Estimates of Detection Rates for LISA Capture Sources

LIST WG1 EMRI taskforce: L. Barak, T. Creighton, C. Cutler,
J. Gair, S. Larson, E.S. Phinney, K.S. Thorne, M. Vallisneri

(Dated: February 2, 2004)

Abstract

DRAFT. We estimate the detection rate for LISA capture sources: extreme-mass-ratio inspirals (EMRIs) of stellar-mass compact objects into massive black holes at the centers of galaxies. We show how the detection rate depends on different assumptions about the instrument (5×10^6 km arms vs. 1.7×10^6km, 3 arms vs. only 2 arms) and assumptions about how well the galactic WD-binary confusion noise can be subtracted out.

I. INTRODUCTION

The LISA detection rate for capture sources basically depends on three factors: i) the actual astrophysical capture rate (in $\mathrm{Mpc}^{-3}\mathrm{yr}^{-1}$), ii) LISA's sensitivity to captures (expressed, e.g., as LISA's matched filtering SNR for a fiducial source at some fiducial distance), and (iii) the SNR required for detection when computationally realistic algorithms are used to dig the capture signals out of the noise. This document summarizes what Working Group I has learned about these 3 factors and synthesizes them to obtain estimated detection rates.

We divide sources into 12 classes: three mass ranges for the MBH multiplied by four types of stellar-mass object– $0.6M_\odot$ WDs, $1.4M_\odot$ NSs and heavy WDs, $10M_\odot$ BHs, and $100M_\odot$ Pop III BHs. We estimate LISA detection rates for each class, for several different assumptions about the LISA instrumental and confusion noise levels.

Section II gives our estimate of capture rates for the different source classes. Sec. III describes our different models of the instrumental and confusion noise. Sec. IV displays LISA's sensitivity to various classes of sources of kludged waveforms (parametrized by MBH mass and captured-object mass), assuming optimal matched filtering. Sec V compares our kludged waveforms to honest full-GR waveforms in the circular orbit cases for which the latter are available. Since optimal matched filtering is computationally impossible, Sec. VI sketches a sub-optimal, *semi-coherent* search method and estimates the effective decrease in LISA's sensitivity (compared to optimal filtering) using this method. An important step in this estimate is counting the number of templates that must be searched over, which determines the search's computational cost. Sec. VII folds the different pieces together to estimate detection rates, and Sec. VIII summarizes the results and makes recommendations.

II. CAPTURE RATES

To compute the normalisation of the present-day EMRI capture rate, we need the distribution of supermassive black hole masses in galactic nuclei and an understanding of the current structure of their nuclei (computing the cosmic evolution of the capture rate additionally requires the even less well known growth and evolution of the black holes and their clusters, though we give a suggested scaling in equation 14 which neglects this evolution).

The tightest black hole mass indicator is the correlation with spheroid velocity dispersion

$$M_\bullet = M_{\bullet,*}(\sigma/\sigma_*)^\lambda \tag{1}$$

where σ is the galaxy velocity dispersion, $\sigma_* = 90\text{km s}^{-1}$ and $h_{65} = H_0/65\text{km s}^{-1}\text{Mpc}^{-1}$, and we adopt $\lambda = 5$ and $M_{\bullet,*} = 3\times10^6 h_{65}^{-1} M_\odot$. Merritt and Ferrarese (2002) find $\lambda = 4.72$ and $M_{\bullet,*} = 3\times10^6 M_\odot$. (Ferrarese and Merrit use dispersion in aperture $r_e/8$; often these are corrected from measured central values -i.e. the same prescription and σ's used in the $\phi(\sigma)$ distribution function below.) Tremaine et al(2002) find $\lambda = 4.02$ and $M_{\bullet,*} = 5\times10^6 M_\odot$, but use a nonstandard definition of dispersion which includes rotation and is averaged over an aperture of r_e, the half-light radius.

We considered two 'observational' sources of the galaxy velocity dispersion function: the Sloan Digital Sky Survey direct measurement (Sheth et al 2003), and indirect inferences (from Luminosity functions and $L-\sigma$ correlations) by Aller and Richstone (2002). The former would have been preferred, but it seems highly suspect in the range of interest, probably because the Sloan spectrometer has inadequate velocity resolution.

The distribution of velocity dispersions (space density per $\ln\sigma$) of early-type galaxies in SDSS is (Sheth et al 2003)

$$\sigma\phi(\sigma) = \frac{dN}{d\ln\sigma} = \phi_* \frac{\beta}{\Gamma(\alpha/\beta)} \left(\frac{\sigma}{\sigma_*}\right)^\alpha \exp[-(\sigma/\sigma_*)^\beta] \tag{2}$$

with (rounded) best fit values of $\phi_* = 0.0018 h_{65}^3 \text{Mpc}^{-3}$ (note that $\phi_* = \int_0^\infty \phi(\sigma)d\sigma$ is the total number density of early-type galaxies), $\alpha = 6.5$, $\beta = 2$, $\sigma_* = 90\text{km s}^{-1}$. $\Gamma(6.5/2) = 2.55$. Note that the Sloan dispersion measurements are of the light within a radius of 1.5 arcsec, and have been corrected to the dispersion at a standard fraction $r_e/8$ of the galaxies effective radius (determined from photometry) using an *average* dispersion profile: $\sigma = \sigma(1.5\text{arcsec})([r_e/8]/1.5\text{arcsec})^{0.04}$ (equation 1 of Bernardi et al), and because of the low spectrograph resolution, are unreliable for $\sigma < 70\text{km s}^{-1}$. Galaxies with lower dispersions (i.e. black hole masses below $9\times10^5 M_\odot$) are not included in the sample (Appendix B of Bernardi et al 2003), and the functional form is observationally unconstrained there.

Temporarily neglecting the dispersion in the $M_\bullet-\sigma$ relation, we translate $N(\sigma)$ to $N(M_\bullet)$, and find

$$M_\bullet dN(M_\bullet)/dM_\bullet = N(\sigma(M_\bullet))d\ln\sigma/d\ln M_\bullet \tag{3}$$

$$= \theta_*(M_\bullet/M_{\bullet,*})^\gamma \exp\left[-(M_\bullet/M_{\bullet,*})^\epsilon\right] \text{Mpc}^{-3}\,, \tag{4}$$

Galaxy Type	$M_{\bullet,*}$ $(10^7 h_{65}^{-1} M_\odot)$	θ_* $(10^{-3} h_{65}^3 \mathrm{Mpc}^{-3})$	γ	ϵ	$\rho_\bullet$ $10^5 h_{65}^2 M_\odot \mathrm{Mpc}^{-3}$
E+S0 (Sheth)	0.3	0.28	1.26	0.4	1.4 ($\lambda = 5$)
E+S0 (Sheth)	0.5	0.35	1.63	0.5	1.3 ($\lambda = 4$)
E (A& R)	17	0.3	0.12	0.8	0.6
S0 (A& R)	5	1.6	0.046	0.8	0.8
Sa-Sb (A& R)	2	1.8	0.32	0.8	0.4
Sc-Sd (A& R)	0.5	0.9	0.03	0.8	0.04

TABLE I: Parameters for black hole space densities in equation 5.

with $\theta_* = \phi_* \beta / (\lambda \Gamma[\alpha/\beta]) \ = 2.8 \times 10^{-4} h_{65}^3$, $\gamma = \alpha/\lambda = 1.26$ and $\epsilon = \beta/\lambda = 2/5$ for $\lambda = 5$ and $1/2$ for $\lambda = 4$.

Aller and Richstone (2002) convert luminosity functions of galaxies as a function of galaxy type to bulge velocity dispersion functions using (assuming no dispersion in the Faber-Jackson relation between luminosity and velocity dispersion) and typical ratios of bulge to disk luminosities as a function of galaxy type, and then use these to derive a black hole mass functions for the various galaxy types. They give this in the form identical to eq 4:

$$M_\bullet dN/dM_\bullet = \theta_* (M_\bullet / M_{\bullet,*})^\gamma \exp(-(M_\bullet / M_{\bullet,*})^\epsilon) \tag{5}$$

with $\epsilon = 3.08/\lambda$, so $\epsilon = 0.8$ for $\lambda = 4$ and $\epsilon = 0.6$ for $\lambda = 5$.

With this functional form, the contribution to the mass density of black holes is

$$\rho_\bullet = \theta_* M_{\bullet,*} \epsilon^{-1} \Gamma([\gamma + 1]/\epsilon) \tag{6}$$

given in the last column of the table for the specified values of λ and hence ϵ (the Aller and Richstone 2002 entries are all for $\lambda = 4$). The total Aller and Richstone 2002 density, for $(H_0 = 65)$ is $1.9 \times 10^5 M_\odot \mathrm{Mpc}^{-3}$ for $\lambda = 4$.

The total E+S0 density from the Sheth measured dispersion function agrees well with the one Aller and Richstone inferred from the luminosity functions. *However*, the actual space densities (see figure 1) are very different in the range relevant for EMRI: the Sheth et al dispersion function has 2×10^{-2} of the density of E+S0 galaxies containing $10^6 M_\odot$ black holes as does the Aller and Richstone one. In the range where the Sheth et al space density

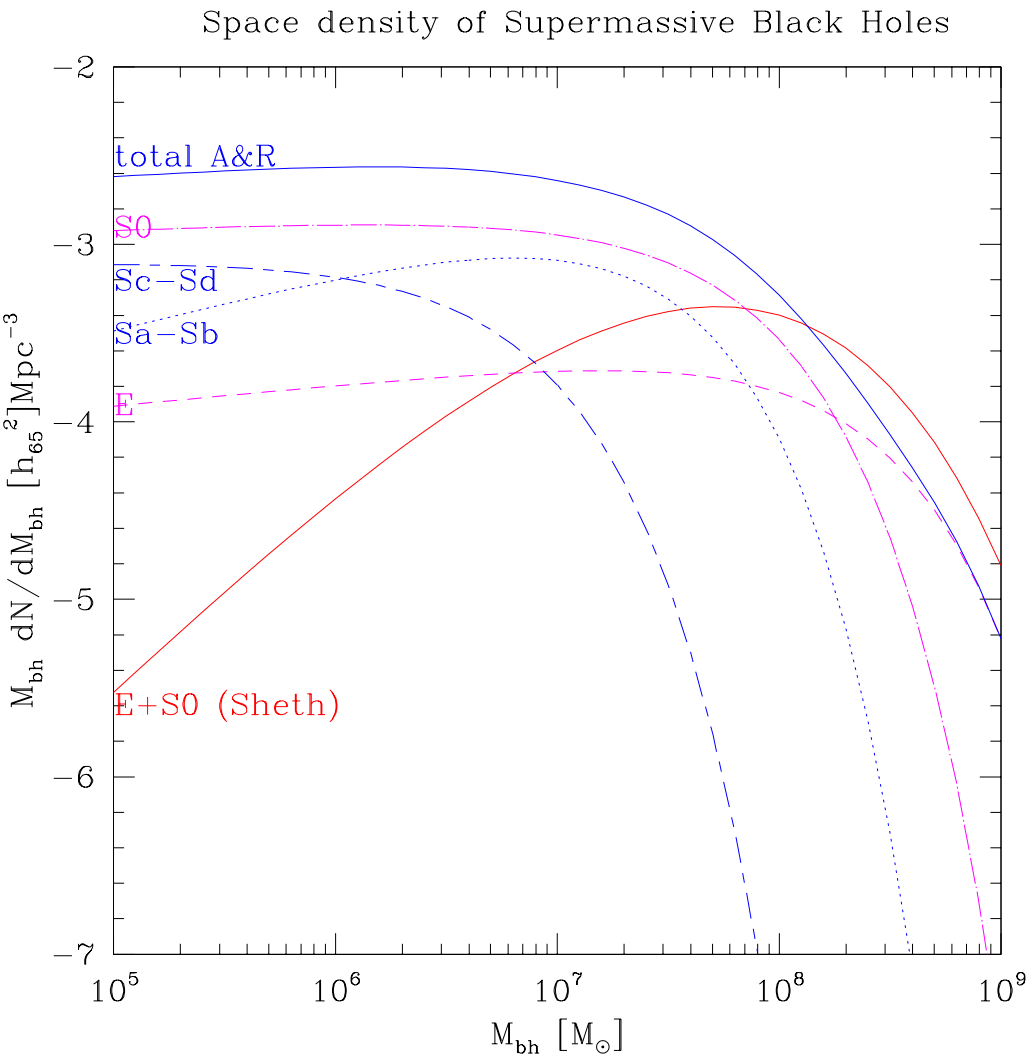

FIG. 1: Black hole space densities derived as described in text. The solid red line labeled E+S0 (Sheth) is based on the SDSS velocity dispersion function of Sheth et al 2003 and the $M_\bullet - \sigma$ relation of Merrit & Ferrarese 2001. The other curves are from the luminosity-dispersion relations for galaxies used in Aller and Richstone 2002, and the $M_\bullet - \sigma$ relation of Tremaine et al 2002.

is observationally constrained ($M_\bullet > 2 \times 10^6 M_\odot$) it lies well below the Aller and Richstone E+S0 curve for $M_\bullet < 7 \times 10^7 M_\odot$ and well above it for $M_\bullet > 7 \times 10^7 M_\odot$. The total E+S0 black hole mass density coincidentally comes out about the same. Because the Sheth et al dispersion function is poorly determined in the range of interest for EMRI, we prefer to use the Aller and Richstone curves, despite their indirect provenance.

