

# Exploring Proxies for the Supermassive Black Hole Mass Function: Implications for Pulsar Timing Arrays

Joseph Simon<sup>1</sup> 

Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO 80309, USA

Received 2022 September 20; revised 2023 May 1; accepted 2023 May 2; published 2023 May 30

## Abstract

Supermassive black holes (SMBHs) reside at the center of every massive galaxy in the local universe with masses that closely correlate with observations of their host galaxy, implying a connected evolutionary history. The population of binary SMBHs, which form following galaxy mergers, is expected to produce a gravitational-wave background (GWB) detectable by pulsar timing arrays (PTAs). PTAs are starting to see hints of what may be a GWB, and the amplitude of the emerging signal is toward the higher end of model predictions. Simulated populations of binary SMBHs can be constructed from observations of galaxies and are used to make predictions about the nature of the GWB. The greatest source of uncertainty in these observation-based models comes from the inference of the SMBH mass function, which is derived from observed host galaxy properties. In this paper, I undertake a new approach for inferring the SMBH mass function, starting from a velocity dispersion function rather than a galaxy stellar mass function. I argue that this method allows for a more direct inference by relying on a larger suite of individual galaxy observations as well as relying on a more “fundamental” SMBH mass relation. I find that the resulting binary SMBH population contains more massive systems at higher redshifts than previous models. Additionally, I explore the implications for the detection of individually resolvable sources in PTA data.

*Unified Astronomy Thesaurus concepts:* Supermassive black holes (1663); Galaxy mergers (608); Gravitational waves (678); Gravitational wave sources (677)

## 1. Introduction

Supermassive black holes (SMBHs), with masses of  $10^6$ – $10^{10} M_{\odot}$ , reside in the nuclei of most, if not all, nearby galaxies. Shared evolution likely gives rise to the close correlation between the mass of the SMBH and several global properties of the host galaxy (Kormendy & Richstone 1995). Following a galaxy merger, the SMBHs from each galaxy sink to the center of the common merger remnant through interactions with the galactic gas, stars, and dark matter (Begelman et al. 1980). Once the SMBHs become gravitationally bound, they will emit strong gravitational waves (GWs) in the nanohertz frequency band.

Utilizing a hierarchical framework of galaxy formation, many theoretical models of SMBH evolution have been used to infer the binary SMBH population and the resulting GW background (GWB; e.g., Rajagopal & Romani 1995; Jaffe & Backer 2003; Wyithe & Loeb 2003; Sesana et al. 2008; Chen et al. 2019). Full cosmological simulations have also been used to great effect from the Millennium simulation (Sesana et al. 2009; Ravi et al. 2012) to the Illustris simulation (Kelley et al. 2017; Siwek et al. 2020). In recent years, focus has turned toward modeling the GWB based on the plethora of well-constrained galaxy observations (Sesana 2013; McWilliams et al. 2014; Ravi et al. 2015; Mingarelli et al. 2017). Simon & Burke-Spoliar (2016, hereafter SB16) undertook an in-depth analysis of how the uncertainty in galaxy observations propagated into the resulting GWB prediction and found that

the inference of the SMBH mass function from host galaxy parameters had the largest impact.

Over the past few decades, precision timing of millisecond pulsars has allowed the creation of a galactic-scale GW observatory, a pulsar timing array (PTA), which is sensitive to GWs at nanohertz frequencies (Perera et al. 2019; Alam et al. 2021). The brightest signature detectable by PTAs is the GWB, which results from the incoherent superposition of GWs from SMBH binaries that form in post-merger galaxies. While the first stages of galaxy mergers have been repeatedly identified through electromagnetic observations (e.g., Duncan et al. 2019; Stemo et al. 2020), bound SMBH binaries, at subparsec separations, remain elusive. In fact, PTAs offer one of the most direct avenues to observing bound SMBH binaries (Burke-Spoliar et al. 2019).

Recently, new PTA data sets are uncovering a signal that may be the first hints of the GWB (Arzoumanian et al. 2020; Chen et al. 2021; Goncharov et al. 2021; Antoniadis et al. 2022). If the signal is of an astrophysical origin, then confirmation of the signal’s GW nature is expected in the next year or two (Pol et al. 2021). The emerging signal is toward the higher end of predicted amplitudes (Rosado et al. 2015; Kelley et al. 2017; Chen et al. 2019), which may indicate that binary SMBHs are more massive than previously thought (Middleton et al. 2021) and/or that the local number density of SMBHs is larger than expected (Casey-Clyde et al. 2022). To help understand these implications, it is worth revisiting the inference of the SMBH mass function from galaxy observations.

Inference of the SMBH mass function is predicated on measurements of host galaxy relationships in the local universe (e.g.,  $M_{\bullet}$ – $M_{\text{bulge}}$  and  $M_{\bullet}$ – $\sigma$ ). Previous work has used observations of the galaxy stellar mass function (GSMF) as the basis for the host galaxy relation used to infer the SMBH mass

<sup>1</sup> NSF Astronomy and Astrophysics Postdoctoral Fellow.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

function (Sesana 2013; SB16); however, this method requires some assumption to be made about the fraction of the total stellar mass of the galaxy contained in the bulge,  $f_{\text{bulge}}$ . While the majority of contributions to the GWB is expected to be from bulge-dominated, early-type galaxies where  $f_{\text{bulge}} \approx 1$  (Sesana 2013), it is increasingly unclear how galaxy morphology tracks SMBH mass at higher redshift (Chen et al. 2020), where galaxies tend to be smaller and clumpier (Conselice 2014).

A more direct inference could be made using a galaxy's stellar velocity dispersion ( $\sigma$ ), which is measured from a galaxy's spectra. In addition to removing the need for an  $f_{\text{bulge}}$  assumption, recent work suggests that  $M_{\bullet}-\sigma$  is a more “fundamental” relationship than  $M_{\bullet}-M_{\text{bulge}}$  (e.g., Kormendy & Ho 2013; van den Bosch 2016; de Nicola et al. 2019; Marsden et al. 2020) and therefore more accurately predicts the SMBH mass. In particular, the discovery of relic galaxies, such as NGC1271 and NGC1277 (Walsh et al. 2015, 2016), which formed at  $z > 2$  and have evolved passively ever since (Ferré-Mateu et al. 2015), points to the limitations of the  $M_{\bullet}-M_{\text{bulge}}$  relation at higher redshifts. These relic galaxies are much smaller than most present-day early-type galaxies, which is consistent with the strong redshift evolution of galaxy size (van der Wel et al. 2014); however, they host extremely massive SMBHs. While these galaxies are significant outliers on the  $M_{\bullet}-M_{\text{bulge}}$  relation (Ferré-Mateu et al. 2015), they are consistent within the intrinsic scatter for the  $M_{\bullet}-\sigma$  relation (van den Bosch 2016).

