New horizons for fundamental physics with LISA

K. G. Arun · Enis Belgacem · Robert Benkel · Laura Bernard · Emanuele Berti · Gianfranco Bertone et al. [full author details at the end of the article]

Received: 17 February 2022 / Accepted: 26 April 2022 / Published online: 30 June 2022
© The Author(s) 2022

Abstract
The Laser Interferometer Space Antenna (LISA) has the potential to reveal wonders about the fundamental theory of nature at play in the extreme gravity regime, where the gravitational interaction is both strong and dynamical. In this white paper, the Fundamental Physics Working Group of the LISA Consortium summarizes the current topics in fundamental physics where LISA observations of gravitational waves can be expected to provide key input. We provide the briefest of reviews to then delineate avenues for future research directions and to discuss connections between this working group, other working groups and the consortium work package teams. These connections must be developed for LISA to live up to its science potential in these areas.

Keywords LISA · Gravitational waves · Fundamental physics · Tests of general relativity

Contents

1 Introduction .................................................................................................................. 2
  1.1 Abbreviations ......................................................................................................... 4
2 Tests of general relativity ............................................................................................ 6
  2.1 Tests of Gravity and its fundamental principles ..................................................... 6
  2.2 Testing GR with compact objects ......................................................................... 7
  2.3 Burning questions and needed developments ..................................................... 14
3 Tests of the nature of BHs ......................................................................................... 15
  3.1 The Kerr hypothesis ......................................................................................... 15
  3.2 Deviations from the Kerr hypothesis .................................................................. 16
  3.3 Observables and tests ....................................................................................... 20
  3.4 Burning questions and needed developments ................................................... 26
4 Dark matter and primordial BHs ............................................................................. 27

Extended author information available on the last page of the article
1 Introduction

Gravity is at the forefront of many of the deepest questions in fundamental physics. These questions include the classical dynamics and quantum nature of black holes (BHs), the matter and antimatter asymmetry of the observable universe, the processes at play during the expansion of the universe and during cosmological structure formation, and, of course, the intrinsic nature of dark matter and dark energy, and perhaps of spacetime itself. These questions are cross-generational and their answers are likely to require cross-disciplinary explorations, instead of being the purview of a particular sub-discipline.

The recent observations of gravitational waves (GWs) are beginning to help us dig deeper into these questions, opening new research paths that enable the fruitful interaction of fundamental theory and observation. GWs have the potential to examine largely unexplored regions of the universe that are otherwise electromagnetically obscure. Examples of these regions include the vicinity of BH horizons, early phases in the formation of large-scale structure, and the hot big bang. Moreover, GWs can complement electromagnetic (EM) observations in astronomy and cosmology, thus enabling “multi-messenger” astrophysics and creating a path toward a better understanding of our universe.

The Laser Interferometer Space Antenna (LISA) has the potential to contribute enormously in this quest, as this instrument is uniquely positioned to observe, for the first time, long-wavelength GWs (Amaro-Seoane et al. 2017), and therefore, new
sources of GW radiation. Why is this? Because LISA can provide new information about gravitational anomalies, perhaps related to the quantum nature of gravity, about quantum-inspired effects in BH physics, and even about beyond the standard model, particle physics. Examples of the information that could be gained abound, but one concrete example is the following. LISA has the potential to constrain (or detect) the activation of scalar or vector degrees of freedom around black holes, which arise in certain modified gravity theories inspired by quantum gravity. This is because these fields typically carry energy-momentum away from BH binaries, forcing them to spiral faster into each other than predicted in GR. This faster rate of inspiral then modifies the time evolution of the GW frequency, and therefore, its phase, which LISA is particularly sensitive to.

GW observations with LISA will complement the information we have gained (and will continue to gain) with ground-based GW interferometers, such as the Laser Interferometer Gravitational-wave Observatory (LIGO), Virgo and KAGRA. This complementarity arises because ground-based instruments operate at higher frequencies than LISA, and therefore, they can hear GWs from a completely different type of sources. This, in turn, implies LISA is uniquely positioned to detect supermassive BH (SMBH) mergers, extreme mass-ratio inspirals (EMRIs), galactic binaries, and SGWB, none of which ground-based detectors are sensitive to. The GWs emitted by some of these sources (such as those emitted by SMBH mergers) will be extremely loud, generating signal-to-noise ratios (SNRs) in the thousands, which will allow us to prove fundamental physics deeply. Other GWs will be less loud, but they will be extremely complex, as is the case for waves generated by EMRIs; these waves will contain intricate amplitude and phase modulations that will encode the BH geometry in which the small compact object zooms and whirls. The observation of GWs with LISA will also complement observations with pulsar timing arrays and EM observations of the B-mode polarization in the cosmic microwave background (CMB), which can inform us about fundamental physics at even lower frequencies.

The goal of this white paper is threefold. First, we aim to identify those topics in fundamental physics beyond the current standard models of particle physics, gravity and cosmology that are particularly relevant for the scientific community, in order to delineate and sharpen LISA's potential in each of these areas. The organization of this article reflects this identification of topics, with each section covering one area. The range of areas is illustrated in Fig. 1. Second, within each section we summarize the state of the art of each of these areas, both from a theoretical and an observational point of view. In order to keep this article both manageable and useful, rather than presenting an extensive detailed review of each topic, we have opted for sufficiently concise "reviews", referring to excellent recent articles in Living Reviews in Relativity and elsewhere for more details (see e.g., Barack et al. 2019). Put differently, the philosophy behind this white paper is to strive towards completion in terms of what are relevant branches of fundamental physics for LISA that we think ought to be explored further, rather than providing extensive background and technical details on each of the topics. Third, each section then assesses what must be done in order for LISA to live up to its scientific potential in these areas of fundamental physics. This is primarily where the living part of this article enters, which should be viewed in a LISA-specific context. Finally, we bring these different strands together and identify possible synergies in a roadmap section at the end.
In closing, let us note that this article is part of a series of LISA working group (WG) articles. Some of the topics discussed here are of interest to several WGs organized within the LISA Consortium. Strong synergies exist between the Fundamental Physics WG, the Cosmology, the Astrophysics, and the Waveform modelling WGs, each of which approaches the topics discussed in this article from different, complementary angles. One of the intended goals of this article is to promote these synergies and connect the WGs, specially the fundamental physics one to the LISA work package groups, so that the ideas presented here can be implemented and deployed in future LISA data analysis.