The Aller and Richstone (2002) black hole space densities have similar shape, but are

about 1/3 those of Salucci et al (1999; note that Salucci et al use $H_0 = 75\mathrm{km\ s^{-1}Mpc^{-1}}$) in the $10^5 < M_\bullet < 10^7 M_\odot$ range of interest.

To good approximation, the total Aller and Richstone black hole space density in the range for $M_\bullet < 5 \times 10^6 M_\odot$ is simply

$$M_\bullet dN_{AR}/dM_\bullet = 3 \times 10^{-3} h_{65}^2 \mathrm{Mpc}^{-3} \tag{7}$$

This is contributed mainly by S0 galaxies. If Sc-Sd galaxies are removed (at least some of these –e.g. M33 (Gebhardt et al 2001) and NGC 4395 (Filippenko and Ho 2003) have central black holes with masses much lower than would be derived from their luminosities using Aller and Richstone's procedure), the space density would be lowered by a factor of about 2.

This lowered value we adopt as our reference black hole space density:

$$M_\bullet dN_{ESP}/dM_\bullet = 1.5 \times 10^{-3} h_{65}^2 \mathrm{Mpc}^{-3} \tag{8}$$

valid for $M_\bullet < 5 \times 10^6 M_\odot$. The total space densities in black holes in a range $\Delta \log_{10} M_\bullet = 0.5$ are just 1.15 times $M_\bullet dN_{ESP}/dM_\bullet$, i.e. $1.7 \times 10^{-3} h_{65}^2 \mathrm{Mpc}^{-3}$.

Freitag's Milky Way simulation (which coincidentally has just about $\sigma = \sigma_*$ and an $M_\bullet = 4 \times 10^6 M_\odot$, has present-day capture rates for $9M_\odot$ black holes of $10^{-6}\mathrm{y}^{-1}$, for $1.4M_\odot$ neutron stars of $10^{-6}\mathrm{y}^{-1}$, and for $0.6M_\odot$ white dwarfs of $5 \times 10^{-6}\mathrm{y}^{-1}$. For the $\sim 100M_\odot$ black hole remnants of Population III stars, we take the dynamical friction rate of 2 per Gyr in the Milky Way from Madau and Rees (2001), and optimistically assume that half of these are captured by gravitational radiation (the rest being direct captures).

To scale these to other galaxies, we make use of the observation (Gebhardt et al 1996, Rest et al 2001) that bright galaxies (with $M_\bullet > 10^8 M_\odot$) have shallow stellar cusps $\rho_* \propto r^{-1}$, while fainter galaxies (with $M_\bullet < 10^8 M_\odot$), i.e. those relevant to LISA detection, have singular isothermal cusps

$$\rho_* = \sigma_*^2/(2\pi G r^2) \propto r^{-2}\ . \tag{9}$$

We scale the central black hole mass as $\tilde{M} \equiv M/(5 \times 10^6 M_\odot) \simeq M/M_{\bullet,*}$. The black hole dominates the dynamics inside

$$r < r_{gr} \equiv GM_\bullet/\sigma_*^2 = 2.5\tilde{M}^{1/2}\mathrm{pc} \tag{10}$$

The relaxation time for stars of mass m is

$$t_{rel}(r) \simeq 0.3 \frac{\sigma_*^3}{G^2 m \rho_*(r) \ln \Lambda} \ . \tag{11}$$

This is also the dynamical friction time for bodies of mass m much larger than the mean. For $r > r_{gr}$, substituting the singular isothermal $\rho_*(r)$ we get

$$t_{rel} \ \simeq \ 2 \frac{\sigma_* r^2}{G m \ln \Lambda} \tag{12}$$

$$\simeq \ 2 \times 10^{10} \mathrm{y} \tilde{M}^{1.25} (r/r_{gr})^2 \tag{13}$$

for $r > r_{gr}$.

Thus only $M_\bullet < M_{\bullet,*}$ black holes have fully relaxed cusps ($r = r_{gr}$).

The mass in relaxed stars is $M_{rel} = M_\bullet \tilde{M}^{-5/8} = M_{\bullet,*} \tilde{M}^{3/8}$, the mass from which stars of mass m larger than the mean have sunk to the center by dynamical friction is $M_{df} = M_{rel}(m/\langle m \rangle)^{1/2}$. If the mass function was independent of radius, the rate of supply of heavy objects to the center by dynamical friction scales just as

$$\dot{M} \sim M_{df}(t)/t \propto r_{df}/t \propto t^{-1/2} \tag{14}$$

This could be an appropriate redshift scaling for black hole EMRI out to modest redshifts (~ 1) at which the central black hole masses and stellar cusp composition were not very different from the present.

Since these heavy bodies are captured mainly by large-angle scattering (see Sigurdsson & Rees 1997) the present rate of gravitational capture is comparable to the direct plunge rate (since they come from a comparable range of initial pericenters). Thus the capture rate is approximately half of the dynamical friction rate. Furthermore the capture time is short compared to a Hubble time, so the EMRI rate is limited by the dynamical friction rate. If we denote by f the fraction of the total stellar mass in the remnants of interest (with a Kroupa IMF and standard initial-final mass relations, $f = 0.2$ for $0.7 M_\odot$ white dwarfs and $f = 0.03$ for $10 M_\odot$ black holes, while $f = 4 \times 10^{-5}$ for $\sim 100 M_\odot$ Pop III black holes (cf. Madau & Rees 2001), we predict an EMRI rate in the galaxy today of

$$\frac{1}{2} \frac{f M_{df}}{m t} \simeq (f M_{\bullet,*}/2mt) \tilde{M}^{3/8} (m/\langle m \rangle)^{1/2} \simeq 10^{-4} f \tilde{M}^{3/8} (m/M_\odot)^{-1/2} \mathrm{y}^{-1} \ , \tag{15}$$

where we used $\langle m \rangle = 1 M_\odot$. For $M = M_{\bullet,*}$, the EMRI rates predicted by this simple equation agree with the rate from Freitag's simulation for $10 M_\odot$ black holes, and within a factor of

$M_\bullet$	space density	Merger rate $\mathcal{R}$			
$M_\odot$	$10^{-3}h_{65}^2\mathrm{Mpc}^{-3}$	$\mathrm{Gpc}^{-3}\mathrm{y}^{-1}$			
		$0.6M_\odot$ WD	$1.4M_\odot$ MWD/NS	$10M_\odot$ BH	$100M_\odot$ PopIII
$10^{6.5\pm0.25}$	1.7	8.5	1.7	1.7	1.7×10^{-3}
$10^{6.0\pm0.25}$	1.7	6	1.1	1.1	10^{-3}
$10^{5.5\pm0.25}$	1.7	3.5	0.7	0.7	7×10^{-4}

TABLE II: Rates of EMRI merger for three ranges of supermassive black hole mass and four types of compact objects, based on Aller and Richstone 2002 E+S0+Sa/Sb black hole space densities, Freitag 2001 merger rates for $10^{6.5}M_\odot$ black hole, and rate scaling with black hole mass from equation 15.

two with the Madau-Rees estimate for $100M_\odot$ black holes. Not surprisingly, for white dwarfs eq 15 overpredicts the rate by a factor of 4 compared to Freitag's simulations (the capture time for WD is not less than the Hubble time, and mass segregation discriminates against the lower mass white dwarfs), so a much larger fraction of captures are direct, not gravitational bremsstrahlung). However, the $\tilde{M}^{3/8}$ scaling of the present rate with black hole mass should be fairly accurate for all stellar components, so we have used this to scale Freitag's rates for the Milky Way (used to generate the top line of table II) to the other table entries.

What are the uncertainties of the rates in Table II? The space densities for the dominant E+S0 galaxies implied by the SDSS dispersion distribution are more than an order of magnitude lower (see fig 1). But the fit is not to be trusted in this mass range because of the poor resolution of SDSS spectra. Hils and Bender 1995 [14] estimated a white dwarf EMRI rate in M32 of $2\times10^{-8}\mathrm{y}^{-1}$, compared to our extrapolation from Freitag of $3\times10^{-6}\mathrm{y}^{-1}$. However Hils and Bender assumed a cusp grown adiabatically $\rho\propto r^{-1.5}$, rather than a relaxed cusp ($\rho\propto r^{-1.75}-r^{-2}$), and only half as many white dwarfs as modern IMFs predict. Sigurdsson and Rees (1997) assumed relaxed cusps, and estimated a white dwarf EMRI rate in the Milky Way of $10^{-7}\mathrm{y}^{-1}$, a factor of 50 below Freitag's rate. However their cusps were not fully self-consistent, and did not include mass-segregation in a multi-mass system including black holes. Nevertheless, Freitag's simulations have not been independently verified, and they have a number of oversimplifications and limitations (e.g. inability to treat large-angle scattering; scattering turned off if gravitational capture time less than scattering time).

Conservative rate estimates for white dwarfs would therefore be 10^{-2} of those in Table II. The rate estimates for black holes are more robust to stellar dynamics, but depend on the mass fraction of stellar mass black holes —i.e. the IMF, initial-final mass relation and possible natal kick velocities. The increasing number of galactic black hole binaries give some confidence that these are not wildly different from our assumptions. But allowing an order of magnitude reduction in rate might be prudent.

III. THE LISA NOISE CURVE

We consider in this document the standard Pre-phase A LISA with 5×10^6km arms. We also consider a "short LISA" with arms of 1.6×10^6km. The white dwarf confusion noise we estimate in an 'optimistic' way: assuming that every white dwarf not sharing a 1/5yr frequency bin with another can be removed perfectly, and a 'pessimistic' way: using Cornish and Larson's gCLEAN algorithm on simulated frequency ranges with 3-years of LISA data, and measuring the residual noise after souces are fitted.

A. The Instrumental Noise

Our treatment of the instrumental noise is based on the noise budget of the LISA Pre-Phase A Report. When counting templates (as opposed to calculating SNRs), we use simple analytic fits (sums of power laws) to the actual noise.

The plots shown in Figures 2 and 3 do not look like the usual LISA sensitivity plots. This is because they are are for the raw instrumental noise spectrum processed through the A (or E or T) TDI variable. The raw instrumental noise is taken to be the combination of proof-mass (sometimes called "acceleration") noise N_{pm}, and the aggregate optical path noise N_{op} (broken up in the LPPA report as a "postion noise budget", for which laser shot noise is the largest single contributor).

The spectral densities of these individual noises are given by

$$N_{pm} = \left(\frac{\sqrt{S_{pm}} L}{c^2} \right)^2 , \tag{16}$$

and

$$N_{op} = \left(\frac{\sqrt{S_{op}}}{L} \right)^2 \ , \tag{17}$$

where L is the characteristic armlength of the interferometer ($L = 5 \times 10^9$ m for the standard LISA), c is the speed of light, and

$$\sqrt{S_{pm}} = 3.0 \times 10^{-15} \mathrm{ms}^{-2} \mathrm{Hz}^{-1/2} \ , \tag{18}$$

and

$$\sqrt{S_{op}} = 2.0 \times 10^{-11} \mathrm{mHz}^{-1/2} \ , \tag{19}$$

are the noise specifications for the observatory (the values given here are the baseline LPPA values).