Until recently, there were no spectroscopic surveys dedicated to measuring the galaxy velocity dispersion function (VDF) at higher redshifts; however, the Large Early Galaxy Astrophysics Census (LEGA-C) survey now provides the first direct spectroscopic measurement of the VDF from  $0.6 < z < 1.0$  (Taylor et al. 2022). The LEGA-C result is consistent with earlier attempts to indirectly measure the VDF using an *inferred* velocity dispersion, calculated from photometric data (Bezanson et al. 2011, 2012), especially at large  $\sigma$ .

Previous attempts to use the  $M_{\bullet}-\sigma$  relation in GWB calculations have started directly from the GMSF and used a galaxy stellar mass to  $\sigma$  conversion derived from observations in the local universe (Sesana 2013; Sesana et al. 2016); however, these results neglected to include the strong redshift evolution in the galaxy size–mass relation (van der Wel et al. 2014). Thus, by incorporating more observational data, the VDFs used in this work provide fundamentally different starting points for calculating the SMBH mass function than anything that has been attempted before.

In this paper, I will use the VDFs from both Bezanson et al. (2012) and Taylor et al. (2022) as the basis for inferring the SMBH mass function and compare the resulting binary SMBH population and GWB signal to that from the standard GMSF inference method. In Section 2 I describe my methods, including how to infer the SMBH mass function with these different approaches. I show my results in Section 3 and close with a summary.

## 2. Methods

Following SB16, this paper calculates the characteristic strain spectrum  $h_c^2(f)$  from a cosmic population of binary SMBHs by integrating over all the sources emitting in an observed GW frequency bin  $df$  multiplied by the square strain

of each source in that bin (e.g., Sesana et al. 2008),

$$h_c^2(f) = f \int \int \int dz \, dM_{\bullet} \, dq_{\bullet} \, h_s^2 \frac{d^4N}{dz \, dM_{\bullet} \, dq_{\bullet} \, df}, \quad (1)$$

where  $d^4N$  is explicitly the number of binaries in a given redshift range  $dz$ , primary black hole mass range  $dM_{\bullet}$ , and black hole mass ratio range  $dq_{\bullet}$ , which are emitting in a given Earth-observed GW frequency range  $df$ , and  $h_s$  is the polarization- and sky-averaged strain contribution from each binary (e.g., Thorne 1987),

$$h_s = \sqrt{\frac{32}{5}} \left( \frac{GM_c}{c^3} \right)^{5/3} \frac{(\pi f_r)^{2/3} c}{D_c}, \quad (2)$$

where  $M_c = M_{\bullet}(q_{\bullet}^3/(1+q_{\bullet}))^{1/5}$  is the chirp mass of the binary,  $D_c$  is the proper (comoving) distance to the binary, and  $f_r$  is the frequency of the GWs emitted in the rest frame of the binary. The Earth-observed GW frequency  $f = f_r/(1+z)$ .

The number of binaries  $d^4N$  is directly related to the comoving number density per unit redshift, primary black hole mass, and binary mass ratio,

$$\frac{d^4N}{dz \, dM_{\bullet} \, dq_{\bullet} \, df} = \frac{d^3n}{dz \, dM_{\bullet} \, dq_{\bullet}} \frac{dV_c}{dz} \frac{dz}{dt} \frac{dt}{df}. \quad (3)$$

The conversion above, Equation (3), takes the number of binaries per comoving volume shell,  $dV_c$ , and converts it to the number of binaries per Earth-observed GW frequency bin,  $df$ , by first converting to redshift and then to Earth-observed time. Once the binary hardens and decouples from the surrounding galactic environment, the orbital changes of the system become dominated by the emission of gravitational radiation at a rate of (Peters & Mathews 1963),

$$\frac{df_{\text{orb}}}{dt} = \frac{96}{5} \left( \frac{GM_c}{c^3} \right)^{5/3} (2\pi)^{8/3} f_{\text{orb}}^{11/3}. \quad (4)$$

For binaries in a circular orbit, the frequency of GWs emitted in the rest frame of the binary  $f_r = 2f_{\text{orb}}$ .

The frequency dependence of  $h_c$  is captured in both  $h_s$  and  $dt=df$ ; therefore, by combining Equations (1)–(4),  $h_c(f)$  can be written as a simple power law with dimensionless amplitude  $A_{\text{yr}}$  referenced to the frequency of an inverse year  $f_{\text{yr}} = 1 \text{ yr}^{-1}$  (Jenet et al. 2006):

$$h_c(f) = A_{\text{yr}} \left( \frac{f}{f_{\text{yr}}} \right)^{-2/3}. \quad (5)$$

However, this only applies for circular binaries in GW-radiation-driven inspirals. If the inspirals were instead driven by some environmental coupling in the lowest frequencies of the PTA band, this would cause a turnover in the power-law spectral shape (Sampson et al. 2015). Environmental coupling can also impact the eccentricity of the binary and therefore the spectral shape of the characteristic strain spectrum in the PTA band (Ravi et al. 2014; Huerta et al. 2015). For the purposes of this project, where I am solely focused on the SMBH mass function, I ignore the specifics of environmental coupling as well as eccentricity and assume that all binaries remain in circular orbits while emitting GWs in the PTA band.

### 2.1. Determining the Number Density of Binary SMBHs

The number density of binary SMBHs is a combination of two functions: the SMBH mass function  $\Phi_*(z, M_*)$  and the binary SMBH merger rate  $\mathcal{R}_*(z, M_*, q_*)$ ,

$$\frac{d^3n_*}{dz dM_* dq_*} = \Phi_*(z, M_*) \mathcal{R}_*(z, M_*, q_*). \quad (6)$$

There are currently no direct observational constraints on the demographics of binary SMBHs, so these functions are not explicitly known. Instead, they need to be inferred from other data. As in SB16, I use the demographics of galaxy mergers as a proxy.

The binary SMBH merger rate is directly related to the galaxy merger rate,  $\mathcal{R}_* \sim \mathcal{R}_{\text{Gal}}$ , but offset in redshift due to the binary evolution timescale. The standard way to calculate the galaxy merger rate,  $\mathcal{R}_{\text{Gal}}$ , directly from observations is by combining a galaxy close pair fraction,  $f_{\text{pair}}$ , with a merger timescale,  $\tau$ . The galaxy close pair fraction is an astronomical observable, while the merger timescale, which approximates the amount of time galaxy close pairs are observable, must be determined through simulations (SB16).