1.1 Abbreviations

We here list a set of abbreviations commonly used in this review article:

- BH = Black hole
- BBH = Binary black hole
- BMS = Bondi–Van der Burg–Metzner–Sachs
- CMB = Cosmic microwave background
- DM = Dark matter
- dCS = Dynamical Chern–Simons

Fig. 1 Schematic diagrams of topics of relevance to the Fundamental Physics WG of the LISA Consortium. Some of these topics overlap naturally with other WGs, such as the Waveform modelling WG, the Astrophysics WG and the Cosmology WG.
• ECO = Exotic compact object
• EdGB = Einstein-dilaton-Gauss–Bonnet
• EdM = Einstein-dilaton-Maxwell.
• EEP = Einstein equivalence principle
• EM = Electromagnetic
• EMRI = Extreme mass-ratio inspiral
• EsGB = Einstein-scalar-Gauss–Bonnet gravity
• FLRW = Friedmann–Lemaître–Robertson–Walker
• GR = General relativity
• GW = Gravitational waves
• IMBH = Intermediate-mass black hole
• IMR = Inspiral-merger-ringdown
• LIGO = Laser Interferometer Gravitational-wave Observatory
• LISA = Laser Interferometer Space Antenna
• LLI = Local Lorentz invariance
• LPI = Local position invariance
• MCMC = Markov Chain Monte Carlo
• MBH = Massive black hole
• NS = Neutron star
• QNM = Quasinormal model
• PBH = Primordial black hole
• PN = Post-Newtonian
• ppE = Parameterized post-Einsteinian
• ppN = Parametrized post-Newtonian
• SCO = Stellar compact object
• SEP = Strong equivalence principle
• SGWB = Stochastic gravitational wave background
• SMBH = Supermassive black hole
• SNR = Signal-to-noise ratio
• TeVeS = Tensor-vector-scalar gravity
• WD = White dwarf
• WG = Working group
• WEP = Weak equivalence principle

Henceforth, we use geometric units in which $G = 1 = c$. 
2 Tests of general relativity

2.1 Tests of Gravity and its fundamental principles

GWs and some of their major sources—BHs and relativistic stars—are among the most groundbreaking predictions of GR. It is hence rather intuitive that GW observations should offer unique insights into the workings of gravity and the fundamental principles that underpin GR.

2.1.1 The equivalence principle

The various versions of the equivalence principle are commonly used as a theoretical foundation of GR and as guiding principles for tests of gravity (Will 1993).

The Weak Equivalence Principle (WEP) postulates that the trajectory of a freely falling test body is independent of its structure and composition, providing the fact that there are no external forces acting on this body such as electromagnetism. The Einstein Equivalence Principle (EEP) goes one step further and combines the WEP, Local Lorentz Invariance (LLI), and Local Position Invariance (LPI). The LLI is connected with the assumption that the outcome of any local non-gravitational test experiment is independent of the velocity of the freely-falling frame of reference where this experiment is performed. The LPI requires that the outcome of such an experiment is independent of where and when it is performed. The Strong Equivalence Principle (SEP) is equivalent to the EEP with the WEP extended to self-gravitating bodies and the LLI and LPI to any experiment.

The EEP dictates the universality of couplings in the standard model because of LLI and LPI. The SEP takes this even further and implies that the gravitational coupling is fixed as well. Testing the SEP is often considered synonymous to testing GR itself (Will 1993). If the SEP is fulfilled, the only field that is a mediator of the gravitational interaction should be the spacetime metric. Conversely, the existence of new fields mediating gravity would generically lead to violations of the SEP (which might appear only in specific systems, depending on how elusive the fields are).

2.1.2 Lovelock’s theorem and GR uniqueness

The discussion above already suggests strongly that testing GR (classically) amounts to looking for new fields. An independent way to reach the same conclusion is to start from Lovelock’s theorem (Lovelock 1972): Einstein’s field equations are unique, assuming that we are working in four dimensions, diffeomorphism invariance is respected, the metric is the only field mediating gravity, and the equations are second-order differential equations. Violating any of these assumption can circumvent Lovelock’s theorem and lead to distinct alternative theories of gravity. However, the vast majority of them1 will share one property: they will contain one or more additional fields (see e.g., Sotiriou 2015b). Demonstrating this

---

1 See Barausse et al. (2008); Flanagan (2004); Pani and Sotiriou (2012); Pani et al. (2013b) for discussions regarding attempts to circumvent Lovelock’s theorem without adding new fields.
mathematically might require compactifying spacetime down to 4-dimensions, introducing additional fields to restore diffeomorphism invariance, and performing field redefinitions.

The realization that testing the SEP and looking for deviations from GR largely amounts to looking for new fields allows one to understand intuitively how compact objects and GWs can be used to probe such deviations. When the new fields have a nontrivial configuration around a BH or compact star, the latter can be thought of as carrying a ‘charge’ (this might not be a conserved charge associated to a gauge symmetry). Field theory intuition tells us that accelerating charges radiate. Hence, binaries beyond GR will exhibit additional GW polarizations. Emission in extra polarizations affects the rate of energy loss and, in turn, the orbital dynamics. So, the pattern of emission of conventional polarizations will also be affected. In Sect. 2.2.1 we discuss how the structure of BHs—primary sources for LISA—can be affected by new fields, whereas in the rest of Sect. 2.2 we discuss how deviations from GR can be imprinted in different parts of the waveform or affect GW propagation.

