBD, BVR, VHMhopk) predict a GW stochastic background whose characteristic amplitude varies by less than factor of 2 at $f \sim 10^{-8}$ Hz, despite the radically different evolutionary paths of MBHBs and the number and mass of the initial seeds. These very same models can also be used to estimate the total number of coalescences of MBHBs detectable by *LISA* (Bender et al. 1998) in the frequency range 0.1–0.1 Hz, therefore approximately five orders of magnitude higher than the frequency window probed by PTAs. In the case of *LISA*, the numbers of coalescences of MBHBs differ by more than a factor of 100. As a specific example, the BVRhf and KBD models predict ~ 10 and $\sim 10^3$ *LISA* events, respectively (Sesana et al. 2007a), with MBHB coalescence taking place at frequency ~ 1 mHz. This is in stark contrast with predictions for a GW stochastic background, whose strength shows just a ≈ 10 per cent difference in the nHz range. As we pointed out in Section 6, these results reflect the fact that all our models reproduce by construction the local M_{BH}–σ relation, and being the main contributors to the very low frequency signal massive
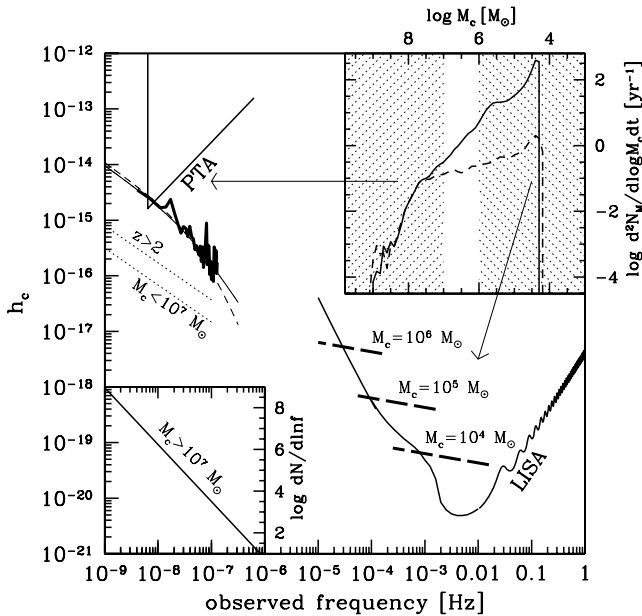
Figure 14. A representative summary of GW observations of the cosmic history of MBHBs with PTAs (through the detection of a GW stochastic background) and *LISA* (via the direct detection the final stage of the coalescence of individual binary systems). In the main panel, we show the characteristic amplitude of GW signals as a function of frequency, compared to the sensitivity of the complete Parkes PTA and *LISA*. The thin lines in the main panel represent the GW stochastic background (according to equation 14 and Table 1) for a MBH formation model in which seeds are abundant (KBD model, solid line), and one in which seeds are rare (BVRhf model, dashed line). A specific Monte Carlo realization of the total signal at frequencies accessible to PTAs is the thick-jagged line generated by nearby ($z < 2$) massive ($\mathcal{M} > 10^7\,\mathrm{M}_\odot$) sources: the thin dotted lines show, in fact, that the contribution by MBHBs at $z > 2$ or with $\mathcal{M} < 10^7\,\mathrm{M}_\odot$ is indeed negligible. In the *LISA* frequency band, one will observe the coalescence of individual sources, and most of the events involve binaries with $10^4\,\mathrm{M}_\odot < \mathcal{M} < 10^6\,\mathrm{M}_\odot$ at $z > 3$ (the thick long-dashed lines show the trace of the final stage of the inspiral for selected masses). The upper right-hand panel shows the MBHB coalescence rate $d^2 N_M / d\log \mathcal{M} dt$ as a function of $\mathcal{M}$ for the two models KDB and BVRhf, which is the same for $\mathcal{M} \gtrsim 10^8\,\mathrm{M}_\odot$ but differs as much as a factor of $\sim$100 for $\mathcal{M} \sim 10^4\,\mathrm{M}_\odot$. The lower left-hand panel shows the number of sources with $\mathcal{M} > 10^7\,\mathrm{M}_\odot$ contributing to the GW background. The slope of the curve is such that the signal is stochastic at very low frequencies, and that the *LISA* detection rate of binaries with $\mathcal{M} > 10^7\,\mathrm{M}_\odot$ is less than 1 yr^{-1}.