The noise in the TDI A channel[32] is given by

$$S_A = N_{pm} \mathcal{R}_{pm} + N_{op} \mathcal{R}_{op} \ , \tag{20}$$

where $\mathcal{R}_{pm}$ and $\mathcal{R}_{op}$ are respectively the noise transfer functions for proof-mass noise and optical-path noise in the A channel. They are given by

$$\mathcal{R}_{pm} = 2\mathrm{sinc}^2 \left[\frac{u}{2} \right] \cdot 2(3 + 2\cos(u) + \cos(2u)) \ , \tag{21}$$

and

$$\mathcal{R}_{op} = 2\mathrm{sinc}^2 \left[\frac{u}{2} \right] \cdot u^4 (2 + \cos(u)) \ , \tag{22}$$

parameterized by the dimensionless parameter $u = 2\pi f L/c$.

Figures 2 and 3 show plots of Eq. (20) composited with estimates of the astrophysical noise background due to galactic binaries, which we now describe.

B. The WD-binary Confusion Noise

We use two alternative versions of the WD confusion noise curve. The first, which follows Barack & Cutler (2003), is a rather optimistic version that assumes the WD binaries can basically be subtracted out to the level suggested by Shannon's Theorem. The second, more conservative, version uses the same assumptions about background populations as the first, but assumes the WD noise is reduced only to the level currently attainable with the gCLEAN algorithm developed by Cornish & Larson (2003).

The relevant section of Barack & Cutler (2003) is reprinted here for convenience.

Any isotropic background of indistinguishable GW sources represents (for the purpose of analyzing *other* sources) a noise source with spectral density [1]

$$S_h^{\rm conf}(f) = \frac{3}{5\pi} f^{-3} \rho_c \Omega_{GW}(f)\,, \tag{23}$$

where $\rho_c \equiv 3H_0^2/(8\pi)$ is the critical energy density needed to close the universe (assuming it is matter-dominated) and $\Omega_{GW} \equiv (\rho_c)^{-1} d\rho_{GW}/d(\ln f)$ is the energy density in gravitational waves (expressed as a fraction of the closure density) per logarithmic frequency interval[33]. For the extragalactic WD background, Farmer and Phinney [7] estimate that, for f near 1 mHz, $\Omega_{GW}(f) = 3.6 \times 10^{-12} (f/10^{-3}{\rm Hz})^{2/3}$ [at $H_0 = 70{\rm km}/({\rm sec}\cdot{\rm Mpc})$], so

$$S_h^{\text{ex. gal}} = 4.2 \times 10^{-47} \left(\frac{f}{1{\rm Hz}}\right)^{-7/3} {\rm s}. \tag{24}$$

Note Eq. (24) is not a good fit to $S_h^{\text{ex. gal}}$ for $f \gtrsim 10^{-2}$Hz, where mergers cause the spectrum to decrease more sharply. However, at such high frequencies, instrumental noise dominates the total noise in any case, so for our purposes the extrapolation of Eq. (24) to high frequencies is harmless.

A recent calculation of the galactic confusion background by Nelemans *et al.* [21] yields an Ω_{GW}^{gal} that is 5.0×10^1 times larger than $\Omega_{GW}^{ex.\ gal}$ (near 1 mHz) [7]; therefore [34]

$$S_h^{\rm gal}(f) = 2.1 \times 10^{-45} \left(\frac{f}{1{\rm Hz}}\right)^{-7/3} {\rm s}\,. \tag{25}$$

This is larger than instrumental noise in the range $\sim 10^{-4}$–10^{-2} Hz. However, at frequencies $f \gtrsim 3\times 10^{-3}\,$Hz, galactic sources are sufficiently sparse, in frequency space, that one expects to be able to "fit them out" of the data. An estimate of the resulting noise is [17]

$$S_h^{\rm inst+gal}(f) = \min\left\{ S_h^{\rm inst}(f)/\exp(-\kappa T^{-1} dN/df),\ \ S_h^{\rm inst}(f) + S_h^{\rm gal}(f) \right\}\,. \tag{26}$$

Here dN/df is the number density of galactic white dwarf binaries per unit GW frequency, T is the LISA mission lifetime (so $\Delta f = 1/T$ is the bin size of the discretely Fourier transformed data), and κ is the average number of frequency bins that are "lost" (for the purpose of analyzing other sources) when each galactic binary is fitted out (κ is larger than one because LISA's motion effectively smears the signal from each binary over several frequency bins). The factor $\exp(-\kappa T^{-1} dN/df)$ is therefore the fraction of "uncorrupted"

bins, where instrumental noise still dominates. For dN/df we adopt the estimate [17]

$$\frac{dN}{df} = 2 \times 10^{-3}\,\mathrm{Hz}^{-1} \left(\frac{1\,\mathrm{Hz}}{f}\right)^{11/3} \tag{27}$$

and take $\kappa T^{-1} = 1.5/\mathrm{yr}$ (corresponding to $T \approx 3\,\mathrm{yr}$ and $\kappa \approx 4.5$ [4]). To obtain the *total* LISA noise, we just add to Eq. (26) the contribution from the extragalatic confusion background, Eq. (24):

$$S_h(f) = S_h^{\mathrm{inst+gal}}(f) + S_h^{\mathrm{ex.\ gal}}(f)\,. \tag{28}$$

IV. MONTE CARLO SIMULATION OF LISA DETECTIONS

The Monte Carlo evaluation of S/Ns for the detection of gravitational captures by LISA was performed as follows:

1. A set of GW signals (h_+ and $h_\times$ strains at the Solar-system baricenter) was computed for the last five years before plunge of 18 EMRI systems, with parameters

$$(M/M_\odot, m/M_\odot, \mathrm{final}\, e) = \begin{pmatrix} 3\cdot 10^5 \\ 10^6 \\ 3\cdot 10^6 \end{pmatrix} \times \begin{pmatrix} 0.6 \\ 10 \\ 100 \end{pmatrix} \times \begin{pmatrix} 0.25 \\ 0.40 \end{pmatrix}. \tag{29}$$

   The signals were computed using the Hughes–Glampedakis–Kennefick–Gair–Creighton kludge scheme. For each system, the source was assigned to ecliptic latitude and longitude 0, and three different values (0, $\pi/2$, π) of source inclination to the line of sight were considered. The other parameters of the system were fixed as follows: $a/M = 0.8$; initial orbital inclination $= \pi/4$; polarization rotation angle $\psi = 0$; initial semilatus rectum P compatible with the final eccentricity requested; initial eccentricity e compatible with the final eccentricity requested; initial phases $\psi_0 = \pi/3$, $\chi_0 = \pi/5$, $\phi_0 = \pi/7$.

2. The signals were then filtered through *Synthetic LISA* (a simulator of the LISA TDI process currently under development at JPL by M. Vallisneri and J. Armstrong) to produce TDI series for the (*first-generation*) single-interferometer unequal-arm-Michelson combination X, and for the optimal interferometric combinations A, E, and T [see Prince et al., *Phys. Rev. D* **66**, 122002 (2002)]. These series are expressed in terms of

fractional laser frequency fluctuations (the standard TDI Doppler observables). The filtering was repeated for the baseline LISA configuration, and for short LISA.

LISA was modeled as a rigid equilateral triangle of three spacecraft on circular, inclined orbits. These orbits represents a very good approximation, as far as the GW signals are concerned, to the actual eccentric orbits. The LISA array was taken to be initially located along the x axis of a Solar-system-baricentric, ecliptic system, with spacecraft 1 sitting along the negative y axis.

3. *Synthetic LISA* was also used to generate synthetic time series of the LISA secondary noises (proof mass and optical path) for the TDI observables of interest. We computed a periodogram of the time series, and averaged it to reduce the variance of the resulting spectra. For short LISA, the noise parameters were adjusted by reducing photon shot noise appropriately and by changing the time delays in the TDI combinations. Galactic WD confusion noise was added consistently with Cutler and Barack's optimistic scenario (for 5 yr subtraction) and with a pessimist scenario based on fitting the residual noise after gCLEANing the data (for 3 yr subtraction). Specifically, the confusion noise was added, at each frequency, by dividing Cutler and Barack's LISA sensitivity curve *with* confusion noise by the same curve *without* confusion noise, and by multiplying the *Synthetic LISA* spectrum by the resulting ratio. For short LISA, a fit to Larson's online sensitivity generator was used instead of Cutler and Barack's curve.

4. The S/Ns of the signals were computed in the Fourier domain, using the noise spectra obtained at step 3, and are shown in Tables III to VI. The columns correspond to the S/N integrated for the last 5 yr, 3 yr, 1 yr, 3 mo, 1 mo, and 1 wk before plunge. The S/Ns shown are cube-root-cube averaged (see captions), to yield numbers that can be used directly to estimate detection rates (see below).

5. An estimate of the effect of averaging over source positions in the sky was obtained by computing S/N for 3-month stretches of data, for various ecliptic latitudes and longitudes, and by cube-root-cube averaging the resulting S/N. While the results are not definite enough to quote a correction factor to the numbers of Tables III to VI, the correction should be within $\pm 10\%$.

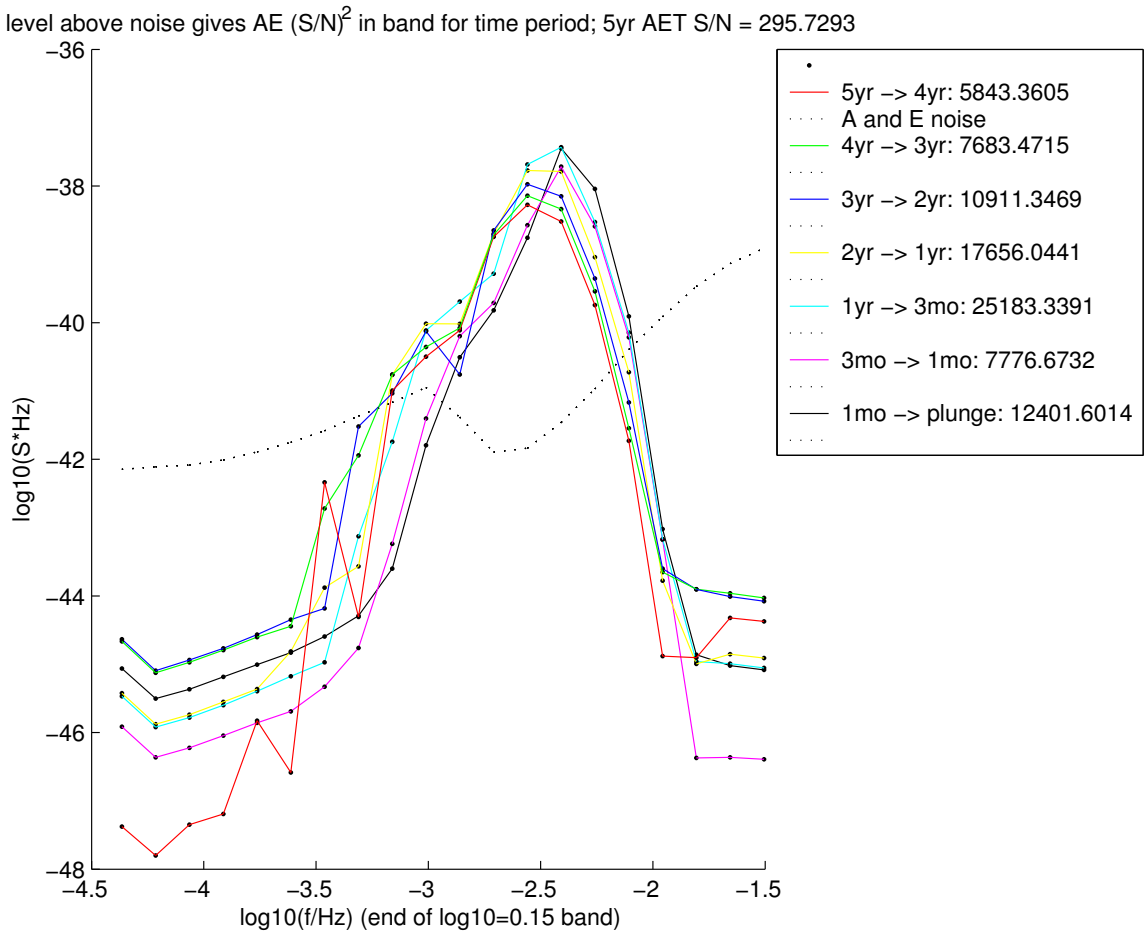

FIG. 2: The height above the TDI Doppler instrument+white dwarf noise (dotted) gives the contribution to the numerically-kludged (full Kerr geodesics, but quadrupole power only -see figures 4 and 5) signal to noise in the AE TDI combination as a function of frequency for the indicated time intervals of the inspiral of a $10M_\odot$ black hole into a $3 \times 10^6 M_\odot$ black hole, with eccentricity at plunge of 0.4. Source is in the ecliptic plane, viewed along the black hole spin axis, and orbit is inclined at 45 degrees prograde to $a/M = 0.8$ black hole.