The SMBH mass function,  $\Phi_*$ , is more difficult to derive and needs to be indirectly inferred by populating each galaxy with a central SMBH assuming some relationship between SMBHs and their host galaxies (e.g.,  $M_*$ - $\sigma$  relation,  $M_*$ - $M_{\text{bulge}}$  relation, etc.). The standard approach to inferring the SMBH mass function from galaxy surveys is to start from an observed GSMF, then make some assumption about the fraction of the total stellar mass of the galaxy contained in the bulge,  $f_{\text{bulge}}$ , and finally calculate the SMBH mass from the  $M_*$ - $M_{\text{bulge}}$  relation (Sesana 2013; Ravi et al. 2015; SB16). Using this approach, the number density of binary SMBHs is directly inferred from the number density of galaxy mergers,

$$\frac{d^3n_*}{dz dM_* dq_*} = |\Phi_{\text{Gal}} \mathcal{R}_{\text{Gal}}|_{\text{Gal} \rightarrow *}, \quad (7)$$

where  $(\text{Gal} \rightarrow *)$  represents the step where the SMBH is populated into the galaxy assuming some galaxy number density parameterization  $\Phi_{\text{Gal}}$ . When using a GSMF,  $\Phi_{\text{Gal}} = \Phi_{\text{GSMF}}(z, M)$ , where  $M$  is the total stellar mass of the primary galaxy. The companion galaxy then has a mass of  $q_M M$ , where  $0.25 < q_M < 1$  is set to capture major merger events, and the  $(\text{Gal} \rightarrow *)$  step in Equation (7) utilizes the  $M_*$ - $M_{\text{bulge}}$  relation and includes the determination of  $f_{\text{bulge}}$  for each galaxy.

While the specific prescription for  $f_{\text{bulge}}$  can change the predicted amplitude of the GWB by a factor of 2 or more, the largest source of uncertainty in this inference of the SMBH mass function inference comes from uncertainty in the GSMF itself (SB16). This is to be expected given discrepancies over the number density of the most massive galaxies in the local universe (Bernardi et al. 2013). In addition to the issues with properly identifying massive galaxies, there is increasing evidence that SMBH evolution and bulge evolution track differently at higher redshifts ( $z > 1$ ; Chen et al. 2020); however, there is no significant evidence that these different evolutionary tracks are caused by a change in the relationship between SMBH mass and other host galaxy properties (Shen et al. 2015; Suh et al. 2020). Ideally, one would like a more robust method for

inferring the SMBH mass function, particularly something that does not require an  $f_{\text{bulge}}$  determination and does not have as much dependency on the underlying mass-to-light (M/L) model used to determine stellar mass (Bernardi et al. 2013).

The stellar velocity dispersion ( $\sigma$ ) appears to allow for a more direct inference of SMBH mass. Recently, it has been suggested that  $\sigma$  is a more fundamental property of a galaxy than it is stellar mass (van den Bosch 2016; de Nicola et al. 2019; Marsden et al. 2020) and that the  $M_*$ - $M_{\text{bulge}}$  relation is just an extension of the more fundamental  $M_*$ - $\sigma$  relation (Wake et al. 2012). However, it has been difficult to obtain VDFs, specifically at higher redshifts ( $z > 0.3$ ), due to a lack of complete spectroscopic surveys to use as a separate starting point for inferring the SMBH mass function. Instead, when  $\sigma$  has been explored in the context of simulating the binary SMBH population, it has been inferred from galaxy total stellar mass, and therefore, the results have unsurprisingly been consistent (Sesana 2013; Sesana et al. 2016). Instead, in an effort to reduce uncertainty in the inferred SMBH mass function, we are pursuing a novel approach to modeling  $M_*$  using the VDFs from both Bezanson et al. (2012), which are derived from an *inferred* velocity dispersion ( $\sigma_{\text{inf}}$ ), and Taylor et al. (2022), which are derived from spectroscopic observations at  $0.6 < z < 1.0$ .

### 2.2. The Velocity Dispersion Function

When using a VDF in Equation (7),  $\Phi_{\text{Gal}} = \Phi_{\text{VDF}}(z, \sigma)$ , where  $\sigma$  is the velocity dispersion of the primary galaxy. The companion galaxy then has a velocity dispersion of  $q_\sigma \sigma$ , where  $q_\sigma$  is set to capture major mergers the same way that  $q_M$  was in the case of GSMFs. Finally, the  $(\text{Gal} \rightarrow *)$  step in Equation (7) now utilizes the  $M_*$ - $\sigma$  relation and no longer requires any additional steps as it did in the case of GSMFs. Taylor et al. (2022) measures the VDF directly through spectroscopic observations. Bezanson et al. (2011) use a suite of photometric data to calculate an *inferred* VDF, which is shown to be comparable with a spectroscopic measurements in the local universe.

In Bezanson et al. (2011), the *inferred* velocity dispersion,  $\sigma_{\text{inf}}$ , is measured from a combination of a galaxy's photometric properties. Starting from the virial theorem and building from the observational evidence that for local galaxies stellar mass is proportional to dynamical mass (Taylor et al. 2010), the central velocity dispersion,  $\sigma$ , of a galaxy can be effectively predicted based on its size (effective radius,  $r_e$ ), shape (Sérsic index,  $n$ ), and total stellar mass ( $M_*$ ).

$$\sigma_{\text{inf}} = \sqrt{\frac{G M_*}{K_*(n) r_e}}, \quad (8)$$

where  $K_*(n) \equiv K_v(n)(M_*/M_{\text{dynamical}})$ .  $K_v(n)$  is a Sérsic-dependent virial constant (Bertin et al. 2002), and since Taylor et al. (2010) have shown that  $K_*(n)$  only weakly depends on mass, the average ratio of stellar to total mass is adopted and calibrated such that the median  $\sigma_{\text{inf}}$  equals the median-measured  $\sigma$  in SDSS data.

Bezanson et al. (2011) finds a scatter in the  $\sigma_{\text{inf}}-\sigma$  relationship of 0.06 dex, which has the effect of increasing the high- $\sigma$  end of the  $\sigma_{\text{inf}}$  distribution. This scatter must be properly taken into account when inferring the “true” VDF in order to not overpredict the number density of massive

galaxies. Numerical fits for the “true” VDF were not published in either Bezanson et al. (2011) or Bezanson et al. (2012) and instead were obtained through private communication with the author (R. Bezanson 2016, private communication). The fits are done using a modified Schechter function, and given the strong correlation between  $\sigma_*$ ,  $\alpha$ , and  $\beta$  (Sheth et al. 2003), uncertainties on the fit were obtained by setting  $\alpha$  and  $\beta$  to their maximum likelihood values and allowing  $\sigma_*$  to absorb the combined uncertainties in those three parameters.

### 2.3. Observational Constraints

Unlike previous work (e.g., SB16), this paper is focused on keeping as many things constant in the calculation of the GWB as possible in order to highlight the differences in the SMBH mass function and the implications for PTA experiments. As such, this work only uses the galaxy close pair fraction from Keenan et al. (2014) in its calculation of  $\mathcal{R}_{\text{Gal}}$ , as opposed to Robotham et al. (2014) since SB16 showed them to be consistent. For  $\tau$ , the same formulation as SB16 is used, which combines the mass and redshift dependence from Kitzbichler & White (2008) with the results from Lotz et al. (2011).