($\mathcal{M} \gtrsim 10^8\,\mathrm{M}_\odot$), nearby ($z \lesssim 2$) binary systems, their population is fairly well constrained by this assumption. On the other hand, very different MBH formation scenarios are compatible with the observed M_{BH}–σ; being the majority of the *LISA* sources lighter binaries at high z (up to $z \gtrsim 10$), current observations do not place stringent constraints on their abundance, and there is much more room for speculation.

LISA and PTA observations provide therefore fully complementary information on MBHs and their formation history. As we have shown in Section 5, the level of the stochastic background in the PTA frequency range is sensitive (a factor of $\approx$5) to the uncertainties in the MBH mass function at low redshift, to the halo merger rate variance and to the accretion prescription, but is fairly insensitive to the nature and abundance of the MBH seeds. In Fig. 14, we provide a summary sketch of such complementary. We consider the KBD

and the BVRhf assembly models which produce GW backgrounds at very low frequencies whose characteristic amplitudes differ by less than 10 per cent (main panel). The models are characterized by mergers of low-mass MBHs at high redshift; the merger rates (shown in the upper right-hand panel as a function of chirp mass) differ by two orders of magnitude. The lower left-hand side box shows that while the low-frequency background is composed by $\sim$$10^3$–$10^6$ sources with $\mathcal{M} > 10^7\,\mathrm{M}_\odot$, they are basically absent in the *LISA* sensitivity window, with a detection rate $\lesssim$1 yr^{-1}.

It is clear that PTA and *LISA* observations will place very different constraints on the assembly history of MBHBs and the nature of the MBH seeds. It also follows that the detection of a GW background with PTAs will bound very weakly the expected number of MBHBs observable with *LISA*. It is equally straightforward to verify that the reverse is also true. The vast majority of the *LISA* detections is associated with the first round of coalescence events after the MBH seed formation (Sesana 2007). The subsequent merger and accretion history of black holes has a minor impact, by a factor of $\lesssim$2, on the number of events, once the mass and the abundance of the seeds are defined (Sesana et al. 2007). It is not difficult to envisage two models with the same seed MBH population but with different assumptions regarding, e.g. the accretion or the relation between the MBH and the host galaxy; such models would predict similar *LISA* number counts but a different GW background level in the PTA frequency range. As an example we find that allowing accretion on both MBHs during each merger, instead of the standard accretion on to the more massive one, could affect the very low frequency GW level by a factor of 2, leaving essentially unchanged the *LISA* number counts.

8 SUMMARY

We have carried out a systematic study of the GW stochastic background generated by populations of MBHBs in the frequency range accessible to PTAs. We have considered a number of models that qualitatively encompass the whole range of assembly scenarios proposed so far and quantitatively covers a broad spectrum of predictions for MBH formation. Our analysis shows that, regardless of the actual model, the stochastic background is not described by a simple power law across the whole frequency spectrum (as considered so far) but by the superposition of two power laws with a knee frequency around 10^{-8} Hz. Above such a frequency, the amplitude of the background decreases more rapidly than $f^{-2/3}$. This is a direct consequence of the discrete nature of sources. A phenomenological parametrization of the amplitude of the GW stochastic background is given by equation (14). We have also carefully considered the present uncertainties that surround the actual values of the key parameters that set the level of the stochastic background and have provided an upper and a lower limit (based on the current theoretical understanding and observational constraints) to the strength of such signal: this provides useful benchmarks for the planning of future PTA observations.

As a byproduct of our analysis, we have compared the observed massive (luminous) galaxy merger rates at $z < 1$ derived by a number of authors with the merger rates inferred using the EPS approach adopted in this paper. We find good agreement with merger rates consistent with constraints provided by observations when we adopt standard prescriptions for the conversion of the galaxy luminosity into a MBH mass.