2.2 Testing GR with compact objects

2.2.1 BHs beyond GR and theories that predict deviations

There exist no-hair theorems for a wide range of alternative theories of gravity (Bekenstein 1997; Doneva and Yazadjiev 2020; Hawking 1972; Herdeiro and Radu 2015; Hui and Nicolis 2013; Sotiriou 2015a; Sotiriou and Faraoni 2012) stating that the quiescent, isolated BHs are indistinguishable from their GR counterpart. If the additional fields can be excited, quasinormal mode (QNM) ringing can still be distinct from GR (Barausse and Sotiriou 2008; Molina et al. 2010; Tattersall and Ferreira 2018) and hence offer a possibility of test theories that are covered by no-hair theorems. Models that manage to circumvent no-hair theorems are expected to lead to more prominent deviations from GR in GW signals, as they can posses additional charges and exhibit additional interactions that can affect all parts of the waveform.

Scalar fields coupled to the Gauss–Bonnet invariant or the Pontryagin density (dynamical Chern–Simons gravity) are known to lead to BH hair (Benkel et al. 2017, 2016; Campbell et al. 1992; Delgado et al. 2020b; Delsate et al. 2018; Kanti et al. 1996; Sotiriou and Zhou 2014a, b; Stein 2014; Yagi et al. 2012c; Yunes and Pretorius 2009a; Yunes and Stein 2011) which can be observed with LISA (Maselli et al. 2020a; Yagi and Stein 2016; Yagi et al. 2012b). Such couplings are expected in low-energy limits of quantum gravity (see e.g., Ashtekar et al. 1989; Jackiw and Pi 2003; Metsaev and Tseytlin 1987) and are part of the Horndeski class of scalar-tensor theories (Kobayashi 2019). A coupling to the Pontryagin density is the leading-order parity violating term in gravity in the presence of a (pseudo) scalar (Alexander and Yunes 2009; Jackiw and Pi 2003), while the linear coupling with the Gauss–Bonnet invariant is the only term inducing hair for shift-symmetric (aka massless) scalars (Delgado et al. 2020b; Saravani and Sotiriou 2019; Sotiriou and Zhou 2014a) and the leading correction for GW emission (Witek et al. 2019). A nonlinear coupling with the Gauss–Bonnet invariant has recently been shown to give rise to BH spontaneous...
scalarization triggered by curvature (Blázquez-Salcedo et al. 2018; Collodel et al. 2020; Cunha et al. 2019; Doneva and Yazadjiev 2018; Macedo et al. 2019; Silva et al. 2018, 2019) or spin (Berti et al. 2021; Dima et al. 2020; Doneva et al. 2020a, b; Herdeiro et al. 2021; Hod 2020). The full set of Horndeski theories that can give rise to BHs scalarization has been identified in Antoniou et al. (2018).

Attempts have also been made to circumvent no-hair theorems by relaxing their assumptions. For example, superradiance can support long-lived scalar clouds for very light scalars (Arvanitaki and Dubovsky 2011; Brito et al. 2015b) or lead to hairy BHs for complex scalars with a time dependent phase (Collodel et al. 2020; Delgado et al. 2019; Herdeiro et al. 2018a; Herdeiro and Radu 2014b; Herdeiro et al. 2015; Kleihaus et al. 2015); long-lived scalar “wigs” can be formed around a Schwarzschild BH (Barranco et al. 2012, 2014); configurations with time-dependent scalar fields can be supported by some non-trivial cosmological boundary conditions (Babichev and Charmousis 2014; Berti et al. 2013; Clough et al. 2019; Hui et al. 2019; Jacobson 1999); BHs immersed in an inhomogeneous scalar field were considered in Healy et al. (2012); scalarization can be induced by matter surrounding a BH (Cardoso et al. 2013a, b; Herdeiro et al. 2018b).

BHs in Lorentz-violating theories, such as Einstein-aether theory (Jacobson 2007; Jacobson and Mattingly 2001) and HoVision Res.ava gravity (Blas et al. 2010; Horava 2009b; Sotiriou 2011), will generically carry hair, as the field that breaks Lorentz symmetry will have to be nontrivial and may backreact on the geometry (see, however, Adam et al. 2021; Ramos and Barausse 2019). In such theories, new fields can propagate superluminally or even instantaneously (Bhattacharyya et al. 2016a; Blas and Sibiryakov 2011), even when all known constraints are satisfied (Emir Gümrükçüoğlu et al. 2018; Sotiriou 2018). Indeed, BHs in Lorentz-violating theories can have a nested structure of different horizons for different modes (Barausse et al. 2011; Eling and Jacobson 2006) and potentially a universal horizon that traps all signals (Barausse et al. 2011; Bhattacharyya et al. 2016b; Blas and Sibiryakov 2011). GW observations can probe this richer causal structure and yield novel constraint on Lorentz symmetry breaking.

Hairy BHs have been studied in a variety of other scenarios and theories, including generalized Proca theories (Babichev et al. 2017; Heisenberg et al. 2017; Herdeiro et al. 2016; Kase et al. 2018; Minamitsuji 2017; Rahman and Sen 2019; Santos et al. 2020) and massive gravity theories (Berezhiani et al. 2012; Comelli et al. 2012; Rosen 2017).

2.2.2 Tests with GW propagation

The propagation of GWs provides a clean test of their kinematics. In GR, gravitons are massless spin-2 particles, while in alternative gravity theories, they might be massive (Abbott et al. 2016b; Hassan and Rosen 2012; de Rham et al. 2011; Will 1998), or even have a Lorentz-violating structure in the dispersion relation (Blas et al. 2010; Horava 2009b; Jacobson 2007; Jacobson and Mattingly 2001; Kostelecký and Mewes 2016; Mirshekari et al. 2012; Shao 2020; Sotiriou 2011). Modifications to the dispersion relation could lead to frequency-dependent, polarization-dependent, direction-dependent propagating velocities of GWs. Eventually, while these
components travel at different velocities, the gravitational waveforms received on the Earth are distorted, with respect to their original chirping structure at generation (Kostelecký and Mewes 2016; Will 1998). Therefore, a matched-filter analysis allowing the possibility of the modified GW propagation in the waveform will reveal the nature of graviton kinematics (see e.g., Abbott et al. 2016b) and provide stringent constraints on the mass of the graviton and on Lorentz symmetry violations. The binary neutron star (NS) merger GW170817 has already provided a strong, double-sided bound on the speed of GWs to a part in $10^{15}$ (Abbott et al. 2017e). However, Lorentz-violating theories have multidimensional parameter spaces and generically exhibit additional polarizations (Sotiriou 2018). The speed of these other polarizations remains virtually unconstrained (Emir Gümrükçuoğlu et al. 2018; Oost et al. 2018). A combination of multiple GW events can also be used to simultaneously constrain a set of beyond-GR parameters (Shao 2020).