6. A breakdown of the S/Ns among a set of frequency bins in the LISA band are shown in Figs.

   As explained in detail in section III, the dotted curves shows the spectra of fractional-frequency-fluctuation noise in the TDI observable (the bump in the middle corresponds to WD confusion noise, while the flat low-frequency behavior is explained by the TDI transfer function $\propto f^2$ for proof-mass noise, which by itself is $\propto f^{-2}$). The height of

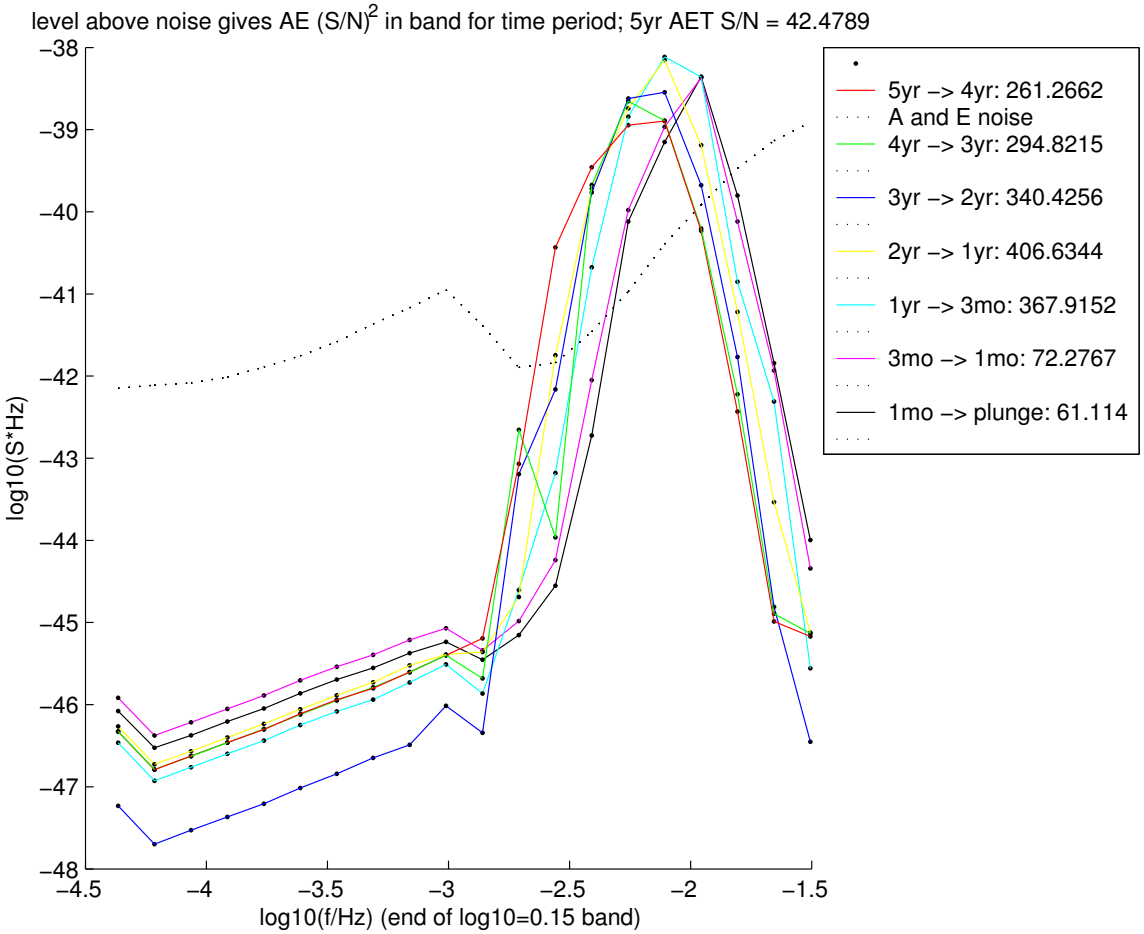

FIG. 3: As in figure 2, but for a $0.6M_\odot$ white dwarf inspiraling into a $1 \times 10^6 M_\odot$ black hole, with eccentricity at plunge of 0.25.

the dots over the noise curve shows the integrated $(\mathrm{S/N})^2$ for a given frequency bin of logarithmic width 0.15, *ending* at the frequencies on the abscissa. Curves of different colors correspond to different, nonoverlapping epochs in the 5 yrs before plunge, as explained in the legend (which shows also cumulative $(\mathrm{S/N})^2$ for the epochs). The $(\mathrm{S/N})^2$ curves are suppressed at low frequencies because of the inherent signal power there, and because of the LISA transfer function.

M	m	e (final)	S/N(AET)						S/N(X)					
			(1wk)	(1mo)	(3mo)	(1yr)	(3yr)	(5yr)	(1wk)	(1mo)	(3mo)	(1yr)	(3yr)	(5yr)
$3 \cdot 10^5$	0.6	0.25	1.1	3.0	5.1	10.2	16.8	20.4	0.6	1.6	2.2	5.8	10.2	12.6
$3 \cdot 10^5$	10	0.25	27.8	60.3	80.4	119.0	149.0	162.0	16.6	38.0	48.8	74.7	95.4	104.0
$3 \cdot 10^5$	100	0.25	277.0	440.0	508.0	591.0	626.0	633.0	188.0	300.0	338.0	391.0	414.0	419.0
10^6	0.6	0.25	3.7	7.3	10.0	18.5	29.0	34.9	2.5	4.9	6.3	12.0	19.0	23.0
10^6	10	0.25	58.2	109.0	140.0	205.0	252.0	271.0	40.5	75.5	92.9	136.0	168.0	181.0
10^6	100	0.25	477.0	752.0	860.0	989.0	1060.0	1090.0	338.0	532.0	595.0	678.0	727.0	743.0
$3 \cdot 10^6$	0.6	0.25	3.1	6.0	8.0	14.1	21.2	24.9	2.2	4.2	5.4	9.5	14.3	16.7
$3 \cdot 10^6$	10	0.25	45.7	81.8	102.0	138.0	158.0	164.0	32.7	57.8	69.8	93.9	107.0	111.0
$3 \cdot 10^6$	100	0.25	344.0	508.0	559.0	590.0	601.0	604.0	244.0	360.0	391.0	411.0	418.0	420.0
$3 \cdot 10^5$	0.6	0.4	1.3	3.0	4.5	8.2	11.6	13.1	0.9	2.0	2.6	4.8	6.8	7.7
$3 \cdot 10^5$	10	0.4	24.9	46.7	56.7	69.2	75.3	76.6	16.1	29.1	33.8	41.3	45.1	45.9
10^6	0.6	0.4	3.4	6.7	9.2	16.6	25.3	30.0	2.3	4.5	5.8	10.7	16.4	19.5
10^6	10	0.4	52.8	96.9	122.0	177.0	223.0	241.0	36.4	66.3	80.3	116.0	147.0	159.0
10^6	100	0.4	405.0	639.0	743.0	871.0	926.0	938.0	284.0	445.0	504.0	586.0	623.0	631.0
$3 \cdot 10^6$	0.6	0.4	3.2	6.1	8.1	14.5	22.4	27.0	2.3	4.2	5.4	9.7	15.1	18.1
$3 \cdot 10^6$	10	0.4	46.3	84.5	108.0	162.0	208.0	226.0	32.7	59.6	73.2	109.0	140.0	152.0
$3 \cdot 10^6$	100	0.4	370.0	596.0	696.0	826.0	898.0	921.0	264.0	422.0	481.0	566.0	614.0	629.0

TABLE III: Baseline LISA, optimistic WD subtraction (5yr), volume-inc-averaged: $\overline{\mathrm{S/N}} = (1/2 \int (\mathrm{S/N})^3 \, d\cos\iota)^{1/3}$.All values are for $\beta = 0$, $\lambda = 0$. Effect of position-in-the-sky averaging is $\pm 10\%$.

V. ACCURACY OF THE NUMERICAL WAVEFORMS

The numerical kludge waveforms used to estimate the signal to noise ratios in the preceding section are only approximate. They are computed by calculating the quadrupole radiation from a sequence of geodesics of the Kerr geometry. The parameters of the geodesic are evolved using post-newtonian radiation reaction expressions to generate an inspiral [12]. The actual radiation from an extreme mass ratio inspiral can be computed using the Teukolsky formalism. At present, Teukolsky codes exist only for eccentric equatorial [13] or circular

M	m	e (final)	S/N(AET)						S/N(X)					
			(1wk)	(1mo)	(3mo)	(1yr)	(3yr)	(5yr)	(1wk)	(1mo)	(3mo)	(1yr)	(3yr)	(5yr)
$3 \cdot 10^5$	0.6	0.25	1.1	3.0	5.1	10.2	16.5	19.8	0.6	1.6	2.2	5.7	10.0	12.1
$3 \cdot 10^5$	10	0.25	27.7	59.8	78.9	114.0	141.0	151.0	16.6	37.7	47.9	71.1	89.3	96.2
$3 \cdot 10^5$	100	0.25	269.0	418.0	479.0	548.0	575.0	581.0	182.0	284.0	318.0	361.0	379.0	383.0
10^6	0.6	0.25	3.6	7.1	9.7	17.7	26.9	31.5	2.5	4.8	6.2	11.5	17.6	20.7
10^6	10	0.25	56.3	104.0	130.0	174.0	199.0	207.0	39.1	71.4	86.3	115.0	132.0	138.0
10^6	100	0.25	431.0	631.0	693.0	744.0	769.0	777.0	304.0	445.0	481.0	514.0	530.0	535.0
$3 \cdot 10^6$	0.6	0.25	2.1	4.0	5.2	8.4	11.3	12.4	1.5	2.8	3.5	5.6	7.6	8.3
$3 \cdot 10^6$	10	0.25	28.3	46.6	54.4	62.2	66.2	67.9	19.9	32.5	37.4	42.4	45.0	46.1
$3 \cdot 10^6$	100	0.25	169.0	212.0	222.0	235.0	244.0	249.0	118.0	147.0	154.0	162.0	168.0	171.0
$3 \cdot 10^5$	0.6	0.4	1.3	3.0	4.4	8.0	11.1	12.5	0.9	2.0	2.6	4.6	6.4	7.3
$3 \cdot 10^5$	10	0.4	24.5	45.1	54.4	66.6	72.6	73.9	15.8	27.8	32.0	39.3	43.0	43.9
10^6	0.6	0.4	3.3	6.4	8.8	15.6	23.4	27.6	2.2	4.3	5.5	10.0	15.1	17.9
10^6	10	0.4	49.9	90.3	113.0	162.0	200.0	214.0	34.3	61.4	74.0	106.0	131.0	141.0
10^6	100	0.4	371.0	583.0	671.0	767.0	806.0	815.0	258.0	404.0	455.0	515.0	541.0	547.0
$3 \cdot 10^6$	0.6	0.4	2.4	4.6	6.0	10.2	14.9	17.3	1.7	3.2	4.0	6.8	10.0	11.6
$3 \cdot 10^6$	10	0.4	33.2	58.2	71.7	95.0	110.0	115.0	23.2	40.6	48.7	64.1	73.9	77.4
$3 \cdot 10^6$	100	0.4	236.0	344.0	379.0	414.0	433.0	439.0	166.0	240.0	262.0	285.0	297.0	301.0

TABLE IV: Baseline LISA, pessimistic WD subtraction (gCLEAN, 3yr), volume-inc-averaged: $\overline{\mathrm{S/N}} = (1/2 \int (\mathrm{S/N})^3 \, d\cos\iota)^{1/3}$. All values are for $\beta = 0$, $\lambda = 0$. Effect of position-in-the-sky averaging is $\pm 10\%$.

inclined orbits [16], and were too computationally intensive to be used for scoping out the data analysis. It is useful to know how well the kludged inspirals approximate true inspirals when interpreting our results.