We have shown that the estimated GW background level is close – within a factor of $\approx$3 for the most optimistic predictions – to

the upper limit provided by current PTAs. The complete Parkes PTA should be able to detect the background predicted by a number of assembly models and even upper limits would already be able to rule out selected scenarios, constraining the MBH formation history. The planned SKA observatory will detect the background in a broad frequency band, including the steeper part of the spectrum. The timing of 20 pulsars for 10 yr with a $\delta t_{\rm rms} \simeq 10$ ns should allow us to study the frequency dependency of the GW background in the frequency range 3×10^{-9}–3×10^{-8} Hz, providing unique information about the MBHB demographics at low redshift ($z < 2$). However, the data analysis approach – as well as the number and stability of pulsars available for such studies – will need to be considered in detail to assess precisely the sensitivity level achievable by any given survey. This is beyond the scope of this paper and we have not tried to provide a detailed discussion of these complex issues.

In the frequency range $\sim 10^{-8}$–10^{-7} Hz, the overall GW signal also shows prominent peaks generated by individual sources that are particularly bright and that may be resolved with PTA observations. We have not considered these signals here – our work has been entirely focused on the GW stochastic contribution – but they are potentially of great interest; this is an area that deserves a systematic and quantitative study and we plan to return to this issue in the future.

The population of MBHB systems is a primary target for both PTAs and the *LISA*. We have shown that PTA observations of a GW stochastic background generated from such a population are 'orthogonal' and complementary to *LISA* observations of individual events. PTAs and *LISA* will therefore *jointly* provide unique information about MBH assembly throughout cosmic history.