Bezanson et al. (2012) calculates the VDF at  $0.3 < z < 1.5$  using photometric measurements of galaxy stellar mass, effective radii, and Sérsic index for the galaxies in the NMBS COSMOS (Whitaker et al. 2011) and the UKIDSS Ultra-Deep Survey fields (Williams et al. 2009). In an attempt to make sure that the galaxy samples are as consistent as possible, this paper compares the VDF derived in Bezanson et al. (2012) to the GSMF from Tomczak et al. (2014) since it is partially based on NMBS COSMOS. It is important to also note that the stellar mass determination in Bezanson et al. (2012) and Tomczak et al. (2014) uses the same software pipeline, which should reduce the amount of bias inserted into these calculations from different M/L determinations. When using the spectroscopic VDF from Taylor et al. (2022) for  $0.6 < z < 1.0$ , the *inferred* VDF is used at  $z > 1$  since the *inferred* VDF matches the spectroscopic VDF.

For inference of the SMBH mass function at  $z < 0.3$ , this paper uses the GSMF in Moustakas et al. (2013) calculated from SDSS data since that is what was used in SB16, and it was also used as the local universe comparison in Tomczak et al. (2014). To minimize introducing additional bias, the VDF from Bernardi et al. (2010) is used at  $z < 0.3$  since it was directly derived from similar SDSS data as Moustakas et al. (2013).

To infer the SMBH in each galaxy, the  $M_{\bullet}$ – $M_{\text{bulge}}$  relation is used with the GSMF while the  $M_{\bullet}$ – $\sigma$  relation is used with the VDF. The one place this paper deviates from holding many things constant is when determining which  $M_{\bullet}$ – $M_{\text{bulge}}$  and  $M_{\bullet}$ – $\sigma$  relations to use since this parameter has been shown to have the largest impact on the predicted amplitude of the GWB (SB16). Thus, relations are used from McConnell & Ma (2013, hereafter MM13), Kormendy & Ho (2013, hereafter KH13), and de Nicola et al. (2019, hereafter dN19) as these papers are representative of the range of values these relations have been shown to have. It is worth noting that the range of measurements for the  $M_{\bullet}$ – $\sigma$  relation is smaller than that for the  $M_{\bullet}$ – $M_{\text{bulge}}$  relation, adding further evidence for the  $M_{\bullet}$ – $\sigma$  relation to be considered more “fundamental” than the  $M_{\bullet}$ – $M_{\text{bulge}}$ .

### 2.4. Calculating the GWB

As in SB16, Equations (1), (3), and (7) are combined to calculate  $h_c$ . When using a GSMF,

$$h_c^2 = f \int \int \int \frac{\Phi_{\text{GSMF}}}{\tau} \frac{df_{\text{pair}}}{dq_M} \frac{dV_c}{dz} \left( \frac{dt}{df} h_s^2 \right) dz \, dM \, dq_M. \quad (9)$$

When using the VDF, as described in Section 2.2,

$$h_c^2 = f \int \int \int \frac{\Phi_{\text{VDF}}}{\tau} \frac{df_{\text{pair}}}{dq_{\sigma}} \frac{dV_c}{dz} \left( \frac{dt}{df} h_s^2 \right) dz \, d\sigma \, dq_{\sigma}. \quad (10)$$

To calculate  $A_{\text{yr}}^2$ , the above integrals are used setting  $f = f_{\text{yr}}$ .

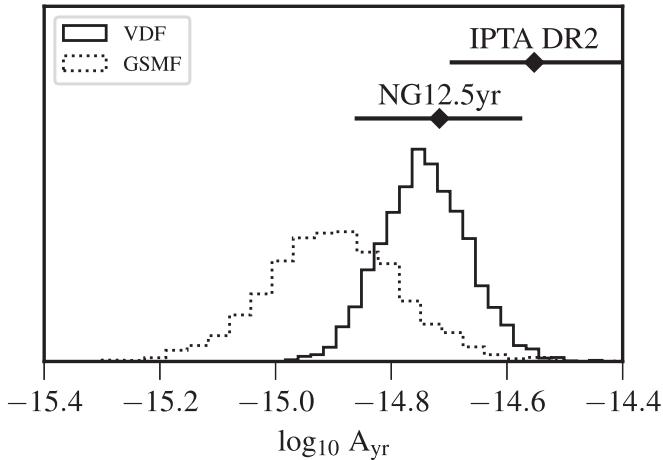
The limits of integration for the above equations are chosen in order to capture the major galaxy mergers in the local universe, which make up the majority of contributing systems to the GWB in the PTA band. For the GSMF (Equation (9)) in SB16, these are shown to be  $0 < z < 1.5$ ,  $10 < \log_{10} M_{\text{Gal}} < 12.5$ , and  $0.25 < q < 1$ . When using the VDF, it is imperative to ensure that the integration happens over a similar range of galaxy mergers. To do so, this work uses the size–mass relation from Williams et al. (2010) because it is based on the same UKIDSS sample that the VDF in Bezanson et al. (2012) is derived from. The corresponding limits of integration for the VDF are then  $0 < z < 1.5$ ,  $1.85 < \log_{10} \sigma < 2.6$ , and  $0.67 < q_{\sigma} < 1$ . While these ranges are smaller, it is important to remember that the corresponding range of  $\sigma$  values is much narrower for a given range of  $M$ . Integration limits derived from the size–mass relation in Shen et al. (2003), which is based on SDSS data, were also tested but produced almost no change in the VDF predictions for  $h_c$ .

## 3. Results

Combining all of the observational constraints from Section 2.3 into the models in Equations (9) and (10), I calculate predictions for the GWB. For  $A_{\text{yr}}$  predictions, I generate 1000 realizations each for the VDF and GSMF under each SMBH–host galaxy relation. In each realization, all parameters are allowed to vary within their observed uncertainties.

### 3.1. Predicted Amplitude of the GWB

Figure 1 shows the relative distributions predicted for  $A_{\text{yr}}$  from the two different methods (GSMF and VDF) employed in this paper. The predictions based on the VDF are both at a higher amplitude than those from the GSMF and cover a smaller spread, implying that the underlying binary SMBH populations inferred by these two methods are quite different. As discussed in Section 2.3, I have endeavored to minimize any differences in observational inputs for these predictions in order to highlight that the subsequent  $A_{\text{yr}}$  differences are due primarily to differences in the method used to infer the SMBH mass function (discussed in Section 2.1) and not simply due to using different observational inputs to the model. The larger values of  $A_{\text{yr}}$  predicted by the VDF are more in line with the amplitude of the common process starting to appear in PTA data; however, both methods are able to reproduce the signal emerging in PTA data sets. To highlight this, Figure 1 also includes the  $A_{\text{yr}}$  constraints from both the NANOGrav 12.5 year data set (Arzoumanian et al. 2020), as well as IPTA DR2 (Antoniadis et al. 2022). Yet, without the requisite spatial



**Figure 1.**  $A_{\text{yr}}$  predictions calculated using the VDF (solid) and GSMF (dotted) methods to infer the SMBH mass function. The constraints from both the NANOGrav 12.5 year data set (Arzoumanian et al. 2020) and IPTA DR2 (Antoniadis et al. 2022) are shown above the two histograms. Each histogram includes 3000 predictions, with 1000 coming from each of the three SMBH—host galaxy relations used. The 90% interval of the VDF predictions cover 0.24 dex with a mean at  $\log_{10} A_{\text{yr}} = -14.74$ , while the GSMF predictions cover 0.4 dex over the 90% interval with a mean at  $\log_{10} A_{\text{yr}} = -14.9$ . Both methods are able to reproduce the emerging signal, but the VDF predictions tend toward higher amplitudes than the GSMF predictions.

correlations, the origin of the signal appearing in PTA data remains unclear.

