Graviton mass bounds from space-based gravitational-wave observations of massive black hole populations

Emanuele Berti,1,2 Jonathan Gair,3,4 and Alberto Sesana1,4

1Department of Physics and Astronomy, The University of Mississippi, University, MS 38677, USA
2California Institute of Technology, Pasadena, CA 91109, USA
3Institute of Astronomy, University of Cambridge, Cambridge, CB3 0HA, UK
4Max-Planck Institut für Gravitationsphysik (Albert-Einstein-Institut), Am Mühlenberg 1, D-14476, Potsdam, Germany
(Dated: October 6, 2011)

Space-based gravitational-wave (GW) detectors, such as LISA or a similar ESA-led mission, will offer unique opportunities to test general relativity. We study the bounds that space-based detectors could realistically place on the graviton Compton wavelength $\lambda_g = \hbar/(m_g c)$ by observing multiple inspiralling black hole (BH) binaries. We show that while observations of individual inspirals will yield mean bounds $\lambda_g \sim 3 \times 10^{15}$ km, the combined bound from observing $\sim 50$ events in a two-year mission is about ten times better: $\lambda_g \approx 3 \times 10^{14}$ km ($m_g \approx 4 \times 10^{-26}$ eV). The bound improves faster than the square root of the number of observed events, because typically a few sources provide constraints as much as three times better than the mean. This result is only mildly dependent on details of BH formation and detector characteristics. The bound achievable in practice should be one order of magnitude better than this figure (and hence almost competitive with the static, model-dependent bounds from gravitational effects on cosmological scales), because our calculations ignore the merger/ringdown portion of the waveform. The observation that an ensemble of events can sensibly improve the bounds that individual binaries set on $\lambda_g$ applies to any theory whose deviations from general relativity are parametrized by a set of global parameters.

The formulation of gravitational theories with nonzero mass for the graviton that are consistent with cosmological observations is an important open problem. Attempts to construct such theories led to well-known conceptual difficulties, such as the so-called van Dam-Veltman-Zakharov (vDV-Z) discontinuity 1–3, due to the fact that the helicity-0 component of the graviton does not decouple from matter when the putative mass of the graviton $m_g \rightarrow 0$. To circumvent pathologies related to the vDV-Z discontinuity, various versions of Lorentz-violating massive graviton theories have been proposed in recent years 4. Massive graviton signatures in the CMB and possible constraints on $m_g$ from cosmological observations are an active area of research (see e.g. 2).

In this paper we are interested in hypothetical massive graviton theories as “straw men” for alternative theories of gravity in which the propagation speed of gravity differs from that of electromagnetic waves, leading to a modified dispersion relation. Therefore we will adopt a phenomenological point of view and ask the following question: if the graviton mass were nonzero, what upper bounds on $m_g$ could we set by gravitational-wave (GW) observations of inspiralling compact binaries with future space-based detectors? Using $\lambda_g = \hbar/(m_g c)$, upper limits on the graviton mass $m_g$ (in eV) can be expressed as lower limits on its Compton wavelength $\lambda_g$ (in km):

$$\lambda_g [\text{km}] \times m_g [\text{eV}] = 1.24 \times 10^{-9}. \quad (1)$$

Our analysis will show that, in hierarchical models of massive BH formation 6, 7, the bound on $\lambda_g$ from GW observations of a population of merger events is about one order of magnitude better than the mean bound from GW observations of individual mergers.