ACKNOWLEDGMENTS

REFERENCES

Allen B., Romano J. D., 1999, Phys. Rev. D, 59, 102001
Babusci D., Foffa S., Losurdo G., Maggiore M., Matone G., Sturani R., 2001, in Ciufolini I., Gorini V., Moschella U., Fr P., eds, Gravitational Waves. Institute of Physics Publishing, Bristol, UK, p. 179
Barnes E. J., Hernquist L., 1992, ARA&A, 30, 705
Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289
Bell E. F., Phleps S., Somerville R. S., Wolf C., Borch A., Meisenheimer K., 2006, ApJ, 652, 270
Bender P.-L. et al., 1998, LISA Pre-Phase A Report, Second Edition, MPQ 233
Berczik P., Merritt D., Spurzem R., Bischof H. P., 2006, ApJ, 642, 21
Bertotti B., Carr B. J., Rees M. J., 1983, MNRAS, 203, 945
Borch A. et al., 2006, A&A, 453, 869
Burkey J. M., Keel W. C., Windhorst R. A., Franklin B. E., 1994, ApJ, 429, 13
Campanelli M., Lousto C. O., Zlochower Y., Merritt D., 2007, Phys. Rev. Lett., 98, 1102
Carlberg R. et al., 2000, ApJ, 532, 1
Colin P., Schramann D. N., Peimbert M., 1994, ApJ, 426, 459
De Propris R., Conselice C. J., Driver S. P., Liske J., Patton D., Graham A., Allen P., 2007, ApJ, 666, 212
Detweiler S., 1979, ApJ, 234, 1100
Enoki M., Inoue K. T., Nagashima M., Sugiyama N., 2004, ApJ, 615, 19
Ferrarese L., Merritt D., 2000, ApJ, 539, 9
Gebhardt K. et al., 2000, ApJ, 543, 5
Haehnelt M. G., 1994, MNRAS, 269, 199
Haehnelt M. G., Rees M. J., 1993, MNRAS, 263, 168
Haring N., Rix H. W., 2004, ApJ, 604, 89
Hellings R. W., Downs G. S., 1983, ApJ, 265, 39
Herrmann F., Hinder I., Shoemaker D., Laguna P., Matzner R. A., 2007, ApJ, 661, 430
Hopkins P. F., Hernquist L., Cox I. J., Di Matteo T., Robertson B., Springel V., 2005, ApJ, 632, 81
Ilbert O. et al., 2006, A&A, 457, 841
Jaffe A. H., Backer D. C., 2003, ApJ, 583, 616
Jenet F. A., Lommen A., Larson S. L., Wen L., 2004, ApJ, 606, 799
Jenet F. A., Creighton T., Lommen A., 2005a, ApJ, 627, L125
Jenet F. A., Hobbs G. B., Lee K. J., Manchester R. N., 2005b, ApJ, 625, 123
Jenet F. A. et al., 2006, ApJ, 653, 1571
Koushiappas S. M., Bullock J. S., Dekel A., 2004, MNRAS, 354, 29
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lauer T. R., Tremaine S., Richstone D., Faber S. M., 2007, ApJ, 670, 249
Lin L. et al., 2004, ApJ, 617, 9
Lodato G., Natarajan P., 2006, MNRAS, 371, 1813
Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169
McLure R. J., Dunlop J. S., 2002, MNRAS, 331, 795
McLure R. J., Jarvis M. J., Targett T. A., Dunlop J. S., Best P. N., 2006, MNRAS, 368, 1395
Magorrian J. et al., 1998, AJ, 115, 2285
Manchester R. N., 2008, in Bassa C., Wang Z., Cumming A., Kaspi V. M., eds, Proc. AIP Conf. Vol. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More. Am. Inst. Phys., New York, p. 584
Marulli F., Crociani D., Volonteri M., Branchini E., Moscardini L., 2006, MNRAS, 368, 1269
Naab T., Khochfar S., Burkert A., 2006, ApJ, 636, 81
Patton D. R. et al., 2002, ApJ, 565, 208
Peng C. Y., Impey C. D., Ho L. C., Barton E. J., Rix H. W., 2006, ApJ, 640, 114
Peebles P. J. E., 1982, ApJ, 263, 1
Phinney E. S., 2001, preprint (astro-ph/0108028)
Press W. H., Schechter P., 1974, ApJ, 187, 425
Rajagopal M., Romani R. W., 1995, ApJ, 446, 543
Redmount I. H., Rees M. J., 1989, Comments Astrophys., 14, 165
Richstone D. et al., 1998, Nat, 395, 14
Robertson B., Hernquist L., Cox T. J., Di Matteo T., Hopkins P. F., Martini P., Springel V., 2006, ApJ, 641, 90
Sazhin M. V., 1978, Soviet Astron., 22, 36
Schnittman J. D., Buonanno A., 2007, ApJ, 662, 63
Sesana A., 2007, MNRAS, 382, 6
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., Volonteri M., Haardt F., 2007a, MNRAS, 377, 1711
Sesana A., Haardt F., Madau P., 2007b, ApJ, 660, 546
Sheth R. K., Tormen G., 1999, MNARS, 308, 119
Sheth R. K. et al., 2003, ApJ, 594, 225
Shields G. A., Gebhardt K., Salviander S., Wills B. J., Xie B., Brotherton M. S., Yuan J., Dietrich M., 2003, ApJ, 583, 124
Soltan A., 1982, MNRAS, 200, 115
Thorne K. S., 1987, in Hawking S., Israel W., eds, 300 Years of Gravitation. Cambridge Univ. Press, Cambridge, p. 330
Tichy W., Marronetti P., 2007, Phys. Rev. D, 76, 1502
Tremaine S. et al., 2002, ApJ, 574, 740
Treu T., Malkan M. A., Blandford R. D., 2004, ApJ, 615, 97

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

Volonteri M., Salvaterra R., Haardt F., 2006, MNRAS, 373, 121

White S. D. M., Rees M. J., 1978, MNRAS, 310, 645

Woo J.-H., Treu T., Malkan M. A., Blandford R. D., 2006, ApJ, 645, 900

Wyithe J. S. B., Loeb A., 2003, ApJ, 590, 691

Zibetti S., Gavazzi G., Scodeggio M., Franzetti P., Boselli A., 2002, ApJ, 579, 261

This paper has been typeset from a TEX/LATEX file prepared by the author.