Interestingly, the VDF predictions for  $A_{\text{yr}}$  appear to be less sensitive to the choice of SMBH—host galaxy relation than the GSMF method. In Figure 2, the three panels show the  $A_{\text{yr}}$  predictions for each SMBH—host galaxy relation used in this work. In all three cases, the VDF predictions are nearly identical and almost perfectly match the combined distribution seen in Figure 1. In contrast, the mean value of the GSMF prediction changes significantly depending on which SMBH—host galaxy relation is utilized. This behavior of the GSMF predictions has been previously noted in SB16.

One of my aims for pursuing this new method of inferring the SMBH mass function was to reduce the uncertainty on  $A_{\text{yr}}$  that is present under the GSMF method. Some of the reduction in uncertainty likely comes from removing the intermediate step of calculating  $f_{\text{bulge}}$  for each galaxy.  $f_{\text{bulge}}$  covered a broad range of values due to limited observational evidence; however, most systems that strongly contribute to the GWB likely have  $f_{\text{bulge}} \sim 1$  (Sesana 2013; Kelley et al. 2017). Some of the reduction in uncertainty is due to the narrower spread of measurements for the  $M_{\bullet}-\sigma$  relation in comparison to the  $M_{\bullet}-M_{\text{bulge}}$  relation. Either way, it appears that by including observations of both a galaxy’s size and Sérsic index along with stellar mass in the inference of the central SMBH, the VDF provides a more direct avenue to the SMBH mass function.

To help understand where the differences in the  $A_{\text{yr}}$  predictions are coming from, Figure 3 shows the 2D differential contribution to  $A_{\text{yr}}^2$  from  $M_c$  and  $z$ . The most striking difference is the contribution to  $A_{\text{yr}}^2$  that the VDF receives at higher redshifts ( $z > 0.5$ ). Figure 3 also shows that the additional contribution at higher redshift is present regardless of which SMBH—host galaxy relation is used with the VDF method. Additionally, one sees that the different SMBH—host galaxy relations impact the predicted SMBH masses from each method in opposite ways. For instance, the

relations in MM13 predict higher SMBH masses when using  $M_{\bullet}-\sigma$  than when using  $M_{\text{bulge}}$ , and that is reversed when the relations in KH13 are used, with the GSMF method producing more massive SMBH systems, especially in the local universe ( $z < 0.2$ ). It is obvious just from looking at the different scaling in the color maps between the two GSMF models that the predicted amplitude using MM13 is lower than KH13. This is in contrast to the similarities in the amplitude predictions from the VDF models, which can be seen in Figure 2. Overall, one can infer that the SMBH mass function implied by the VDF includes more massive SMBHs at higher redshifts than the mass function implied by the GSMF.

The results from using the VDF model are similar to those found in Casey-Clyde et al. (2022), which uses a quasar-based model to predict the binary SMBH population. In that case, the SMBH mass function has more support at larger mass values than the GSMF models from Sesana (2013) that they compare to, the amplitude of the GWB is more in line with the NANOGrav 12.5 yr data set, and the redshift contribution from the quasar-based model is more evenly distributed like the VDF model shown here.

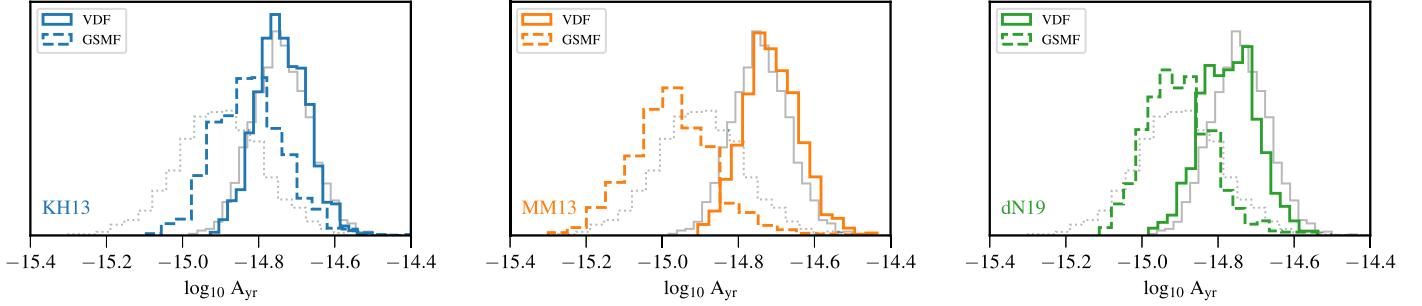
### 3.2. Spectroscopic versus Inferred VDF

As discussed in Section 2.2, this paper primarily uses the *inferred* VDF from Bezanson et al. (2012); however, given the recent publication of the spectroscopic VDF observed with the LEGA-C survey (Taylor et al. 2022), I calculate  $A_{\text{yr}}$  using both functions. Figure 4 shows the relative distributions predicted for  $A_{\text{yr}}$  under each model; however, unlike earlier comparisons, this one only uses the dN19 SMBH—host galaxy relation. Because spectroscopic observations are only available for  $z < 1$ , I use the same *inferred* VDF functions for  $1 < z < 1.5$ ; however, Figure 3 shows that this redshift range is not a significant contributor to the overall amplitude of the GWB.