GWs travel over cosmological distances before they reach the detector. On a Friedmann–Lemaître–Robertson–Walker (FLRW) background, we can describe GWs by $h_{ij}$, the transverse and traceless perturbation of the spatial metric, $g_{ij}(x, t) = a^2(t) \delta_{ij} + h_{ij}(x, t)$. In Fourier space, the most general modification of the GW propagation equation can be written as (assuming spatially flat models) (Barausse et al. 2020; Ezquiaga 2021; Ezquiaga and García-Bellido 2018)

$$\ddot{h}_{ij}(k, t) + [3H(t) + \Gamma(k, t)]h_{ij}(k, t) + \left[c_T^2(t)k^2 + D(k, t)\right]h_{ij}(k, t) = 0,$$

where $H \equiv \dot{a}/a$ is the Hubble rate. The parameters $c_T$, $\Gamma$ and $D$ respectively describe the speed of propagation of the wave, the damping of its amplitude and additional modifications of the dispersion relation. Scale-independent modifications, described by a $k$-independent $\Gamma$, are discussed in Sect. 5. One expects a single observation to exclude $\Gamma \gtrsim L^{-1}$ and $D \gtrsim fL^{-1}$, where $L$ is the distance to the source and $f$ the GW frequency, although the detailed constraints are also controlled by the frequency dependence of these quantities. $c_T$, $\Gamma$ and $D$ will depend on the theory of gravity and on the cosmological background. Hence, one can translate bounds on these parameters into bounds on a given model. For example, the speed of GW bound has severely constrained generalized scalar-tensor theories (Ben Achour et al. 2016; Crisostomi et al. 2016; Deffayet et al. 2011; Gleyzes et al. 2015; Horndeski 1974; Langlois and Noui 2016; Zumalacárregui and García-Bellido 2014) under the assumption that they account for dark energy (Baker et al. 2017; Creminelli and Vernizzi 2017; Ezquiaga and Zumalacárregui 2017; Lombriser and Taylor 2016; McManus et al. 2016; Sakstein and Jain 2017) (see also de Rham and Melville 2018). Relaxing this assumption (Antoniou et al. 2021; Franchini and Sotiriou 2020; Noller et al. 2020) can lift the constraint.

GW propagation tests can also reveal potential couplings and decays of gravitons into other particles (e.g., Creminelli et al. 2020) or oscillations between different states (analogous to neutrino oscillations). The latter are expected in bigravity models (Hassan and Rosen 2012), where a massless and a massive spin-2 fields interact in a specific way to avoid ghost degrees of freedom. The lightest tensor mode is the one that couples to matter and its speed can be constrained as above if there is a prompt EM counterpart. Its amplitude determines the ratio between the luminosity distance

---

**Springer**
of GWs and the one of EM radiation, which oscillates as a function of redshift. Using several astrophysical population models for the population of massive black hole binaries (Barausse 2012; Klein et al. 2016) and performing a $\chi^2$ analysis, it was found that oscillation effects can be observed for masses $m \gtrsim 2 \times 10^{-25}\text{eV}$ (Belgacem et al. 2019c). Thus LISA will provide a $\sim 3$ order of magnitude improvement in mass sensitivity over the current LIGO/Virgo limit, which probes $m \gtrsim 10^{-22}\text{eV}$ (Max et al. 2017), due to the larger oscillation baseline and the lower detection frequency.

### 2.2.3 Tests of GR with MBH coalescence

#### a. Inspiral: LISA will observe very long inspiral phases and it will therefore allow for high precision tests of gravity (Berti et al. 2005). This stage can be modelled using approximate techniques such as the low-velocity, weak-field PN expansion (Poisson and Will 2014), or the parameterized post-Einsteinian approach (ppE) (Yunes and Pretorius 2009b), that is better suited in certain cases for alternative theories of gravity. Even though tests of the GR nature of the waveforms can be performed without linking to specific alternative theories of gravity within the PN or ppE approaches, connecting the predictions of the different generalizations of Einstein’s theory with the possible deviations in the GR expectation values of the PN or ppE parameters is an inseparable part of testing the possible violations of the GR fundamental symmetries (Berti et al. 2018).

Using the inspiral observations, LISA will be able to improve the constraints on different non-GR predictions, such as the scalar dipole radiation, Lorenz symmetry, mass of the graviton, etc., by several orders of magnitude better compared to LIGO/Virgo (Chamberlain and Yunes 2017). Many alternative theories of gravity predict non-zero tidal deformation of BHs (Cardoso et al. 2017) and that can be also tested using the inspiral waveforms.

Even more intriguing is the possibility for multiband GW observations of BBH mergers. The separation between the two members of a BBH determines the frequencies of the emitted GW signal. A binary that will be in the observational band of LISA at large separation can several years later enter the band of ground-based detectors. Using LISA observations and assuming GR, one can then obtain high-accuracy predictions of the time when the binary will become observable by ground-based detectors as well as its position on the sky (Sesana 2016). Any deviation of the former, observed by a ground based detector, would imply a breakdown of GR. For example, joint observations of a LIGO/Virgo and LISA of a GW150914-like event could improve constraints on BBH dipole emission by 6 orders of magnitude (Barausse et al. 2016; Toubiana et al. 2020). The possibility of observing the same system more than once for a prolonged period will offer the opportunity to test completely nonlinear predictions of some generalized scalar-tensor theories, such as dynamical BH scalarization (Khalil et al. 2019).