<table>
<thead>
<tr>
<th>Current bounds</th>
<th>$\lambda_g [\text{km}]$</th>
<th>$m_g [\text{eV}]$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary pulsars</td>
<td>$1.6 \times 10^{10}$</td>
<td>$7.6 \times 10^{-20}$</td>
<td>8</td>
</tr>
<tr>
<td>Solar System†</td>
<td>$2.8 \times 10^{12}$</td>
<td>$4.4 \times 10^{-22}$</td>
<td>9, 10</td>
</tr>
<tr>
<td>Clusters†</td>
<td>$6.2 \times 10^{15} h_0$</td>
<td>$2.0 \times 10^{-29} h_0^{-1}$</td>
<td>11</td>
</tr>
<tr>
<td>Weak lensing†</td>
<td>$1.8 \times 10^{22}$</td>
<td>$6.9 \times 10^{-32}$</td>
<td>12</td>
</tr>
<tr>
<td>Proposed bounds</td>
<td>$\lambda_g [\text{km}]$</td>
<td>$m_g [\text{eV}]$</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------</td>
<td>------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Pulsar timing#</td>
<td>$4.1 \times 10^{14}$</td>
<td>$3.0 \times 10^{-23}$</td>
<td>13</td>
</tr>
<tr>
<td>White dwarfs*</td>
<td>$1.4 \times 10^{14}$</td>
<td>$8.8 \times 10^{-24}$</td>
<td>14</td>
</tr>
<tr>
<td>EM counterparts*</td>
<td>$10^{-15} - 10^{-16}$</td>
<td>$10^{-24} - 10^{-25}$</td>
<td>15</td>
</tr>
</tbody>
</table>

TABLE I. Graviton mass bounds. For proposed methods we quote the best achievable bounds. In the notation of the main text, a dagger † denotes static bounds; a number sign ‡, dynamical bounds; an asterisk *, bounds that could be achieved by comparing GW and electromagnetic observations.

To put our results in context, in Table II we present a non-exhaustive summary of current and proposed bounds on $\lambda_g$ that do not rely solely on GW observations.

These bounds can be roughly divided into three classes. i) Static bounds. If the graviton has nonzero mass, it is reasonable to expect that the Newtonian gravitational potential will be modified to the Yukawa form...
in the nonradiative near zone of any body of mass \( M \): 
\[ V(r) = \frac{GM}{r} \exp(-r/\lambda_g). \]
Talmadge et al. investigated deviations from Kepler’s third law for the inner planets of the Solar System. By translating the uncertainties in these measurements in terms of accelerations of test bodies resulting from a Yukawa potential, Will found that the strongest bound on \( \lambda_g \) comes from the very nearly Keplerian orbit of Mars: \( \lambda_g > 2.8 \times 10^{12} \text{ km} \).

Some bounds on \( m_g \) quoted by the Particle Data Group (PDG) similarly assume Yukawa corrections to the Newtonian potential in the weak-field limit. The strongest bounds are naturally obtained from observations on the largest (cosmological) scales. As early as 1974, Goldhaber and Nieto assumed that the graviton mass would produce a Yukawa-type correction to the standard Newtonian potential and argued that the evidence for bound clusters and tidal interactions between galaxies should imply a range for gravity at least as large as a few Mpc, so that \( \lambda_g > 6.2 \times 10^{15} h_0 \text{ km} \). However, the need to include dark matter to explain galactic rotation curves indicates that there are complications in the nature of gravity on those scales that are not necessarily well characterized by a Yukawa-type potential. An even stronger bound \( \lambda_g > 1.8 \times 10^{22} \text{ km} \) comes from weak gravitational lensing, because no distortions are observed in the measured values of the variance of the power spectrum. Because of uncertainties in the amount and dynamics of dark matter in the Universe (and in the absence of a consistent massive graviton theory compatible with cosmology) these bounds should be regarded as model-dependent and viewed with some caution.

ii) Dynamical bounds. All bounds listed in part i) are static, in the sense that they do not probe features related to the propagation of the gravitational interaction when \( m_g \neq 0 \). The best available dynamical bounds come from the (indirect) observations of GWs from binary pulsars, that are in excellent agreement with general relativity. Finn and Sutton observed that the consistency of the orbital decay rates of binary pulsars PSR B1534+12 and B1913+16 with general relativistic predictions yields a “dynamical” bound \( \lambda_g > 1.6 \times 10^{10} \text{ km} \). A recent idea is to set bounds on \( m_g \) using Pulsar Timing Arrays. The best bound achievable in the near future would be \( \lambda_g \approx 4.1 \times 10^{13} \text{ km} \), but this figure could worsen by an order of magnitude depending on the number of pulsars used for the test, timing accuracy and observation time.