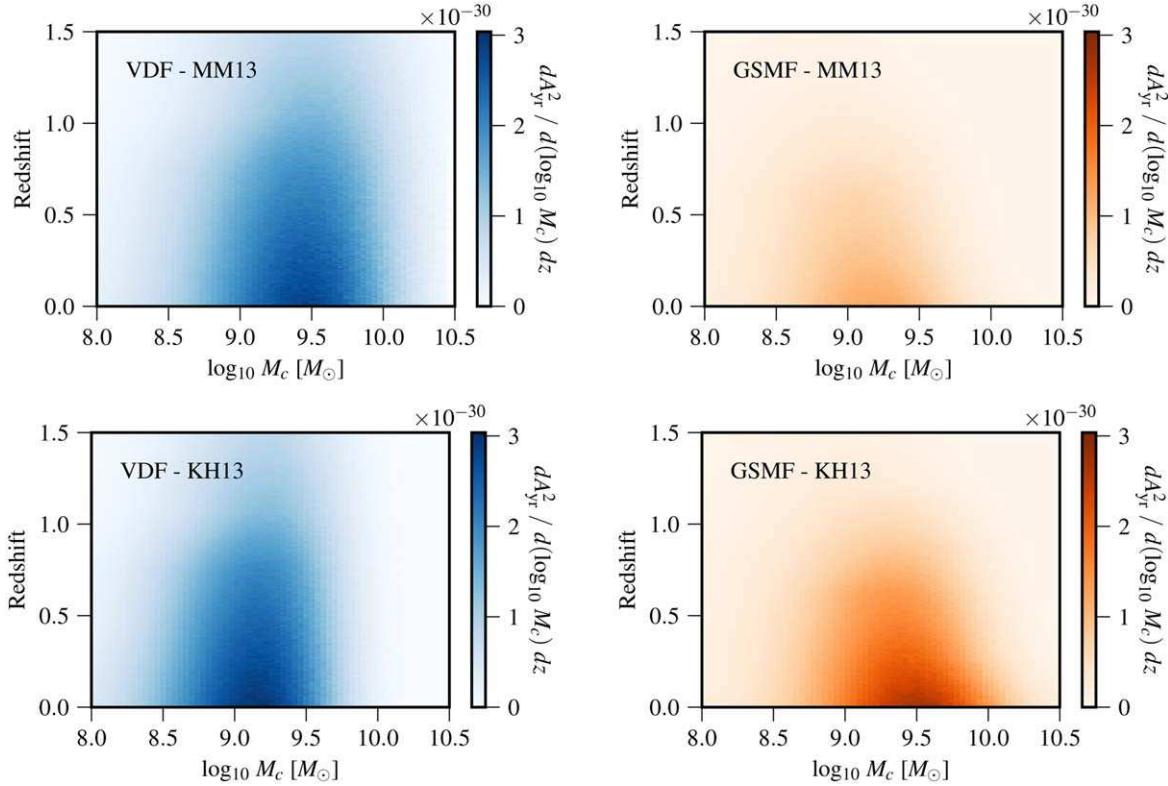
The consistency between the spectroscopic and the *inferred* VDF predictions is promising, implying that the *inferred* VDF is a good proxy for the spectroscopic VDF. As shown in the LEGA-C survey results (Taylor et al. 2022), the spectroscopic VDF is on the higher end of the error region covered by the *inferred* VDF. Thus, it is not surprising that the spectroscopic VDF predictions are consistent with the higher end of the *inferred* VDF predictions, and the smaller spread is directly attributable to the smaller error bars on the spectroscopic VDF.

### 3.3. Prevalence of Individual Sources

Beyond the amplitude of the GWB, the presence of more massive systems at higher redshifts predicted by the VDF method also has implications for the prevalence of individually resolvable, or continuous-wave (CW), sources in PTA data sets. These systems are unique from the GWB, which is compromised of the GW contributions from the entire population of SMBH binary systems in the universe. CWs are individual systems, and once one is detected, it will provide a unique multimessenger opportunity to study a binary SMBH system. The CWs detectable by PTAs are different than those detectable by other GW experiments because these systems are typically a long way away from coalescence, and so their frequency evolution should be negligible, allowing PTAs to probe a part of the SMBH binary evolution that is complimentary to other GW experiments.



**Figure 2.**  $A_{\text{yr}}$  predictions calculated using the VDF (solid) and GSMF (dashed) methods to infer the SMBH mass function. Each panel shows the predictions for a different SMBH—host galaxy relation: KH13 is shown in blue in the left panel, MM13 is shown in orange in the center panel, and dN19 is shown in green in the right panel. Additionally, in all three panels, the full distribution of predictions from Figure 1 for all three relations is shown in gray. The VDF predictions appear to not be sensitive to the choice of host galaxy relation in the same way that the GSMF predictions are, and as seen in Figure 1 the VDF predictions are always at a higher amplitude. However, the magnitude of the difference between the VDF and GSMF predictions is dependent on the specific host galaxy relation chosen. The largest difference is seen when using MM13, and the smallest difference is seen when using KH13.

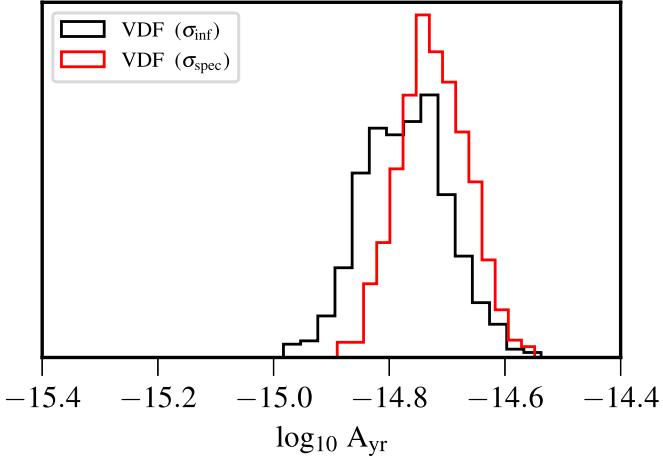


**Figure 3.** 2D differential contribution to  $A_{\text{yr}}^2$  from two different aspects of the simulated binary SMBH population: binary chirp mass ( $M_c$ ) shown on the x-axis and redshift shown on the y-axis. The columns show the VDF inference on the left (in blue) and the GSMF inference on the right (in orange). The rows represent two different SMBH—host galaxy relations with MM13 on the top and KH13 on the bottom. The limits of the color bars are the same in all cases. The biggest difference between the VDF and GSMF is the contribution to the signal at higher redshift ( $z > 0.5$ ). In the case of the GSMF inference, the binaries that contribute significantly to the GWB are contained to the local universe, while the VDF inference shows significant contributions all the way out to  $z \sim 1$  and beyond. One can infer from this that the underlying SMBH mass function implied by the VDF includes more massive systems at higher redshifts than that implied by the GSMF.

Figure 5 shows the redshift distribution of individual binary SMBH systems from 100 different population realizations using the KH13 SMBH—host galaxy relations. I use this host galaxy relation since it predicts the most similar amplitudes of the GWB between the two methods, and CW prevalence is directly linked to GWB amplitude (Rosado et al. 2015). In each case, I have only looked at individual systems that are emitting GWs at a GW frequency close to 8 nHz in order to compare to the expected sky-averaged sensitivity of the IPTA by 2025 (Xin et al. 2021). On average, the VDF method predicts more than twice as many potential CWs as the GSMF, and the redshift distribution of sources is much broader with the VDF

method than with the GSMF. This is to be expected, given the differences in the underlying SMBH populations seen in Figure 3.

The GSMF results shown here are very similar to those in Rosado et al. (2015), which finds that a GWB is more likely to be detected before an individually resolvable source and that the first detected individual binary system is likely to be very close by ( $z \lesssim 0.5$ ). The VDF results are much more optimistic due to the increase in massive SMBH systems at higher redshift. Notably, the VDF method predicts that potentially resolvable sources are fairly evenly distributed across redshift, which is in contrast to other CW predictions, which find that the first



**Figure 4.**  $A_{\text{yr}}$  predictions calculated using the *inferred* VDF ( $\sigma_{\text{inf}}$ , black line) and the spectroscopic VDF ( $\sigma_{\text{spec}}$ , red line) to infer the SMBH mass function. In the case of the  $\sigma_{\text{spec}}$  predictions, spectroscopic data from SDSS (Bernardi et al. 2010) and the LEGA-C survey (Taylor et al. 2022) were used at  $z < 1$ , while the *inferred* VDF was used to fill in from  $1 < z < 1.5$ . Given the consistency that the LEGA-C survey finds between the spectroscopic and *inferred* VDFs (Taylor et al. 2022), it is not surprising that these results are so consistent. While it would be preferred to use only spectroscopic VDFs for this work, the similarity in these results show that the *inferred* VDFs are an acceptable alternative.