#### b. Ringdown: In GR, numerical simulations have shown that the end state of a BBH merger is a Kerr BH (e.g., Boyle et al. 2019; Healy and Lousto 2020; Jani et al.
in agreement with the earlier analytical studies proving that the Kerr metric is the unique stationary, asymptotically flat, axially symmetric BH spacetime (Bunting 1983; Carter 1971; Mazur 1982; Robinson 1975) (see Chrusciel et al. 2012; Heusler 1996 for a review). Before reaching the final Kerr state, the BH emits GWs in QNMs, which are labeled by their overtone numbers, \( n \), and angular numbers \((l, m)\) and are determined by these numbers \((l, m, n)\) plus the mass \( M \) and spin parameter \( a \) of the Kerr BH (Berti et al. 2009). Measuring several of these QNMs would allow for “BH spectroscopy,” in which a perturbed Kerr BH is identified by the spectrum of QNMs in its ringdown GWs (Berti et al. 2006; Detweiler 1980a; Dreyer et al. 2004).

Specifically, for a given BBH merger, the number and amplitude of QNMs excited during the ringdown is determined by the individual BHs and their orbital parameters prior to merger (Kamaretsos et al. 2012). If at least two of the QNMs can be measured from the ringdown of a BH, then by knowing which modes were excited, one can verify that the two QNM frequencies and two QNM damping times are both consistent with the same mass \( M \) and spin \( a \) parameters of the remnant Kerr BH, within the errors of the observation (Berti et al. 2006; Dreyer et al. 2004). For a high SNR massive BBH event measured by LISA, this “no hair” (or “final-state”) test can be performed to verify the remnant is consistent with a Kerr BH to high precision (e.g., Gossan et al. 2012).

If the underlying gravity theory differs from GR, the ringdown of a BBH merger will also typically differ from the ringdown in GR. Computing the QNMs in modified theories is typically challenging, as is performing NR simulations of BBH mergers in modified gravity theories (both of which have only been performed in a handful of cases, e.g., Blázquez-Salcedo et al. 2016; Cardoso and Gualtieri 2009; Okounkova 2020; Okounkova et al. 2020). This makes it difficult to look for deviations in a particular modified theory. Parametrized tests are instead used to look for deviations. In Tattersall et al. (2018), deviations from linear perturbations about a Schwarzschild BH in GR were parametrized at the level of a diffeomorphism-invariant action that encompassed a large range of possible theories that lead to second-order equations of motion. A more phenomenological approach to the test for deviations from GR during ringdown is to perform a parametrized test in which the QNM frequencies are given by the Kerr values plus small deviations, in a “post-Kerr” expansion (Glampedakis et al. 2017). There are also procedures to try to combine the deviation parameters from multiple BBH events measured by LISA (Maselli et al. 2020b). With more detailed predictions from specific theories, parametrized constraints could be converted into constraints on a particular theory.

**c. Merger:** The merger phase of a BBH inspiral is arguably when the nonlinear effects of gravity truly manifest themselves. This makes it a very challenging regime to model, requiring full nonlinear evolutions of the field equations. A model-independent self-consistency test, such as the IMR consistency test (Ghosh et al. 2016a, 2017) could be used in order to avoid having to model the merger in alternative theories of gravity. However, having theory-specific waveforms that include the merger is essential. It can provide more stringent constraints and it is necessary for interpreting them physically. It can also be used to quantitatively explore deviations from GR that do not affect other parts of the waveform—e.g. new
fields that are highly excited by nonlinear effects and decay rapidly—and guide and calibrate parametrizations.

Numerical simulations beyond GR have only been performed in a handful of cases, most notably for scalar nonminimally coupled to curvature invariants (Benkel et al. 2016, 2017; Cayuso and Lehner 2020; East and Ripley 2021; Okounkova et al. 2017, 2019b, 2020; Silva et al. 2021b; Witek et al. 2019) and for Einstein–Maxwell–dilaton theory (Hirschmann et al. 2018). Establishing whether the initial value problem is well-posed in alternative theories is particularly challenging (Cayuso et al. 2017; Kovács and Reall 2020; Papallo and Reall 2017; Ripley and Pretorius 2019b; Sarbach et al. 2019). So far, known simulations have circumvented this problem by adopted Effective Field Theory inspired treatments: either working perturbatively in the new coupling constants (Benkel et al. 2016, 2017; Okounkova et al. 2017, 2019b, 2020; Witek et al. 2019) or adopting a scheme similar to the Israel-Stewart treatment of viscous relativistic hydrodynamics, in which additional fields are introduced to render the system hyperbolic and then exponentially ‘damped’ so that the evolution equation match the initial system asymptotically (Cayuso et al. 2017). Nonlinear evolution beyond GR is one of the major challenges in GW modelling.

2.2.4 Test of GR with EMRIs (non-null tests)

GW observations of EMRIs offer the opportunity to probe gravity in a mass range which is unique to LISA. In such systems a stellar mass object orbits around a more massive component, with typical mass ratios of the order of \( q \sim 10^{-5} - 10^{-7} \), leading to a large number of accumulated cycles before the merger, proportional to \( \sim 1/q \). The signal produced by the slow inspiral provides a detailed map of the spacetime that will be able to pinpoint deviations from GR predictions (if any) and requires detailed knowledge of the emitted waveform to avoid spurious systematics (Barack and Pound 2019; Pound 2015a).