iii) Comparisons of gravitational and electromagnetic observations. If \( m_g \neq 0 \), the modified dispersion relation for GWs would result in different arrival times of GWs and electromagnetic waves emitted by the same astrophysical source. Cutler et al. proposed to correlate electromagnetic observations and future space-based GW observations of white dwarf binaries. The best bound that could be obtained in this way is \( \lambda_g \approx 1.4 \times 10^{14} \text{ km} \), but realistic bounds would probably be worse by about one order of magnitude.

Kocsis et al. proposed to correlate LISA observations of GWs from massive BH binary mergers with their possible electromagnetic counterparts. An intrinsic limitation of this method is related to timing uncertainties in the GW burst, which are comparable to the dynamical timescale for the coalescing binary during merger, and can be estimated as the inverse of the orbital frequency at the innermost stable circular orbit. For binaries in the mass range \( M = 10^5 - 10^7 \) at redshift \( z = 1 \), this uncertainty leads to a best bound \( \lambda_g \sim 10^{15} - 10^{16} \text{ km} \), worse than bounds coming from GW observations alone (as we will see below). Systematic, model-dependent uncertainties in the electromagnetic counterpart will further weaken graviton mass bounds achievable in this way.

### I. Bounds on \( \lambda_g \) from Individual Gravitational-Wave Observations

Tight dynamical bounds on the graviton mass can be obtained using GW observations in space. This is due to two reasons: (1) the larger mass of observable BH binaries, as compared to ground-based GW observations; (2) the statistical increase in the bound that would result from observing populations of individually resolved binaries. To illustrate point (1), in this section we review existing work on massive graviton bounds from GW observations of individual binaries with LISA. In section II we will present the first attempt to quantify the statistical improvement of bounds on \( \lambda_g \) that could be achievable in reality by observing several events in the context of hierarchical models of massive BH formation.

Will first pointed out that interesting bounds on \( \lambda_g \) could come from a careful monitoring of the phase of GWs emitted by binaries of compact objects, such as BHs and/or neutron stars. In hypothetical massive graviton theories, the GW damping formulae and dispersion relation would be modified. As a consequence, the GW phasing \( \Psi_{GR}(f) \) would acquire an additional term:

\[ \Psi_{MG}(f) = \Psi_{GR}(f) - \beta_g(\pi M f)^{-1}, \]

where \( \beta_g = \pi^2 DM/[(1+z)^3] \) and \( D \) is a distance parameter, similar to (but not quite the same as) the luminosity distance \( D_L \) (here and below we assume a \( \Lambda \)CDM model with \( H_0 = 70 \text{ km s}^{-1}\text{Mpc}^{-1} \), \( \Omega_M = 0.3 \), \( \Omega_{\Lambda} = 0.7 \)).

Using the restricted post-Newtonian approximation and neglecting binary spins, Will estimated that stellar-mass binary inspirals to be observed with LIGO would yield a bound \( \lambda_g \approx 5 \times 10^{12} \text{ km} \), only slightly better than Solar System bounds. For an “optimal” LISA system, i.e., an equal-mass BH binary of total mass \( M = 2 \times 10^6 M_\odot \) at \( D_L = 3 \text{ Gpc} \) (\( z \approx 0.5 \)), the bound would be four orders of magnitude better: \( \lambda_g \approx 5 \times 10^{16} \text{ km} \approx 1.6 \text{ kpc} \).