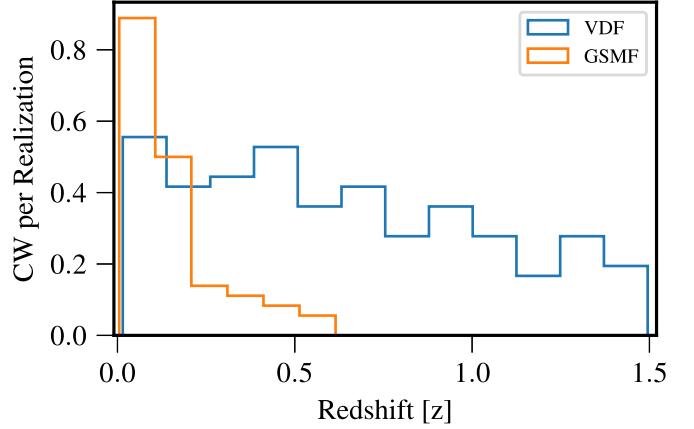
detected system is likely to be very near by Rosado et al. (2015) and Kelley et al. (2018). The larger systems at higher redshifts have intriguing multimessenger potential since if they were to have any variable active-galactic nuclei activity, they would be promising multimessenger candidates given near-future time-domain surveys (Charisi et al. 2022).

#### 4. Summary

In this paper, I undertake a new approach to inferring the SMBH mass function using observations of an *inferred* VDF and compare that to the standard inference method from a GSMF. I also compare the *inferred* VDF to the spectroscopic VDFs that are available within limited redshift ranges. I am careful to use as many of the same observations as possible between the two methods to ensure that the results are solely due to the different method of inference rather than different observational inputs. However, in order to not be biased by the choice of SMBH—host galaxy relation, which is arguably one of the most impactful parameters (Sesana 2013; SB16), I use three separate measurements that span the range of observed values. The VDF method predicts a binary SMBH population that is significantly more massive at higher redshifts than the GSMF method. This population is more in line with observed relic galaxies, like NGC1271 and NGC1277 (Walsh et al. 2015, 2016).

Interestingly, the VDF method produces  $A_{\text{yr}}$  predictions that are at larger amplitudes and cover a narrower spread than the GSMF method although both methods are able to reproduce the amplitude of the emerging signal in PTA data. The narrower spread in the VDF method’s predictions is partially due to the smaller range of  $M_{\bullet} - \sigma$  measurements, compared to the  $M_{\bullet} - M_{\text{bulge}}$  relation. This provides further evidence that the  $M_{\bullet} - \sigma$  relation is a more “fundamental” relation between host galaxy and their SMBHs.

The largest difference between the two methods is from massive binaries at higher redshifts, where the VDF predicts



**Figure 5.** Redshift distribution of individual binary SMBH systems at  $f_{\text{GW}} = 8$  nHz averaged over 100 different population realizations. On average, the VDF method (blue curve) predicts more than twice as many potential individual systems as the GSMF method (orange curve). These systems are expected to be detectable for the IPTA by 2025 (Xin et al. 2021). As seen in previous work, the GSMF method predicts that single sources will be found in the local universe ( $z < 0.5$ ; Rosado et al. 2015). In contrast, the VDF method shows only a slight preference for systems at lower redshift. As seen in Figure 3, the larger-mass systems at higher redshifts that are present in the VDF predictions directly translate to an increase in individually resolvable sources in PTA data for  $z > 0.5$ .

significant contributions, while the GSMF does not. Additionally, the more massive binary SMBHs at higher redshifts predicted by the VDF method increase the number of potential, individually resolvable systems in PTA data as well as broaden the redshift range where those systems are expected to be found. This is in line with recent studies of  $z \sim 1$  galaxy clusters, which appear to host potential CW systems (Wendt & Roman 2023).

It is worth noting that the results presented here are a preliminary investigation into using a VDF to infer the SMBH mass function. The fits to the VDF in Bezanson et al. (2012) are estimates, and a full error analysis has yet to be conducted on the *inferred* VDFs. However, the consistency between the LEGA-C spectroscopic VDF and *inferred* VDFs is a promising development. Additionally, the VDFs used in this paper only went out to  $z = 1.5$ , and as Figure 3 shows, there is reason to think that significant contributions to the GWB from the VDF method may lie at  $z > 1.5$ . Ideally, the VDF would reach out to  $z \sim 2-3$ , and work is currently underway to provide it (Matt et al. 2023).

Finally, I want to reiterate some of the other initial caveats to this work. The model presented here is overly simplistic, looking only at binary SMBHs in circular orbits, ignoring environmental coupling, and assuming a quick, uniform binary evolution timescale. However, these results show that a more direct method of SMBH mass inference is available to binary SMBH populations based on galaxy observations and that the *inferred* SMBH mass function from this new method is fundamentally different than what has been used in the past. While more work is needed, these results hint toward there being a population of more massive SMBHs at higher redshift than previously thought.

I am grateful to the referee, Alberto Sesana, for comments and insights that greatly improved this work. This project was conceived and advanced through fruitful discussions with Sarah Burke-Spolaor, Julie Comerford, Joseph Lazio, Xavier

Siemens, and Sarah Vigeland. The manuscript was improved by comments from Luke Kelley and Kayhan Gürtekin. I am grateful to Rachel Bezanson for providing key insights to the VDF. This work has been supported by the National Science Foundation (NSF) under award 1847938. Additionally, I am supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-2202388.

*Software:* Numpy (Harris et al. 2020), Scipy (Virtanen et al. 2020), MatPlotLib (Hunter 2007).