Using the circular inclined Teukolsky code of Hughes [16], we computed the instantaneous amplitude and inspiral rate from a sequence of geodesics along a kludged inspiral trajectory. These were compared to the kludge waveforms and inspiral rates. The comparison was performed for a $0.6 M_\odot$ black hole falling into a $1 \times 10^6 M_\odot$ black hole with $S/M^2 = 0.8$. Table VII compares the instantaneous inspiral rate $d(r/M)/dt$ of the kludge and Teukolsky

M	m	e (final)	S/N(AET)						S/N(X)					
			(1wk)	(1mo)	(3mo)	(1yr)	(3yr)	(5yr)	(1wk)	(1mo)	(3mo)	(1yr)	(3yr)	(5yr)
$3 \cdot 10^5$	0.6	0.25	2.6	4.7	6.5	13.2	20.3	23.8	1.6	2.9	3.9	7.9	12.2	14.4
$3 \cdot 10^5$	10	0.25	51.5	81.7	99.9	135.0	160.0	170.0	31.8	52.0	61.8	82.6	98.1	104.0
$3 \cdot 10^5$	100	0.25	401.0	526.0	577.0	637.0	661.0	666.0	253.0	335.0	364.0	399.0	413.0	416.0
10^6	0.6	0.25	5.2	8.4	11.0	19.2	27.7	31.7	3.2	5.4	6.8	11.7	16.9	19.4
10^6	10	0.25	76.8	115.0	137.0	173.0	194.0	202.0	47.1	73.3	85.4	107.0	120.0	125.0
10^6	100	0.25	520.0	660.0	707.0	756.0	783.0	791.0	327.0	419.0	446.0	475.0	491.0	496.0
$3 \cdot 10^6$	0.6	0.25	2.3	3.7	4.7	7.8	11.1	12.8	1.4	2.3	2.9	4.8	6.8	7.8
$3 \cdot 10^6$	10	0.25	31.5	46.4	54.9	69.2	77.0	79.2	19.1	29.3	34.2	42.8	47.5	48.9
$3 \cdot 10^6$	100	0.25	212.0	268.0	285.0	299.0	307.0	310.0	130.0	167.0	178.0	186.0	191.0	193.0
$3 \cdot 10^5$	0.6	0.4	2.6	4.4	5.8	10.1	14.0	15.8	1.6	2.8	3.5	6.0	8.4	9.4
$3 \cdot 10^5$	10	0.4	42.0	60.7	70.2	85.4	92.5	94.0	26.3	38.3	43.4	52.1	56.4	57.3
10^6	0.6	0.4	4.8	7.7	10.0	17.5	25.8	30.3	2.9	4.9	6.2	10.7	15.7	18.4
10^6	10	0.4	70.0	105.0	127.0	173.0	207.0	220.0	43.1	67.0	79.0	107.0	127.0	135.0
10^6	100	0.4	496.0	663.0	732.0	813.0	847.0	854.0	309.0	421.0	461.0	508.0	528.0	532.0
$3 \cdot 10^6$	0.6	0.4	2.6	4.1	5.2	9.0	13.3	15.7	1.5	2.6	3.2	5.5	8.2	9.6
$3 \cdot 10^6$	10	0.4	35.5	53.7	65.1	90.6	110.0	118.0	21.4	33.9	40.5	55.8	67.9	72.6
$3 \cdot 10^6$	100	0.4	260.0	351.0	390.0	442.0	471.0	480.0	159.0	220.0	244.0	274.0	292.0	298.0

TABLE V: Short LISA, optimistic WD subtraction (5yr), volume-inc-averaged: $\overline{\mathrm{S/N}} = (1/2 \int (\mathrm{S/N})^3 \, d\cos\iota)^{1/3}$. All values are for $\beta = 0$, $\lambda = 0$. Effect of position-in-the-sky averaging is $\pm 10\%$.

waveforms for one prograde and one retrograde orbit, at a number of points along the inspiral. The points are labeled by the time from plunge, as measured using the kludged inspiral. The Teukolsky results are for the orbit with the same radius and inclination, not an orbit the same time from plunge. The inclination of the orbit ι is defined in terms of the Carter constant [12]. Prograde orbits have $0 \leq \iota < \pi/2$, and retorgrade orbits have $\pi/2 < \iota \leq \pi$. The rate of change of ι is not shown in the table. In the kludge this is set to zero. For the prograde orbit, the Teukolsky equation gives $\dot{i} = 7.7 \times 10^{-11}$ one week from plunge, and $\dot{i} = 2.1 \times 10^{-11}$ three years from plunge. The corresponding rates for the

M	m	e (final)	S/N(AET)						S/N(X)					
			(1wk)	(1mo)	(3mo)	(1yr)	(3yr)	(5yr)	(1wk)	(1mo)	(3mo)	(1yr)	(3yr)	(5yr)
$3 \cdot 10^5$	0.6	0.25	2.6	4.7	6.5	13.2	20.1	23.4	1.6	2.9	3.9	7.9	12.1	14.1
$3 \cdot 10^5$	10	0.25	51.3	81.3	98.9	132.0	155.0	164.0	31.8	51.7	61.2	80.8	95.1	100.0
$3 \cdot 10^5$	100	0.25	394.0	513.0	561.0	614.0	635.0	639.0	249.0	327.0	354.0	385.0	397.0	400.0
10^6	0.6	0.25	5.1	8.3	10.8	18.5	26.1	29.6	3.1	5.3	6.6	11.3	16.0	18.1
10^6	10	0.25	74.7	111.0	130.0	156.0	170.0	174.0	45.8	70.3	81.1	96.9	105.0	107.0
10^6	100	0.25	482.0	592.0	622.0	647.0	660.0	664.0	304.0	375.0	393.0	407.0	415.0	417.0
$3 \cdot 10^6$	0.6	0.25	1.6	2.6	3.2	5.0	6.6	7.3	1.0	1.6	2.0	3.1	4.0	4.5
$3 \cdot 10^6$	10	0.25	20.8	29.1	32.9	39.2	44.4	46.8	12.5	18.1	20.4	24.2	27.3	28.8
$3 \cdot 10^6$	100	0.25	116.0	142.0	153.0	170.0	181.0	186.0	71.4	88.7	95.0	105.0	112.0	115.0
$3 \cdot 10^5$	0.6	0.4	2.6	4.4	5.8	10.0	13.7	15.5	1.6	2.8	3.5	6.0	8.2	9.3
$3 \cdot 10^5$	10	0.4	41.6	59.9	69.2	84.2	91.3	92.8	26.1	37.8	42.7	51.4	55.6	56.5
10^6	0.6	0.4	4.7	7.6	9.8	16.9	24.8	29.0	2.9	4.8	6.0	10.3	15.1	17.7
10^6	10	0.4	67.9	102.0	122.0	165.0	194.0	205.0	41.8	64.7	76.1	101.0	119.0	126.0
10^6	100	0.4	474.0	630.0	691.0	757.0	784.0	790.0	296.0	399.0	435.0	473.0	489.0	492.0
$3 \cdot 10^6$	0.6	0.4	2.1	3.2	4.0	6.7	9.4	10.8	1.2	2.0	2.5	4.1	5.8	6.6
$3 \cdot 10^6$	10	0.4	26.8	39.3	46.3	59.1	68.1	71.8	16.0	24.6	28.7	36.4	41.9	44.1
$3 \cdot 10^6$	100	0.4	177.0	226.0	245.0	270.0	286.0	292.0	108.0	140.0	151.0	166.0	176.0	180.0

TABLE VI: Short LISA, pessimistic WD subtraction (gCLEAN, 3yr), volume-inc-averaged: $\overline{\mathrm{S/N}} = (1/2 \int (\mathrm{S/N})^3 \, d\cos\iota)^{1/3}$. All values are for $\beta = 0$, $\lambda = 0$. Effect of position-in-the-sky averaging is $\pm 10\%$.

retrograde orbit are 1.8×10^{-11} and 7.1×10^{-12}. Over three years, the total change in ι is less than 10^{-3} radians, so the constant inclination angle approximation should be valid.

The fluxes in Table VII are in good agreement, particularly the prograde fluxes. The kludged and Teukolsky fluxes are consistently close over the whole inspiral parameter space, and rarely differ by more than 10% [12]. The retrograde fluxes differ by $\approx 10 - 15\%$ in this example, but this inspiral is somewhat exceptional as it spends a long time in the strong field region near the central black hole.

To estimate how well the kludge waveforms predict the signal to noise, we must compare

ι	Time to plunge	Kludge $d(r/M)/dt$ (s^{-1})	Teukolsky $d(r/M)/dt$ (s^{-1})	Percentage difference
$\frac{\pi}{4}$	1 week	-1.72×10^{-7}	-1.81×10^{-7}	4.6
	1 month	-8.03×10^{-8}	-8.26×10^{-8}	2.8
	3 months	-4.43×10^{-8}	-4.48×10^{-8}	1.1
	1 year	-2.00×10^{-8}	-1.98×10^{-8}	-0.8
	3 years	-1.01×10^{-8}	-9.93×10^{-9}	-1.7
$\frac{3\pi}{4}$	1 week	-9.92×10^{-8}	-1.18×10^{-7}	16.
	1 month	-4.70×10^{-8}	-5.53×10^{-8}	15.
	3 months	-4.43×10^{-8}	-3.10×10^{-8}	14.
	1 year	-1.28×10^{-8}	-1.45×10^{-8}	12.
	3 years	-6.98×10^{-9}	-7.68×10^{-9}	9.0

TABLE VII: Comparison of inspiral rates between kludged inspirals and Teukolsky based inspirals.

the waveform amplitude. For a source at $1Gpc$ we compute the squared amplitude of the plus and cross gravitational wave polarisations, weighted by the LISA noise, i.e.,

$$|X|^2 = 4\,\Re\left[\int \frac{\tilde{X}^*(f)\,\tilde{X}(f)}{S_h(f)}\,\mathrm{d}f\right] \tag{30}$$

In this S_h denotes the LISA noise spectral density, as defined earlier. This amplitude measure was computed using a one day pure geodesic waveform segment at each of the points along the inspiral. The calculation was repeated for two different orientations of the source - $i = 0$ (source viewed from the pole) and $i = \pi/2$ (source viewed from the equatorial plane). These results are summarised in Table VIII.

The amplitudes appear to be in remarkably good agreement. Close to plunge, the kludge waveform somewhat underpredicts the amplitude as viewed from the equator, but appears to get the amplitude as seen from the pole quite accurately. Earlier in the prograde inspiral, the kludge waveform slightly overpredicts the amplitude but the difference is only a few per cent. This suggests that the kludge waveforms do a remarkable job of reproducing the Teukolsky results. However, although the amplitude appears to be similar, the spectral decomposition of the waveforms is quite different. Figure 4 shows the power $(\tilde{h}_+^*\tilde{h}_+ + \tilde{h}_x^*\tilde{h}_x)$ of the two waveforms as a function of frequency, as viewed from the pole and equator.

Figure 4 illustrates that the kludged waveforms contain only a few of the frequency harmonics seen in the Teukolsky waveform. However, the dominant contribution to the

ι	Time to plunge	Inclination	Kludge $\lvert h_+\rvert^2$	Teukolsky $\lvert h_+\rvert^2$	Kludge $\lvert h_x\rvert^2$	Teukolsky $\lvert h_x\rvert^2$
$\frac{\pi}{4}$	1 week	0.	1.60	1.58	1.60	1.58
		$\frac{\pi}{2}$	0.61	0.75	0.79	0.87
	1 month	0.	1.48	1.45	1.48	1.45
		$\frac{\pi}{2}$	0.57	0.67	0.74	0.77
	1 year	0.	0.97	0.95	0.97	0.95
		$\frac{\pi}{2}$	0.40	0.40	0.51	0.47
	3 years	0.	0.61	0.60	0.61	0.60
		$\frac{\pi}{2}$	0.25	0.25	0.34	0.30
$\frac{3\pi}{4}$	1 month	0.	0.18	0.20	0.18	0.20
		$\frac{\pi}{2}$	0.12	0.15	0.17	0.18
	1 year	0.	0.12	0.13	0.12	0.13
		$\frac{\pi}{2}$	0.081	0.096	0.11	0.12
	3 years	0.	0.078	0.082	0.078	0.082
		$\frac{\pi}{2}$	0.052	0.063	0.071	0.080

TABLE VIII: Comparison of waveform amplitudes between kludged inspirals and Teukolsky based inspirals. The amplitude $|X|^2$ is computed using expression (30).