Despite progress within GR (Pound et al. 2020), tests of gravity with EMRIs have been so far limited to systematic calculations in modified theories, in which the different couplings to the gravity sector require, in general, a theory-by-theory analysis. These additional degrees of freedom, introduced as scalar, vector or tensor modes, activate extra emission channels that modify the binary phase evolution. These changes are expected to leave a detectable footprint in the emitted signal and be augmented by the large number of cycles followed by EMRIs. Calculations for non GR theories with non minimally coupled scalar fields have mainly investigated the changes in the emitted flux within the adiabatic approximation, for some orbital configurations (including generic orbits) around spinning BHs (Blázquez-Salcedo et al. 2016; Canizares et al. 2012; Cardoso et al. 2011; Fujita and Cardoso 2017; Pani et al. 2011a; Yunes et al. 2012). These works have shown how, depending on the magnitude of the couplings, in some cases the accumulated GW phase can be large to produce hundreds of cycles of difference in the binary evolution compared to GR. The projected constraints inferred by LISA on the parameters of non-GR theories using approximate waveforms (Canizares et al. 2012; Yunes et al. 2012), and modelling beyond the adiabatic approximation, i.e., taking into account self-force calculations (Zimmerman 2015), requires further work.
Drastic simplifications in the EMRI modelling beyond GR have been recently proven to hold for a vast class of theories, for which no-hair theorems or separations of scales provide a decoupling of the metric and scalar perturbations (Maselli et al. 2020a). This result allows one to describe the background spacetime as in GR, rendering all the modifications induced by the modified theory to be universally captured by the scalar field’s charge only. It has already been used to show that the latter can be measured with unprecedented precision (Maselli et al. 2022). In this framework waveform modelling beyond GR can take advantage of all the efforts devoted so far to study the evolution of scalar charges around Kerr BHs (Barack and Burko 2000; Castillo et al. 2018; Detweiler et al. 2003; Diaz-Rivera et al. 2004; Gralla et al. 2015; Nasipak et al. 2019; Warburton 2015; Warburton and Barack 2010, 2011).

2.2.5 Gravitational memory and BMS symmetry

GR predicts that the passage of a GW causes a permanent displacement in the relative position of two inertial GW detectors. This phenomenon is known as the displacement memory effect (Christodoulou 1991; Ludvigsen 1989; Payne 1983; Thorne 1992) (see Compère 2019a; Strominger 2018 for an account of historical references). Studies on the detection of the displacement memory effects in actual experiments have been presented in Favata (2009b, 2009a, 2010, 2011). More recently, prospects for the memory effect detection with LISA have been discussed in Islo et al. (2019) using SMBH binaries undergoing coalescence (See also Sect. 3). In addition to the displacement memory effect, other and subdominant effects such as the spin memory and center of mass memory effects can also take place (Flanagan et al. 2019; Nichols 2017, 2018; Pasterski et al. 2016). Generalizations of the memory effect, not necessarily motivated or associated to a symmetry, which should in principle be measurable, have also been discussed in Compère (2019b); Flanagan et al. (2019), and further memory effects with logarithmic branches have been inferred from graviton amplitude calculations (Laddha and Sen 2019; Sahoo and Sen 2019).

Measurement of memory effects would not only act as a test of GR (Flanagan et al. 2019; Hou and Zhu 2021; Nichols 2017, 2018; Tahura et al. 2021a, b), but it could also potentially shed light on puzzling infrared aspects of quantum field theories. Ideally isolated systems in GR can be described by asymptotically flat spacetimes. At null infinity, the symmetry group of asymptotically flat spacetimes in GR is known to be larger than the Poincare group of symmetries. It contains the original infinite-dimensional Bondi–Van der Burg–Metzner–Sachs (BMS) group which is a semi-direct product of the Lorentz group and of an infinite-dimensional group of “angle dependent translations” called supertranslations (Bondi et al. 1962; Sachs 1962). Further extensions of the BMS group have also been proposed (Barnich and Troessaert 2010; Campiglia and Laddha 2015). BMS supertranslations can be related to the gravitational memory effect in that the relative positions of two inertial detectors before and after the passage of a GW differ by a BMS supertranslation. In addition, perturbative quantum gravity admits remarkable infrared identities among its scattering amplitudes, the soft theorems, that have been demonstrated to be the
Ward identities of BMS symmetries (Strominger and Zhiboedov 2016; Strominger 2018). The classical limit of the quantum soft theorems are equivalent after a Fourier transformation to the memory effects (Strominger and Zhiboedov 2016). Measuring memory effects therefore directly probes the infrared structure of quantum gravity amplitudes.

The detection of GW bursts with memory with the LISA instrument could be considered from SMBH binary mergers. Over a SMBH binary lifetime, memory undergoes a negligible growth prior to merger (corresponding to the slow time evolution of the binary’s inspiral), rapid accumulation of power during coalescence, and eventual saturation to a constant value at ringdown. Islo et al. (2019) study a simulation-suite of semi-analytic models for the SMBH binary population with a signature of a memory signal from a SMBH binary approximated in the time domain by a step-function centered at the moment of coalescence and assuming a simple power-law model to emulate any environmental interaction which could influence the coalescence timing (e.g., final parsec problem Begelman et al. 1980). SMBH could stall before reaching a regime where GW radiation can drive the binary to coalesce.

Considering SMBH binaries at $z < 3$ with masses in the range $(10^5 - 10^7)M_\odot$ and with mass ratios in the range 0.25–1, LISA prospects could be SNR $\geq 5$ events occurring 0.3–2.8 times per year in the most optimistic environmental interaction model, and less than once per million years in the most pessimistic. As shown in Klein et al. (2016); Sesana et al. (2007) (and discussed in the Astrophysics WG White Paper by Amaro-Seoane et al. 2022), most massive BBHs coalescences whose inspiral signature yields LISA SNR $\geq 5$ lie beyond $z = 3$ for a 3-year LISA lifetime, meaning that the results of this study may be interpreted as lower limits on the number of LISA memory events.

2.3 Burning questions and needed developments

Let us now summarize some important questions that ought to be further investigated in the context of tests of GR, without being however exhaustive.