These initial estimates were refined in various papers. Will and Yunes showed that bounds from binary BH inspirals at \( D_L = 3 \text{ Gpc} \) would range between \( 10^{15} \text{ km} \) and \( 5 \times 10^{16} \text{ km} \) for \( M \) in the range \( 10^4 - 10^7 M_\odot \). They found that the bound on \( \lambda_g \) is proportional to \( \sqrt{L} \) (where \( L \) is the LISA armlength) and to (LISA acceleration noise)\(^{-1/2} \), and that it scales in the following way:
\[ \lambda_g \propto \left( \frac{D}{(1+z)D_L} \right)^{1/2} \frac{M^{11/12}}{S_0^{1/4}f_0^{1/2}}, \]

where \( M = (m_1 m_2)^{3/5} M^{-1/5} \) is the “chirp mass”, \( S_0 \) (in Hz\(^{-1}\)) sets the scale of the noise power spectral density (PSD) of the detector, and \( f_0 \) is a characteristic “knee” frequency where the PSD has a minimum. Distance dependence is weak because the effect of the massive graviton and measurement errors both grow with distance.

Berti et al. [14] performed a more detailed Monte Carlo calculation of LISA’s parameter estimation capabilities. They still considered restricted post-Newtonian inspiral waveforms, but they used Cutler’s model [20] to take into account the motion of the detector (all previous analyses assumed a static LISA constellation). In this paper we will do the same. In Cutler’s model, LISA can be seen as one (two) independent “LIGO-like” Michelson interferometers depending on whether 4 (5/6, respectively) laser links are available between the three satellites forming the constellation. For the “optimal binary” (a nonspinning, equal-mass binary with \( M = 2 \times 10^6 M_\odot \) at \( D_L = 3 \) Gpc), Monte Carlo averages over the binary position and orientation yield a bound \( \lambda_g = 5.0 \times 10^{16} \) (3.7 \( \times 10^{16} \)) km when using two (one) Michelsons, in excellent agreement with Will’s original results. For masses in the range \( M = 2 \times 10^4 \) to \( 2 \times 10^7 M_\odot \), the mean bound using two Michelsons at \( D_L = 3 \) Gpc is in the range \( \sim 5 \times 10^{15} \) to \( 7 \times 10^{16} \) km.

Later work used mostly Cutler’s model and investigated the effect of more complex inspiral waveforms. Arun and Will [21] studied the effect of higher harmonics and PN amplitude corrections. They found that bounds on \( \lambda_g \) improve by a factor of a few for systems of total mass \( \gtrsim 10^6 M_\odot \), and that this improvement is more significant for binaries with large mass ratios. If one ignores spin precession, adding spins to the waveforms generally introduces degeneracies between the binary parameters, degrading parameter estimation accuracy [19]. Stavridis and Will [22] showed that modulations induced by spin precession can break these degeneracies. For the “optimal binary”, for example, the average bound including spin precession is \( \lambda_g \sim 5 \times 10^{16} \) km, basically the same as in the case of nonspinning binaries. Yagi and Tanaka [23] included both spin precession and eccentricity. Performing simulations for a \( 10^7 + 10^9 \) \( M_\odot \) binary at 3 Gpc, they found an average bound \( \lambda_g \sim 3.1 \times 10^{16} \) km. This is again consistent with [14] to within a factor 2.

More recently, Keppel and Ajith [24] revisited this problem using phenomenological waveforms for nonspinning BH binaries that include the merger/ringdown phase. Their analysis of LISA bounds is similar to the original work by Will [10] in that it ignores the motion of the detector, using an effective non-sky-averaged noise PSD. This simplification should not affect their main conclusions: (i) for the “optimal binary”, the bound on \( \lambda_g \) improves by about one order of magnitude, up to \( \sim 4 \times 10^{17} \) km; (ii) a comparable bound is obtained also for binaries of larger mass, up to \( M \sim 10^9 M_\odot \).