signal comes from the quadrupole component, which is included in the kludged waveforms. In fact, the kludge waveform over-estimates the power in the quadrupole harmonics, and this somewhat compensates for the harmonics that it misses. The kludge does best for sources seen from the pole, since the quadrupole component is even more dominant in that direction. These results suggest that while the kludged waveforms would do quite badly (in this strong field regime) at matching a real inspiral waveform, the signal to noises that we are estimating using them are reasonably accurate. This particular example is perhaps unusually favourable, as the majority of the emission is at frequencies over the flat part of the sensitivity curve. For higher mass central black holes, the quadrupole peak would be pushed into the white dwarf noise, and the kludge would then underestimate the signal to noise. The discrepancy between the kludge and the Teukolsky results also decreases as the radius of the orbit is increased. Figure 5 is the equivalent of Figure 4, but for waveforms that are three years from plunge. In this example, the particle is still very close to the black

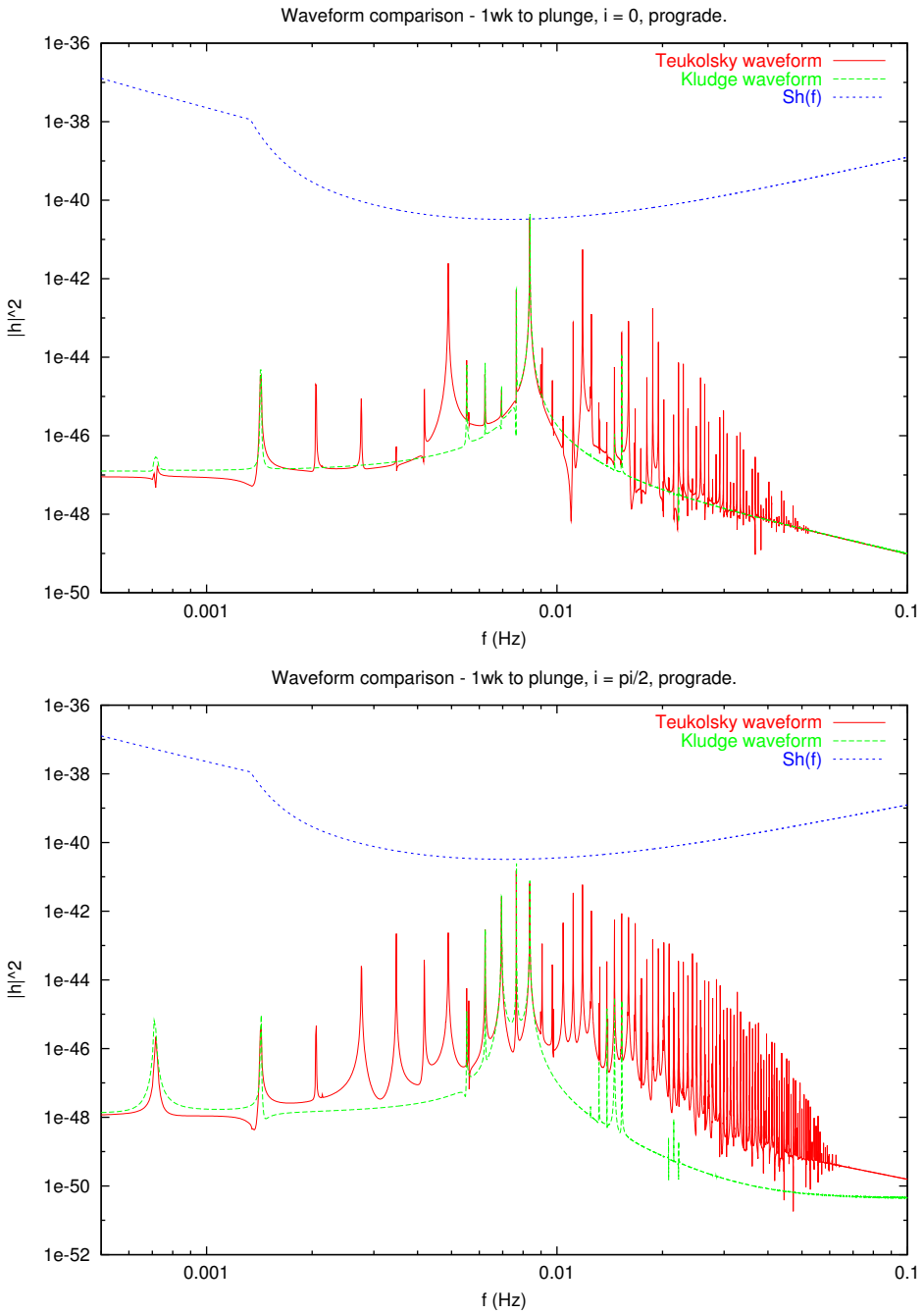

FIG. 4: Frequency spectra of kludged (green) and Teukolsky (red) waveforms, for a polar (top) and equatorial (bottom) viewing angle. Waveforms are generated from the instantaneous geodesic trajectory one week from plunge. The blue curve is the LISA noise spectral density, $S_h(f)$.

hole ($r = 5.8M$) at that time, but the amplitude of the higher harmonics is already being suppressed by an order of magnitude compared to the central quadrupole peak.

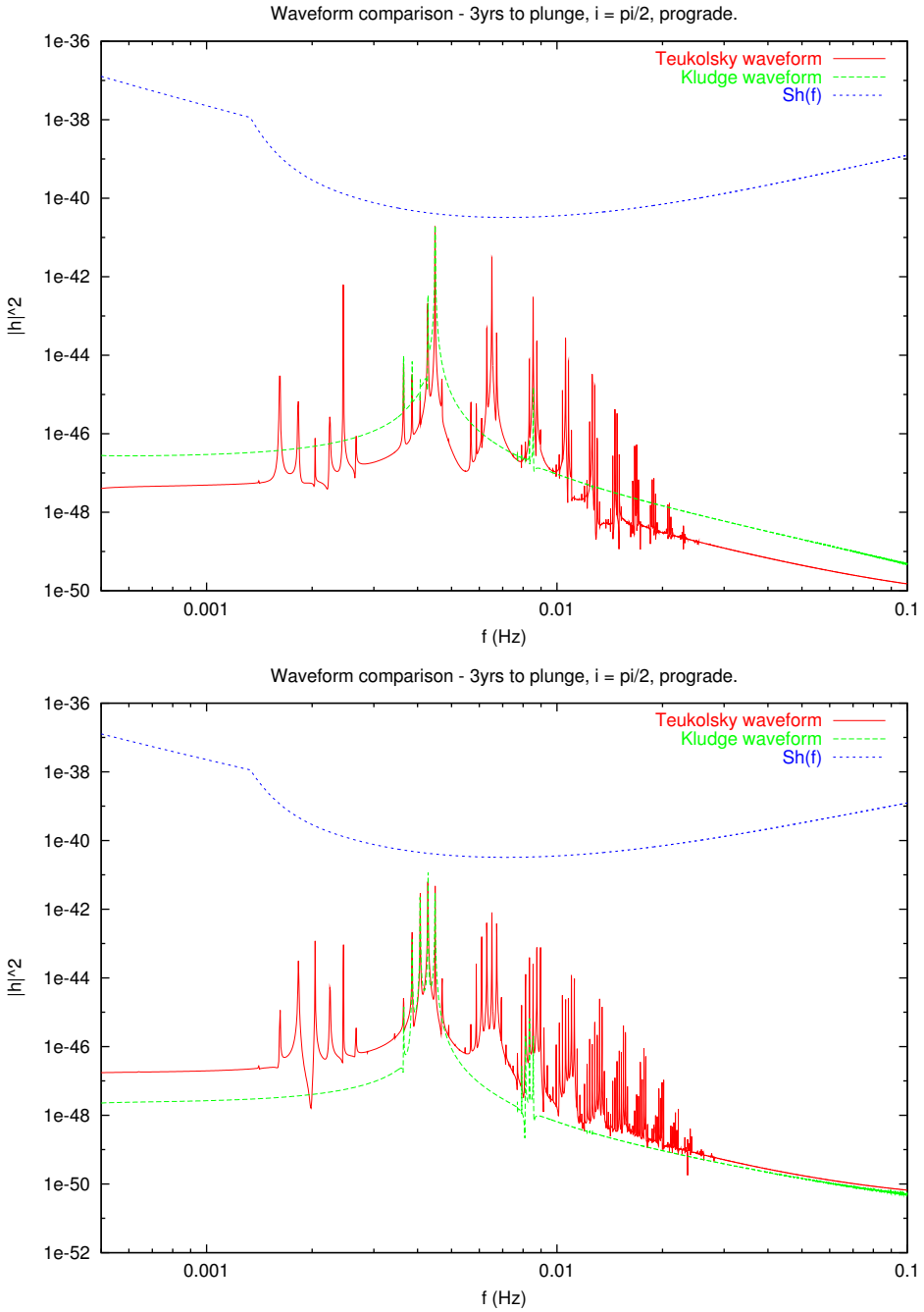

FIG. 5: As for Figure 4, but for waveforms three years from plunge.

VI. SEARCH STRATEGIES, TEMPLATE COUNTS, AND DETECTION THRESHOLDS

A. Basic Search Strategy

Even ignoring the spin and internal structure of the secondary body, it takes 14 parameters to characterize completely an inspiral orbit. These are:

- 3 constants defining the source location: normally taken to be ecliptic latitude β, ecliptic longitude λ, and distance d.
- 2 constants defining the orientation of the supermassive black hole's spin (i.e. the equatorial plane of the source system): for instance, its inclination i to the line of sight, and the angle ψ that its projected spin axis makes with a north-pointing meridian ray in the plane of the sky.
- 3 constant parameters of the system: the component masses M and m, and the magnitude of the primary's spin S.
- 6 phase-space parameters of the secondary body at some specified time t_0. For orbits in Kerr, three of these are "adiabatic" constants, which change gradually due to radiation reaction: for instance, the orbital periapse r_0, eccentricity e_0, and inclination ι_0 to the equatorial plane of the primary (or any other three parameters constructed from these). The other three are dynamical angle variables that change rapidly with the choice of t_0: for instance, the anomaly, apsidal precession, and nodical precession angles, all of which vary dynamically in the Kerr geometry even for test particles.

The data analysis task is simplified by the fact that only 8 of these parameters define the *intrinsic* radiation pattern in the source frame: the three constants of the system, the three "adiabatic" constants of the orbit, and *two* of the dynamical variables of the orbiting body. The remaining parameters are *extrinsic*: they affect only the projection of this radiation pattern onto the detector. The relevance of this distinction, as pointed out by Buonanno, Chen, & Vallisneri(2003), is that extrinsic parameters can be searched over cheaply, while intrinsic parameters are relatively expensive to search over. A quadrupolar gravitational wave can be decomposed into a linear combination of 5 orthonormal waveforms $h_i(\boldsymbol{\lambda}_I; t)$

that depend on intrinsic parameters $\boldsymbol{\lambda}_I$, with constant amplitudes $A_i(\boldsymbol{\lambda}_E)$ that depend on the extrinsic parameters $\boldsymbol{\lambda}_E$. For such a waveform, the optimal matched filter statistic ρ^2 can be obtained simply by summing the squares of the overlaps between the template waveforms $h_i(\boldsymbol{\lambda}_i; t)$ and the data $x(t)$:

$$\rho^2(\boldsymbol{\lambda}_I) = \sum_{i=1}^{5} \langle h_i(\boldsymbol{\lambda}_I), x\rangle^2 \ , \tag{31}$$

where

$$\langle h_i(\boldsymbol{\lambda}_I), x\rangle = 4\left[\int_0^\infty \frac{\tilde{h}_i^*(\boldsymbol{\lambda}_i; f)\tilde{x}(f)}{S_x(f)}\, df\right] \ , \tag{32}$$

$\tilde{}$ denotes the Fourier transform, * denotes complex conjugation, and $S_x(f)$ is the power spectral density of the noise in x. Actually, since LISA's output can be (approximately) represented as *two* synthesized Michelsons at 45° to one another, we have two data streams x_1 and x_2, and our optimal statistic is:

$$\rho^2(\boldsymbol{\lambda}_I) = \sum_{i=1}^{5}\sum_{j=1}^{2} \langle h_i(\boldsymbol{\lambda}_I), x_j\rangle^2 \ , \tag{33}$$

A second simplification we can make is to *fix* one of the "adiabatic" constants of our waveforms, say r_0, to some canonical value, and instead parameterize our waveforms by the time offset t_0 at which it passes through that value. The advantage of this is that time offsets can be included in the frequency-domain form of a template with a simple phase factor $e^{2\pi i f t_0}$, turning Eq. (32) into an inverse Fourier transform: the parameter t_0 (proxy for r_0) can be searched cheaply using fast Fourier transforms.