## ORCID iDs

Joseph Simon  <https://orcid.org/0000-0003-1407-6607>

## References

Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2021, *ApJS*, 252, 4

Antoniadis, J., Arzoumanian, Z., Babak, S., et al. 2022, *MNRAS*, 510, 4873

Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020, *ApJL*, 905, L34

Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, *Natur*, 287, 307

Bernardi, M., Meert, A., Sheth, R. K., et al. 2013, *MNRAS*, 436, 697

Bernardi, M., Shankar, F., Hyde, J. B., et al. 2010, *MNRAS*, 404, 2087

Bertin, G., Ciotti, L., & Del Principe, M. 2002, *A&A*, 386, 149

Bezanson, R., van Dokkum, P., & Franx, M. 2012, *ApJ*, 760, 62

Bezanson, R., van Dokkum, P. G., Franx, M., et al. 2011, *ApJL*, 737, L31

Burke-Spolaor, S., Taylor, S. R., Charisi, M., et al. 2019, *A&ARv*, 27, 5

Casey-Clyde, J. A., Mingarelli, C. M. F., Greene, J. E., et al. 2022, *ApJ*, 924, 93

Charisi, M., Taylor, S. R., Runnoe, J., Bogdanovic, T., & Trump, J. R. 2022, *MNRAS*, 510, 5929

Chen, S., Caballero, R. N., Guo, Y. J., et al. 2021, *MNRAS*, 508, 4970

Chen, S., Sesana, A., & Conselice, C. J. 2019, *MNRAS*, 488, 401

Chen, Z., Faber, S. M., Koo, D. C., et al. 2020, *ApJ*, 897, 102

Conselice, C. J. 2014, *ARA&A*, 52, 291

de Nicola, S., Marconi, A., & Longo, G. 2019, *MNRAS*, 490, 600

Duncan, K., Conselice, C. J., Mundy, C., et al. 2019, *ApJ*, 876, 110

Ferré-Mateu, A., Mezcua, M., Trujillo, I., Balcells, M., & van den Bosch, R. C. E. 2015, *ApJ*, 808, 79

Goncharov, B., Shannon, R. M., Reardon, D. J., et al. 2021, *ApJL*, 917, L19

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, 585, 357

Huerta, E. A., McWilliams, S. T., Gair, J. R., & Taylor, S. R. 2015, *PhRvD*, 92, 063010

Hunter, J. D. 2007, *CSE*, 9, 90

Jaffe, A. H., & Backer, D. C. 2003, *ApJ*, 583, 616

Janet, F. A., Hobbs, G. B., van Straten, W., et al. 2006, *ApJ*, 653, 1571

Keenan, R. C., Foucaud, S., De Propris, R., et al. 2014, *ApJ*, 795, 157

Kelley, L. Z., Blecha, L., & Hernquist, L. 2017, *MNRAS*, 464, 3131

Kelley, L. Z., Blecha, L., Hernquist, L., Sesana, A., & Taylor, S. R. 2018, *MNRAS*, 477, 964

Kitzbichler, M. G., & White, S. D. M. 2008, *MNRAS*, 391, 1489

Kormendy, J., & Ho, L. C. 2013, *ARA&A*, 51, 511

Kormendy, J., & Richstone, D. 1995, *ARA&A*, 33, 581

Lotz, J. M., Jonsson, P., Cox, T. J., et al. 2011, *ApJ*, 742, 103

Marsden, C., Shankar, F., Ginolfi, M., & Zubovas, K. 2020, *FrP*, 8, 61

Matt, C., Gürtekin, K., & Simon, J. 2023, *MNRAS*, submitted

McConnell, N. J., & Ma, C.-P. 2013, *ApJ*, 764, 184

McWilliams, S. T., Ostriker, J. P., & Pretorius, F. 2014, *ApJ*, 789, 156

Middleton, H., Sesana, A., Chen, S., et al. 2021, *MNRAS*, 502, L99

Mingarelli, C. M. F., Lazio, T. J. W., Sesana, A., et al. 2017, *NatAs*, 1, 886

Moustakas, J., Coil, A. L., Aird, J., et al. 2013, *ApJ*, 767, 50

Perera, B. B. P., DeCesar, M. E., Demorest, P. B., et al. 2019, *MNRAS*, 490, 4666

Peters, P. C., & Mathews, J. 1963, *PhRv*, 131, 435

Pol, N. S., Taylor, S. R., Kelley, L. Z., et al. 2021, *ApJL*, 911, L34

Rajagopal, M., & Romani, R. W. 1995, *ApJ*, 446, 543

Ravi, V., Wyithe, J. S. B., Hobbs, G., et al. 2012, *ApJ*, 761, 84

Ravi, V., Wyithe, J. S. B., Shannon, R. M., & Hobbs, G. 2015, *MNRAS*, 447, 2772

Ravi, V., Wyithe, J. S. B., Shannon, R. M., Hobbs, G., & Manchester, R. N. 2014, *MNRAS*, 442, 56

Robotham, A. S. G., Driver, S. P., Davies, L. J. M., et al. 2014, *MNRAS*, 444, 3986

Rosado, P. A., Sesana, A., & Gair, J. 2015, *MNRAS*, 451, 2417

Sampson, L., Cornish, N. J., & McWilliams, S. T. 2015, *PhRvD*, 91, 084055

Sesana, A. 2013, *MNRAS*, 433, L1

Sesana, A., Shankar, F., Bernardi, M., & Sheth, R. K. 2016, *MNRAS*, 463, L6

Sesana, A., Vecchio, A., & Colacino, C. N. 2008, *MNRAS*, 390, 192

Sesana, A., Vecchio, A., & Volonteri, M. 2009, *MNRAS*, 394, 2255

Shen, S., Mo, H. J., White, S. D. M., et al. 2003, *MNRAS*, 343, 978

Shen, Y., Greene, J. E., Ho, L. C., et al. 2015, *ApJ*, 805, 96

Sheth, R. K., Bernardi, M., Schechter, P. L., et al. 2003, *ApJ*, 594, 225

Simon, J., & Burke-Spolaor, S. 2016, *ApJ*, 826, 11

Siwek, M. S., Kelley, L. Z., & Hernquist, L. 2020, *MNRAS*, 498, 537

Stemo, A., Comerford, J. M., Barrows, R. S., et al. 2020, *ApJ*, 888, 78

Suh, H., Civano, F., Trakhtenbrot, B., et al. 2020, *ApJ*, 889, 32

Taylor, E. N., Franx, M., Brinchmann, J., van der Wel, A., & van Dokkum, P. G. 2010, *ApJ*, 722, 1

Taylor, L., Bezanson, R., van der Wel, A., et al. 2022, *ApJ*, 939, 90

Thorne, K. S. 1987, in Gravitational Radiation, ed. S. W. Hawking & W. Israel (Cambridge: Cambridge Univ. Press), 330

Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, *ApJ*, 783, 85

van den Bosch, R. C. E. 2016, *ApJ*, 831, 134

van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, *ApJ*, 788, 28

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, 17, 261

Wake, D. A., van Dokkum, P. G., & Franx, M. 2012, *ApJL*, 751, L44

Walsh, J. L., van den Bosch, R. C. E., Gebhardt, K., et al. 2015, *ApJ*, 808, 183

Walsh, J. L., van den Bosch, R. C. E., Gebhardt, K., et al. 2016, *ApJ*, 817, 2

Wendt, D. K., & Romani, R. W. 2023, *ApJ*, 944, 112

Whitaker, K. E., Labb  , I., van Dokkum, P. G., et al. 2011, *ApJ*, 735, 86

Williams, R. J., Quadri, R. F., Franx, M., et al. 2010, *ApJ*, 713, 738

Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labb  , I. 2009, *ApJ*, 691, 1879

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

Xin, C., Mingarelli, C. M. F., & Hazboun, J. S. 2021, *ApJ*, 915, 97