At \( D_L = 3 \) Gpc, the best bound using the full merger is \( \lambda_g = 5.9 \times 10^{17} \) km for \( M = 4.8 \times 10^7 M_\odot \), to be compared with a best bound \( \lambda_g = 6.3 \times 10^{16} \) km for \( M = 1.9 \times 10^6 M_\odot \) if we consider only inspiral waves. A similar order-of-magnitude improvement is expected for Earth-based detectors.

Del Pozzo et al. [25] recently revisited graviton mass bounds using Bayesian inference (see also [26, 27] for similar work including other possible alternatives to general relativity). They focused on ground-based observations of sources within 150 Mpc and found results consistent with Will’s original analysis [10]; for example, their Fig. 6 shows that binary observations with Advanced LIGO would yield typical bounds \( \sim 2 \times 10^{12} \) km.

In this study we consider quasicircular, nonspinning, restricted post-Newtonian inspiral waveforms. We consider an observation time of two years. We take into account weak lensing errors on the redshift following [29], and we compute the individual graviton mass bounds by generalizing the Fisher matrix formalism described in [31], which takes into account correlations with the other waveform parameters. The results of Refs. [22, 24] suggest that our bounds will be (i) very close to the bounds we would obtain for spinning, precessing inspirals, and (ii) about one order of magnitude worse than the bounds achievable if we used “full” merger waveforms.

II. BOUNDS ON \( \lambda_g \) FROM GRAVITATIONAL-WAVE OBSERVATIONS OF MASSIVE BLACK HOLE POPULATIONS

All studies of massive graviton bounds so far analyzed isolated systems, such as Will’s “optimal binary” of mass \( M = 2 \times 10^6 M_\odot \) at \( z \sim 0.5 \). Unfortunately, hierarchical models of massive BH formation and evolution predict that typical systems observable by space-based interferometers would have masses smaller than this, and be located at redshift \( z > 4 \) or higher (see e.g. [4, 5] and Fig. 2 of [31]). The main uncertainties in these models concern the seeding mechanism and the role of accretion in BH growth [31, 32]. In this work we ignore spins in the gravitational waveforms. Accretion mostly influences the spin magnitude [33], so we will focus on the role of seeding. Following work by the LISA Parameter Estimation Taskforce [34], we will consider two “extreme” scenarios: small seeds, efficient accretion (SE) and large seeds, efficient accretion (LE). These models being extreme, we expect that our results should bracket the constraints that would be obtained using other population models.

Here we consider the “Classic LISA” design along with two different designs for the proposed ESA-led space-based detector, that we will call by the working name of “New LISA”. “Classic LISA” consists of three spacecraft forming an equilateral triangle with laser power \( P = 2 \) W, telescope diameter \( d = 0.4 \) m and armlength...
$L = 5 \times 10^9$ m, trailing $20^\circ$ behind the Earth at an inclination of $60^\circ$ with respect to the ecliptic. The authors are members of a Science Performance Task Force that is considering several different LISA-like configurations with different characteristics and sensitivities. The configurations that we call New LISA C2 (C5, respectively) consist of three spacecraft forming an equilateral triangle with armlength $L = 10^9$ m ($L = 2 \times 10^9$ m), laser power $P = 2$ W and telescope diameter $d = 0.4$ m ($d = 0.28$ m). “New LISA” should be deployed $10^9$ degrees behind the Earth, gradually drifting to $\sim 25^\circ$ behind the Earth in 5 years.

The non sky-averaged noise power spectral densities for all three configurations are shown in Fig. 1; they are related to the sky-averaged power spectral density by $S^{NSA}(f) = \frac{1}{\sigma} S^{SA}(f)$ (see [10] for a discussion of sky averaging). These curves include galactic confusion noise, estimated using methods similar to [34] (which in turn was based on [35]). In our study we consider the noise power spectral density to be infinite below a cutoff frequency $f_{\text{cutoff}} = 10^{-5}$ Hz. As shown in [10], bounds on $\lambda_g$ drop significantly at masses $\gtrsim 2 \times 10^9 M_\odot$ if the noise cannot be trusted below $10^{-4}$ Hz. This assumption has a strong effect on our results, because most binaries in our models have mass lower than this.