- Even in theories where black holes deviate from Kerr, or strong field dynamics deviates from GR, deviation can be very small once consistency and known viability constraints are imposed or once the dependence of any new charge on the mass or the spin are taken into account. Further work is needed to pin down theories and scenarios that could lead to deviations that are observable by LISA.
- LISA inspiral observations will allow to put severe constraints on different non-GR predictions. This, however, will require a more accurate and complete development of GW waveforms including non-GR effects.
- Numerical simulations with a systematic inclusion of non-GR effects for the merger and ringdowns phase have to be performed such as to get more adequate templates to be then used once data will be available.
- Similarly more detailed studies of the EMRIs waveforms including non GR effects have to be performed in a systematic way.
GW propagation tests are sensitive to potential couplings and decays of gravitons into other particles. Due to the larger oscillation baseline and lower frequency range LISA could improve substantially present LIGO/Virgo bounds. Present studies are preliminary and further analysis could be useful.

3 Tests of the nature of BHs

3.1 The Kerr hypothesis

In vacuum GR, the Carter–Robinson (Carter 1971; Robinson 1975) uniqueness theorem, with later refinements (see Chrusciel et al. 2012 for a review), establishes that the Kerr geometry (Kerr 1963) is the unique physically acceptable equilibrium, asymptotically flat BH solution. This led to the more ambitious proposal that, regardless of the initial energy-matter content available in a gravitational collapse scenario, the dynamically formed equilibrium BHs belong to the Kerr family (in the absence of gauge charges) (Ruffini and Wheeler 1971). Accordingly, (near-) equilibrium astrophysical BH candidates are well described by the Kerr metric. This working proposal is the Kerr hypothesis. Testing the Kerr hypothesis is an important cornerstone of strong-gravity research in which GW science, and, in particular, LISA, are expected to give key contributions.

Deviations from the Kerr hypothesis, that we shall refer to as non-Kerrness, require either modified gravity (discussed in Sect. 2 and also below with a different twist) or non-vacuum GR (discussed in Sect. 3.2.1). Moreover, two approaches are, in principle, possible for studying non-Kerrness. The most explored one is theory dependent: to consider specific choices of matter contents or modified gravity models, compute the BH solutions (which, generically, will be non-Kerr), and finally explore the different phenomenology of a given model. The second one is theory agnostic: to consider parametrized deviations from the Kerr metric, regardless of the model they solve (if any). The latter has been fruitfully employed in studying BH phenomenology in stationary scenarios, e.g., Cardoso et al. (2014b); Johannsen and Psaltis (2011), but its application in the study of dynamical properties is more limited, given the potential lack of an underlying theory.

A substantial departure from the Kerr hypothesis is to admit deviations from, or even the absence of, a classical horizon. This leads to hypothetical exotic compact objects (ECOs) with a compactness comparable to that of (classical) BHs (see Cardoso and Pani 2019 for a review). One motivation for such a dramatic scenario is that, from Penrose’s theorem (Penrose 1965), a classical BH (apparent) horizon implies the existence of spacetime singularities, under reasonable energy (and other) conditions. Thus, the absence of (or deviations from) a classical horizon could circumvent the singularity problem. Another motivation is that quantum corrections may be relevant at the horizon scale, even for small-curvature supermassive objects (Almheiri et al. 2013; Giddings 2006; Lunin and Mathur 2002a, b; Mathur 2005, 2009a, b; Mayerson 2020).

In this context it is also interesting that current LIGO/Virgo GW observations (especially the recent GW190814; Abbott et al. 2020d and GW190521; Abbott et al. 2020d).
2020c, e, respectively in the lower-mass and upper-mass gap forbidden for standard stellar-origin BHs) do not exclude the possibility that ECOs might co-exist along with BHs and NSs. Models of ECOs are discussed in Sect. 3.2.2.

3.2 Deviations from the Kerr hypothesis

3.2.1 BHs in non-vacuum GR

There is a large class of models wherein (covariant) matter-energy is minimally coupled to Einstein’s gravity. These models obey the EEP (cf. Sect. 2.1.1) and fall into the realm of GR.

Including minimally coupled matter fields, with standard kinetic terms and obeying some energy conditions (typically the dominant) can be quite restrictive for the admissible BH solutions. In many models it prevents the existence of non-Kerr BHs. This has been typically established by model-specific no-hair theorems. Historically influential examples are the Bekenstein no-scalar and no-massive-vector hair theorems (Bekenstein 1972) (see Herdeiro and Radu 2015 for a review). Nonetheless, non-standard kinetic terms (e.g. Skyrme hair; Luckock and Moss 1986), negative energies (e.g. interacting real scalar hair; Nucamendi and Salgado 2003), non-linear matter models (e.g. Yang–Mills hair; Bizon 1990) or symmetry non-inheritance between the geometry and matter fields (e.g. synchronised bosonic hair; Herdeiro et al. 2016; Herdeiro and Radu 2014b) allow the existence of new families of BHs with hair, 2 co-existing with the vacuum Kerr solution.

The viability and relevance of any “hairy” BH model should be tested by dynamical considerations: besides demanding well posedness of the matter model, the non-Kerr BHs must have a dynamical formation mechanism and be sufficiently stable to play a role in astrophysical processes. Asymptotically flat BHs with Yang-Mills hair, for instance, are known to be perturbatively unstable (Zhou and Straumann 1991) whereas BHs with Skyrme hair are perturbatively stable (Heusler et al. 1992). However, both these fields are best motivated by nuclear physics, in which case the corresponding BH hair is, likely, astrophysically negligible (except, possibly, for small primordial BHs (PBHs)).