1. *Semicoherent method*

Even with these simplifications, it would exceed reasonable computational requirements to search the remaining 7 parameters with enough precision to maintain phase coherence over the full signal duration or mission lifetime. The solution is to break the data up into more manageable stretches and filter these separately. This data can then be recombined to recover much of the original signal-to-noise. What makes the technique *semi*-coherent is that some of the more rapidly-varying parameters are ignored when combining the stretches. This raises the detection threshold, but it also vastly reduces the precision at which one needs to sample the remaining parameters. (An analogy with radio pulsar searches is perhaps useful here. In stacked searches, one does not require the waveform phase to remain coherent from

stack to stack. Similary, in our semi-coherent LISA search, we will not require the three dynamical phase angles to remain coherent from segment to segment.)

One such approach is to divide the template waveforms into stretches of length T, and to overlap them against the LISA data. The shorter T is, the less precise the template parameters need to be to maintain a match with a putative signal, and the coarser we can sample the parameter space. We then explicitly maximize over the remaining two dynamically-varying angular parameters (these are roughly the mean anomaly and perihelion direction; the third is actually an extrinsic parameter and so was searched implicitly in the construction of the ρ^2 statistic), giving a filter output that is a function of the remaining 5 parameters $\vec{\Lambda}_k$ (plus time offsets t_k) for each interval T labeled by $k = 1, \ldots, N$. A given set of overall signal parameters $(\vec{\Lambda}, t_0)$ will then define a *trajectory* through the parameters of subsequent intervals, as the "adiabatic" parameters slowly evolve. To recover an imbedded signal, we select a set of overall signal parameters, and sum the filter output along the specified trajectory. (We note that the overall parameters will need to be selected from a finer sample than the parameters on each interval, in order to specify a trajectory uniquely.) If the parameters correspond to an actual signal in the data, then its contribution will grow linearly with the number of intervals, while the standard deviation of the noise will grow as the square root. This is discussed further in Sec. VI C.

This is only one of several possible semicoherent search methods. One possible change is to divide the templates into stretches not of constant time intervals T, but of constant intervals in periapse r_0, or constant intervals in the fundamental orbital frequency, or even constant increments in integrated signal-to-noise. Another change is to add a pre-threshold between the maximization over dynamical parameters and the summing along trajectories: points along the trajectory will be marginally more likely to exceed the threshold, so a well-chosen threshold can act as a kind of contrast enhancement while simultaneously paring down the data requirements of subsequent operations. It will probably take considerable work to fully optimize the search.

B. Template Counts for Coherent Integrations

Here we discuss the computational cost of the first stage of the hierarchical search: matched-filtered searches for short segments of capture waveforms in the LISA data. We

first briefly describe the formalism for calculating this.

1. *formalism*

Consider the 13-dim space of normalized waveforms $h_\alpha(t)$ (normalized so that $\langle \mathbf{h}|\mathbf{h}\rangle = 1$). The natural metric on this space is $g_{ab} = \frac{1}{2}\langle \partial_a h|\partial_b h\rangle$. We want to place a grid on this parameter space so that the average overlap of any waveform with the nearest template waveform (i.e., the nearest gridpoint) is $\mathcal{A}$. ($\mathcal{A}$ is a number that parametrizes the "coarseness" of the first-stage grid; we'll eventually want to find the most efficient value, but for now we imagine it is $\sim 0.8 - 0.9$.) The total number of template gridpoints this requires (assuming a hypercubic grid) is:

$$\mathcal{N}_{\rm temp} \approx (-12 \ln \mathcal{A})^{-N/2} N^{N/2} \int \sqrt{\Gamma}\; d\lambda^1 ... d\lambda^N, \tag{34}$$

where N is the dimension of the (normalized) parameter space (here $N = 13$).

Actually, for our purposes Eq. (34) must be modified in two ways. First, since the detection threshold will be set in large part by the computational cost of the search, we want to distinguish between the extrinsic parameters, which are "cheap" to search over, and the intrinsic ones. The number of intrinsic-parameter gridpoints to be searched over is

$$\mathcal{N}_{\rm int.temp} = (-12 \ln \mathcal{A})^{-\hat{N}/2} \hat{N}^{\hat{N}/2} \int \sqrt{\gamma}\; d\lambda^1 ... d\lambda^{\hat{N}}\,. \tag{35}$$

where γ_{ab} is average (over extrinsic paramters) of the (orthogonal) projection of Γ_{ab} onto the $\hat{N}(=7)$-dim subspace of intrinsic parameters.

The second modification occurs because in practice γ_{ab} usually has one extremely small eigenvalue, with corresponding eigenvector pointing mostly in the Φ_0 direction (Φ_0 is the orbital phase at t_0). That is, the parameter space is extremely "thin" in this direction, in which case the volume integral undercounts the required number of grid points. To account for this, we replace the nearly-zero eigenvalue(s) by one. We denote the modified determinant as $\sqrt{\hat{\gamma}}$.

Recall that here we are counting *not* entire, multi-year waveforms, but waveforms segments. To specify a segment, we need to specify not only the physical parameters of the full waveform, but also the length T of the segment and where the segment lies within the full waveform. The latter is parametrized by the radial orbital frequency ν_0 at the middle of the segment.

2. Template density

We have not yet had time to a full survey of $\sqrt{\hat{\gamma}}$ on the intrinsic space, but we have done some initial exploration of its dependence on the parameters. We find that $\sqrt{\hat{\gamma}}$ has only weak dependence on S/M^2, $\cos\lambda$ and $\tilde{\gamma}_0$, and has roughly the following scaling with the other physical parameters:

$$\sqrt{\hat{\gamma}} \approx 1.37 \cdot 10^6 \times (T/3\,\text{weeks})^{4.3} (\nu_0/1\,\text{mHz})^{7.7} (\mu/10M_\odot)^1 (M/10^6 M_\odot)^{3.1} (e_0/0.25)^{1.4}. \quad (36)$$

The pre-factor here corresponds to including *both* data channels (with the response of only a single 2-arm detector, the pre-factor drops by a factor ~ 3). This empirical approximate formula applies at integration times between ~ 1 week and ~ 4 weeks, in which range $\hat{\gamma}$ has dimension $\hat{N} = 6$ over most parameter space. We obtained this scaling formula by sampling parameter space around a single "central" point (with $\nu_0 = 0.9$ mHz, $\mu = 10M_\odot$, $M = 10^6 M_\odot$, $S/M^2 = 0.8$, $\cos(\lambda) = 0.5$, $e_0 = 0.25$). However, we also performed a limited Monte-Carlo survey over the entire parameter space, which suggests there are no "surprising" areas with densities that may violate significantly our scaling formula.

3. template count

To implement Eq. (36) in Eq. (35), we consider separately 20 different classes of sources, sorted by their M and μ (see the table below). Each class spans $\Delta \ln M = 1$ and $\Delta \ln \mu = 1$. For each class, we consider sources with all possible e_0, S, $\cos(\lambda)$, and $\tilde{\gamma}_0$. The volume integral in Eq. (35) over these 4 parameters is

$$\int_0^{0.4} (e_0/0.25)^{1.4} de_0 \int_0^1 d(S/M^2) \int_{-1}^1 d(\cos\lambda) \int_0^{2\pi} d\tilde{\gamma}_0 = 0.32 \times 1 \times 2 \times 2\pi = 4. \quad (37)$$

For a given data stretch with duration T and central frequency ν_0, and a given source class with $M_{\min} < M < M_{\max}$ and $\mu_{\min} < \mu < \mu_{\max}$, Eq. (35) becomes

$$\begin{aligned} \mathcal{N}_{\text{int.temp}} &= 1.2 \cdot 10^9 \times (-12 \ln \mathcal{A})^{-3} (T/3\,\text{week})^{4.3} (\nu_0/1\,\text{mHz})^{7.7} \\ &\times \int_{\mu_{\min}}^{\mu_{\max}} (\mu/10M_\odot)(d\mu/\mu) \int_{M_{\min}}^{M\max} (M/10^6 M_\odot)^{3.1} (dM/M), \end{aligned} \quad (38)$$

where we specified $\hat{N} = 6$. Table I gives $\mathcal{N}_{\text{int.temp}}$ for the various source classes, for $T = 3$ weeks, $\nu_0 = 1$ mHz, and $\mathcal{A} = 0.9$. With $\mathcal{A} = 0.8$, all entries are smaller by a factor ~ 10.

	$1.0e5 < M < 2.7e5$	$2.7e5 < M < 7.4e5$	$7.4e5 < M < 2.0e6$	$2.0e6 < M < 5.5e6$
$0.5 < \mu < 1.3$	$2.5 \cdot 10^5$	$5.8 \cdot 10^6$	$1.3 \cdot 10^8$	$2.9 \cdot 10^9$
$1.3 < \mu < 3.7$	$7.6 \cdot 10^5$	$1.7 \cdot 10^7$	$3.8 \cdot 10^8$	$8.7 \cdot 10^9$
$3.7 < \mu < 10$	$2.0 \cdot 10^6$	$4.6 \cdot 10^7$	$1.0 \cdot 10^9$	$2.3 \cdot 10^{10}$
$10 < \mu < 27$	$5.4 \cdot 10^6$	$1.2 \cdot 10^8$	$2.7 \cdot 10^9$	$6.2 \cdot 10^{10}$
$27 < \mu < 74$	$1.5 \cdot 10^7$	$3.4 \cdot 10^8$	$7.4 \cdot 10^9$	$1.7 \cdot 10^{11}$

TABLE IX: Number of templates required for 3 week long integration centered at $\nu_0 = 1$ mHz, for various CO and MBH mass ranges, and with both data channels. All masses are in $M_\odot$. The average match has been set here to 0.9. With $\mathcal{A} = 0.8$, one need only divide all entries by ~ 10. To re-scale for other T and ν_0, use the approximated scaling factors in Eq. (36)

4. Maximal length of time segment

We would prefer to do optimal filtering on segments that are as long as possible. What can we afford? Imagine that our computer in 2013 has effective power of 50 Teraflops, so that in 2 years running time it can perform $\sim 3 \times 10^{21}$ ops, and that roughly a third of this power is expended on the coherent integration stage of the search. Imagine that we have $\tau \sim$ 2yr's worth of data, and that the segments we'll use in the coherent integration have length T. The cost of the coherent stage is then

$$\sim 10 N_{int.temp}(\tau/T)[3 f_{max}\tau \log_2(f_{max}\tau)] \tag{39}$$

where the initial factor of 10 comes from the fact that there are 5 independent components of the quadrupolar waveform and 2 measured polarizations; $f_{max} \sim 3 \times 10^{-2}$Hz is an upper-frequency cut-off we'll impose at the first stage to limit computational cost; and the factor $[3 f_{max}\tau \log_2(f_{max}\tau)]$ comes from the search (using FFTs) over all time-translations of the template. Then we can afford for $N_{int.temp}$ to be $\sim 10^{12}(T/\tau) \sim 10^{10}$:

$$\text{maximal \# of templates in one coherent step} \sim 10^{10}. \tag{40}$$

The table above shows that this allows for $\sim$ 3-week segments for all populations of COs and MBHs, except BHs captured by MBHs heavier than $\sim 2 \cdot 10^6 M_\odot$. For the latter we can only have $T \sim 10 - 14$ days (at $\mathcal{A} = 0.9$). In estimating event rates we will assume that we can use three week long stacks in the 'optimistic' case, but will assume only two week long stacks for the 'pessimistic' case.

C. Detection Thresholds

In the presence of Gaussian noise only, the ρ^2 statistic defined in Eq. (33) is distributed as a χ^2 with 10 degrees of freedom: it has a mean of 10 and a standard deviation of $2\sqrt{5} \approx 4.47$. The statistic P_k is the maximum of ρ^2 along some slice through parameter space (corresponding to the most rapidly-changing variables), with varying degrees of correlation along this slice: the resulting distribution is analytically intractable, but relatively straightforward to estimate using Monte-Carlo simulations. We characterize the distribution by its mean μ_k and standard deviation σ_k, and find that, over a broad range of signal parameters, $\mu_k \sim 18$ and $\sigma_k \sim 4.5$. While these statistics may not be normally-distributed, their sum along a trajectory $P = \sum_{k=1}^{N} P_k$ will tend towards a normal distribution (by the central limit theorem), with a mean $\mu = N\mu_k$ and standard deviation $\sigma = \sqrt{N}\sigma_k$.