For each model we consider 1000 realizations of the Universe. Each of these realizations typically produces $\sim 30 - 50$ events observable with signal-to-noise ratio (SNR) larger than 8. The distribution of bounds on $\lambda_g$ resulting from individual observations is shown in Fig. 2.

The top half of Table II shows that the mean bound over individual observations is $\sim 3 \times 10^{15}$ km for New LISA C2, and only slightly better for the other designs. This conclusion is quite robust, in the sense that numbers vary only mildly for different seeding mechanisms and different detector characteristics.

In most alternative theories, deviations from general relativity can be parametrized by one or more global parameters (such as $\lambda_g$) which are the same for every system. It is natural to expect that one can obtain better constraints on these parameters, as well as other universal constants, by combining multiple observations (see e.g. [25, 28, 29, 36–39]). Assuming that estimates for individual sources are independent and Gaussian posteriors for each source, consistent with the Fisher matrix approximation, the width $\sigma^2$ of the combined posterior on $1/\lambda_g$ is given by $\sigma^2 = \sum_i \sigma_i^{-2}$, where $\sigma_i^2$ is the width of the posterior for the $i$th source. The bound on $\lambda_g$ can thus be obtained by adding the individual bounds in quadrature. The results are shown in Fig. 3 and in the bottom half of Table II. The combined bound obtained from the whole BH population is about one order of magnitude better.
than the average bound obtained from typical observations. A rough estimate would suggest that $N$ identical sources should provide a bound $\sim \sqrt{N}$ times better than the bound from a single source. Our combined bound is typically about 3 times better than the bound from the best event, but the median bound is typically an order of magnitude worse than the best, and hence $\sim 30$ times worse than the combined bound. A typical realisation has $\sim 50$ events, so our analysis shows that we can beat the $\sqrt{N}$ extrapolation from the median bound by a considerable margin. If 5/6 links (two Michelsons) are available instead of 4 links (one Michelson), the bound typically improves by a factor $\sim \sqrt{2}$.

III. CONCLUSIONS AND OUTLOOK.

We assessed the capability of future space-based interferometers, such as “Classic LISA” and the proposed ESA-led “New LISA”, to constrain the mass of the graviton by combining observations of a population of massive BH binaries. We found that: (1) by using a population of merging BH binaries we can obtain a bound on $\lambda_g$ that is $\sim 10$ times better than the mean bound on individual observations; (2) quite independently of the detector’s design and of details of the massive BH formation models, the combined bound from inspiral observations will be $\lambda_g \approx 3 \times 10^{16}$ km. This figure is likely to underestimate the bound achievable in practice by about one order of magnitude, as we have ignored the merger and ringdown portion of the waveform [24], but further work is required to confirm this expectation.

In conclusion, space-based observations of a population of merging BHs should set bounds in the range $\lambda_g \in [2 \times 10^{16} \text{ to } 10^{18} \text{ km}]$ on the graviton Compton wavelength, depending on details of the detector and on the specific waveform model used to set the bounds. This is comparable to the (static and model-dependent) bounds from cosmological-scale observations quoted in Table I but it is very different in nature, because gravitational radiation tests the dynamical regime of Einstein’s general relativity.

Acknowledgments. E.B. is supported by NSF Grant No. PHY-0900735 and by NSF CAREER Grant No. PHY-1055103. J.G.’s work is supported by the Royal Society. The authors wish to thank Martin Elvis and the “New LISA” Science Performance Task Force for discussions, Marta Volonteri for sharing her BH formation models, and the Aspen Centre for Physics (where this work was started) for providing a very stimulating environment.