Our detection threshold can then be characterized by the normal Z-score, $Z = (P-\mu)/\sigma$, that must be achieved for a signal to qualify as a detectable event (or event candidate). Now suppose we want our entire search to have a false alarm probability of 1% from noise alone. The number of independent trajectories that we search is difficult to compute exactly, but will certainly exceed the 10^{10} templates over $\vec{\Lambda}_k$ computed in Sec. VI B. For a 3-year observation, each of these trajectories has a number of independent time offsets t_0 of order $3\text{yr} \times 1\text{mHz} \sim 10^5$. So our false alarm probability per trajectory per time offset should not exceed 10^{-17}, for a Z-score of 8.8. Changing any of the assumptions by an order of magnitude changes the Z-score by less than 0.3 — when you're this far into the tail of the distribution, the threshold is relatively insensitive to large variations in false alarm probability. Even if one assumed a hierarchical search strategy, where millions of candidate events could be detected and then winnowed out by subsequent analyses, the threshold Z-score would only go down to 7. For definitiveness, we will assume that a "typical" Z-score for a search will be around 8.

The next question is: How strong must a signal be to exceed this threshold? Signal strengths are usually measured in terms of their intrinsic signal-to-noise ratio $\text{SNR} = \sqrt{\langle s, s\rangle}$ over the entire observation, with the inner product defined in Eq. (32). However, the signal measured on each stretch of data is reduced by two factors. First, each stretch is only $1/N$ of the total observation, so $\langle s, s\rangle$ is reduced by $1/N$. Second, the coarseness of our template bank introduces a factor $\mathcal{M} < 1$, representing how closely the signal matches the nearest

template in the bank. (This includes not only the factor $\mathcal{A}$ discussed in Sec. VI B, but other reduction factors due to the discreteness of our method, such as finite time sampling rate.) These two factors tend to be complementary: if one chooses a bank with higher $\mathcal{M}$, the templates are placed more densely, so computational resources limit us to shorter data stretches and hence higher N. Our calculations of template densities have assumed an overall match factor of $\mathcal{M} \sim 0.8 - 0.9$.

The presence of a signal will increase the expectation value of ρ^2 for the nearest template from 10 to $10 + (\mathcal{M}/N)\langle s, s\rangle$. Assuming that the maximization over the two dynamic angles will usually arrive at the correct values corresponding to the imbedded signal, then the corresponding P_k's will have a similar value, and their sum P will have N times that value. Setting this equal to the value required to exceed our threshold Z-score $Z_{\rm thresh}$, we obtain the threshold intrinsic signal-to-noise ratio:

$$\mathrm{SNR}_{\rm thresh} = \sqrt{\frac{N}{\mathcal{M}}\left(\mu_k - 10 + Z_{\rm thresh}\frac{\sigma_k}{\sqrt{N}}\right)}\,. \tag{41}$$

The LHS is the (optimal, matched-filtering) signal-to-noise ratio $\sqrt{\langle s, s\rangle}$, coming from *both* synthetic Michelsons, that a signal must have in order to have a good chance of being detected, using our method. The variation due to noise is fairly narrow at this level, so we can assume to good approximation that signals will be detected if and only if they exceed this threshold.

For a typical case of a 3-year observation divided into 3-week stretches, assuming an overall match factor of $\mathcal{M} = 0.8$ and $Z_{\rm thresh} = 8$, we get $SNR_{\rm thresh} = 29$. Changing the threshold Z-score by ± 1 changes this number by only ± 0.7.

Note that even with infinite computing power (which would allow an optimal search), an SNR of ~ 14 would be required to insure a $< 1\%$ false alarm probability (owing to the vast number of templates effectively searched over). Thus one could say that computing-power limitations reduce LISA's reach by a factor ~ 2, for these sources.

VII. SUMMARY OF DETECTION RATES

The final event rate that can be detected by LISA can be computed from the detection signal to noise ratios Tables III, IV, V and VI, and the EMRI rates per unit volume $\mathcal{R}$ in table II. Since for a source at distance D, $SNR(D) \propto D^{-1}$, the number of detected mergers

in time T is, for $D \ll 3.5$Gpc (corresponding to $z = 1$, by when the universe is no longer Euclidean and evolution is important)

$$N_{det} = (4\pi/3) D^3 \mathcal{R} T \left[\frac{SNR(D)}{\text{thresh}} \right]^3 \tag{42}$$

which is independent of the fiducial D (here chosen to be 1 Gpc); thresh is the threshold, set by the requirement of a false alarm rate less than 10^{-4} (for coherent integration, $P_{false} \sim N_{\text{templates}} \exp[-(SNR)^2/2]$).

We consider two cases, one optimistic and one pessimistic:

- For the *optimistic* case, we assume 5 years of LISA data, using the optimal AET signal combination, with the optimistic confusion background estimates, and assume that we can integrate 3-week coherent stretches of data (giving $\text{SNR}_{\text{thresh}} \sim 36$).

- For the *pessimistic* case, we assume 3 years of LISA data, using only the X signal, with the pessimistic confusion background estimates, and assume that we can integrate 2-week coherent stretches of data (giving $\text{SNR}_{\text{thresh}} \sim 34$).

For each case we compute the *total* number of events in the 5- or 3-year mission, for each of the 17 fiducial sources in the S/N tables, for both standard and short-arm LISA designs *assuming that the universe is flat Minkowski space.* We then average the numbers from $e = 0.4$ and $e = 0.25$ because we expect comparable contributions to the EMRI rates from circular (large initial peribothron) and eccentric (small initial peribothron). The resulting source numbers are given in Table X When $SNR(1\text{Gpc}) > 120$ (i.e. $D_{\max} > 3.5 Gpc$), the flat Minkowski space estimate of equation 42 is not to be trusted: the volume of space increases only as $\sim D$, not D^3, and redshifting of frequency and source evolution are important effects, not considered here. Instead in these cases (marked with a * in the table) we have instead entered in the table the crude *lower limit* to the total number of events within $z < 1$, all of which LISA can detect.

$$N_{det} = V_c(z < 1) \mathcal{R} T \tag{43}$$

The comoving volume $V_c(z < 1) = 199\text{Gpc}^3$.

$M_\bullet$	m	LISA		Short LISA	
		Optimistic	Pessimistic	Optimistic	Pessimistic
300 000	0.6	8	0.7	14	1
300 000	10	739	89	902	115
300 000	100	1*	1*	1*	1*
1 000 000	0.6	94	9	80	7
1 000 000	10	1000*	800	1000*	502
1 000 000	100	1*	1*	1*	1*
3 000 000	0.6	67	2	11	0.3
3 000 000	10	1700*	134	816	25
3 000 000	100	2*	1*	2*	1

TABLE X: Columns 3-6 give the number of EMRI events LISA can see for merger of body of mass m (column 2) into supermassive black hole of mass $M_\bullet$ (column 1). Columns 3-4 are for the normal 5×10^6km baseline. Columns 5-6 are for a 1.6×10^6km baseline. *Optimistic* uses all 3 TDI variables for 5 years, with ideal white dwarf background removal and assuming 3 week stretches for the coherent search. *Pessimistic* uses only a single pair of arms for 3 years, with current gCLEAN white dwarf removal and assuming 2 week coherent stretches. Entries marked with a * are $z < 1$ lower limits computed from equation 43, since LISA can detect all sources out to $z \gg 1$, and evolution is unknown. All other entries computed from the Euclidean equation 42, since LISA cannot see the sources to cosmological distances.

VIII. DISCUSSION AND RECOMMENDATIONS

We conclude that there is no strong reason to push for a shorter LISA baseline. With the fiducial numbers presented here, LISA should detect $\sim 10^3$ EMRI events, mainly stellar mass black hole inspirals, even with the loss of one arm. Only tens of white dwarf inspirals will be seen, and the Pop III $100M_\odot$ inspiral numbers are to sensitive to cosmological evolution and redshift effects for the current calculations to give reliable numbers, beyond the fact that one could be seen at $z < 1$. With conservative rates (1/100 the fiducial ones for white dwarfs, and 1/10 the fiducial ones for black holes), the detection of even a single white dwarf event becomes marginal, but the stellar mass black hole signals are robust.

[1] B. Abbott *et al.*2003 "Analysis of First LIGO Science Data for Stochastic Gravitational Waves", gr-qc/0312088

[2] M.C. Aller and D. Richstone 2002 AJ 124 3035

[3] M. Bernardi et al 2003 AJ (astro-ph/0301631)

[4] N.J. Cornish, gr-qc/0304020

[5] T. Creighton 1 Oct 2003 "Semicoherent search methods for EMRI data analysis" at http://www.tapir.caltech.edu/listwg1/EMRI/Meetings/

[6] J-P. De Villiers, J.F. Hawley & J.H. Krolik 2003 "Magnetically Driven Accretion Flows in the Kerr Metric I: Models and Overall Structure," astro-ph/0307260

[7] A.J. Farmer and E.S. Phinney 2003 MNRAS 346, 1197

[8] A.V. Filippenko and L.C. Ho 2003 ApJ 588 L13. "A Low-Mass Central Black Hole in the Bulgeless Seyfert 1 Galaxy NGC 4395"

[9] M. Freitag 2001 "Monte Carlo cluster simulations to determine the rate of compact star inspiraling to a central galactic black hole," Class Quantum Gravity 18, 4033 (astro-ph/0107193).

[10] Gebhardt, K. et al 1996 AJ 112, 105 "The Centers of Early-Type Galaxies With HST. III. Non-Parametric Recovery of Stellar Luminosity Distribution"

[11] Gebhardt, K. et al 2001 AJ 122, 2469 "M33: A Galaxy with No Supermassive Black Hole"

[12] K. Glampedakis, S. Hughes and D. Kennefick, Phys. Rev. D **66**, 064005 (2002).

[13] Glampedakis, K. and Kennefick, D., Phys. Rev. D **66**, 044002 (2002).

[14] Hils, D and Bender, P. 1995 ApJL 445, L7 "Gradual approach to coalescence for compact stars orbiting massive black holes"

[15] S. A. Hughes & R.D. Blandford 2003 "Black hole mass and spin coevolution by mergers," ApJ 585, L101.

[16] Hughes, S., Phys. Rev. D **61**, 084004 (2000), Phys. Rev. D **64** 064004 (2001).

[17] S. Hughes 2002 MNRAS 331, 805

[18] Madau, P. & Rees, M.J. 2001 ApJL 551, L27 "Massive Black Holes as Population III Remnants"

[19] Merrit & Ferrarese 2001 ApJ 547, 140

[20] Miralda-Escudé, J., & Gould, A. 2000, ApJ, 545, 847 "A Cluster of Black Holes at the Galactic

Center"

[21] G. Nelemans, L.R. Yungelson and S.F. Portegies Zwart 2001 A& A 375, 890

[22] Rest, A. et al 2001 AJ 121, 2431 "WFPC2 Images of the Central Regions of Early-Type Galaxies. I. The Data"

[23] P. Salucci et al 1999 MNRAS 307, 637 (erratum 311, 448)

[24] P. Salucci et al 2000 MNRAS 317, 488

[25] P. Salucci and M. Persic 1999 MNRAS 309, 923

[26] R.K. Sheth, M. Bernardi, P. Schechter et al 2003 ApJ 594, 255 (astro-ph/0303092)

[27] Sigurdsson, S., & Rees, M. J. 1997, MNRAS, 284, 318 "Capture of stellar mass compact objects by massive black holes in galactic cusps"

[28] K. S. Thorne 1974 "Disk Accretion onto a Black Hole. II. Evolution of the Hole," ApJ 191, 507.

[29] S. Tremaine et al 2002 ApJ 574, 740

[30] R.F. Webbink and Z. Han 1998 AIPC 456, 61

[31] Q. Yu and S. Tremaine 2002 MNRAS 355, 965

[32] Similar expressions can be generated for other TDI combinations.

[33] Note the RHS of our Eq. (23) is a factor $\frac{3}{4}$ as large as the RHS in Eq. (3.4) in[1]; this difference arises simply because the angle between any two LISA arms is $\pi/3$ (instead of the $\pi/2$ for LIGO's arms), and $\sin^2(\pi/3) = 3/4$.

[34] Note our prefactor 2.1×10^{-45} is a factor ~ 25 lower than the prefactor cited in Hughes [17], based on his private communication with S. Phinney. This large discrepancy seems to be the product of the following two factors. First, it was based on the estimate of Ω_{GW}^{gal} by Webbink and Han [30], which is ~ 3 times larger than the result of Nelemans *et al.* [21]. Second, it contained a factor 20/3 error due to a misunderstanding of Phinney's normalization convention.