Potentially astrophysically relevant hairy BHs in GR occur in the presence of hypothetical (ultra-light) massive bosonic fields (such as the QCD axion, axion-like particles, dark photons, etc). These ultralight fields could be a significant component of the dark matter (Arvanitaki et al. 2010; Essig et al. 2013; Hui et al. 2017; Marsh 2016) and are predicted in a multitude of scenarios beyond the standard model of particle physics (Essig et al. 2013; Hui et al. 2017; Irastorza and Redondo 2018; Jaeckel and Ringwald 2010), including extra dimensions and string theories. They naturally interact very weekly and in a model-dependent fashion with baryonic matter, but their gravitational interaction is universal. The superradiant instability of Kerr BHs (Brito et al. 2015b), in the presence of (complex) ultralight bosonic fields (East and Pretorius 2017; Herdeiro and Radu 2017), or mergers of self-gravitating

2 BH “hair” refers to new macroscopic degrees of freedom that are not associated to gauge symmetries, and cannot be computed by flux integrals at infinity.
lumps of such ultralight bosons (Sanchis-Gual et al. 2020) (known as bosonic stars—see Sect. 3.2.2) form BHs with synchronised bosonic hair. These BHs are themselves afflicted by superradiant instabilities (Ganchev and Santos 2018; Herdeiro and Radu 2014a), but possibly on long timescales, even cosmologically long (Degollado et al. 2018), which can render them astrophysically relevant. Superradiance triggered by real ultralight bosonic fields, on the other hand, leads to other effects, such as a SGWB, continuous GW sources from isolated BHs, effects in compact binaries, BH mass-spin gaps, etc, all relevant for LISA science (cf. Sect. 4.1).

3.2.2 ECOs: deviations from (or absence of) a classical horizon

ECOs (Giudice et al. 2016) is a generic name for a class of hypothetical dark compact objects without a classical BH horizon that, nonetheless, can mimic the phenomenology of BHs at the classical level. They may be described by their compactness (i.e. the inverse of their—possibly effective—radius in units of the total mass), reflectivity (as opposed to the perfect absorption by a classical BH horizon), and possible extra degrees of freedom related to additional fields (Cardoso and Pani 2019). Their compactness should be comparable to that of BHs; they may be (albeit need not be) ultracompact, i.e., possess bound photon orbits, such as light rings. If they do, they could be further classified according to whether the typical light-crossing time of the object is longer or shorter than the instability time scale of circular null geodesics at the photon sphere, which in turn depends on the object compactness and internal composition (Cardoso and Pani 2019).

Several models of ECOs have been conceived in order to overcome conceptual issues associated to BHs, such as their pathological inner structure and the information loss paradox. Under general conditions, Penrose’s theorem (Penrose 1965) implies that an apparent horizon always hides a curvature singularity wherein Einstein’s theory breaks down. Moreover, in the semi-classical approximation, BHs are thermodynamically unstable and have an entropy which is far in excess of a typical stellar progenitor (Hawking 1976). It has been argued that GWs may provide smoking guns for ECOs (Barausse et al. 2018; Cardoso and Pani 2017, 2019; Cardoso et al. 2016a, b; Giudice et al. 2016).

ECOs fall into two classes: some models are solutions of concrete field theories coupled to gravity, with known dynamical properties; other models are ad hoc proposals (to different extents) put forward to test phenomenological responses without a complete embedding in a concrete model. In the former case their maximum compactness is constrained by the Buchdahl’s theorem, when its hypotheses apply (Cardoso and Pani 2019). In the latter case details about the dynamical formation of the ECOs are unknown. But, as a general principle, it has been argued that quantum effects in the near would-be horizon region could prevent the formation of a horizon in a variety of models and theories (Giddings 2006; Lunin and Mathur 2002a, b; Mathur 2005, 2009a, b; Mayerson 2020; Mazur and Mottola 2004).

Amongst the first class of ECOs one of the most studied examples corresponds to *bosonic stars*. These are self-gravitating solitons, composed of either scalar (Jetzer 1992; Kaup 1968; Ruffini and Bonazzola 1969) or vector (Brito et al. 2016), massive
complex fields, minimally coupled to Einstein’s gravity—see also Herdeiro et al. (2019); Herdeiro and Radu (2020); Herdeiro et al. (2017) for comparisons. Bosonic stars\(^3\) arise in families of models with different classes of self-interactions of the bosonic fields, e.g., Colpi et al. (1986); Delgado et al. (2020a); Grandclement et al. (2014); Guerra et al. (2019); Kleihaus et al. (2005); Minamitsuji (2018); Schunck and Mielke (2003) and different field content, e.g., Alcubierre et al. (2018); they may also be generalized to modified gravity, e.g., Herdeiro and Radu (2018). Bosonic stars circumvent Derrick type no-soliton theorems (Derrick 1964) due to a symmetry non-inheritance between matter and geometry, as the latter is static/stationary and the former includes a harmonic time dependence (but with a time-independent energy-momentum tensor). Some bosonic stars are dynamically robust (Liebling and Palenzuela 2017), in particular perturbatively stable, with a known formation mechanism known as gravitational cooling (Di Giovanni et al. 2018; Seidel and Suen 1994). These can be evolved in binaries yielding gravitational waveforms, e.g., Bezares et al. (2017); Liebling and Palenzuela (2017); Palenzuela et al. (2008, 2017); Sanchis-Gual et al. (2019), that can be used—together with PN approximations (Pacilio et al. 2020)—as a basis to produce waveform approximants for GW searches. Their typical GW frequency depends crucially on the mass of the putative ultralight bosonic field and, depending on the range of the latter, the signal can fall in the frequency band of either LIGO/Virgo or LISA. Recently it was argued that one particular GW event, GW190521 (Abbott et al. 2020c), is well mimicked by a very eccentric collision of spinning Proca stars (Bustillo et al. 2021).

Bosonic stars have a cousin family of solitons in the case of real bosonic fields, called oscillatons (Seidel and Suen 1994). They have a weak time-dependence and slow decay, but can be very long lived, at least for spherical stars (Page 2004). Collisions of oscillatons and the corresponding waveforms have also been obtained, e.g., Clough et al. (2018).

Bosonic stars are the prototypical example of ECOs which are not meant to replace all BHs in the universe, but could in principle “co-exist” with them and be exotic sources for LISA. They could also be especially interesting for the BH seed problem at large redshift. Indeed, just like ordinary NSs, bosonic stars have a maximum mass beyond which they are unstable against gravitational collapse and classically form an ordinary BH. Other models that share the same features are anisotropic stars (Bayin 1982; Bowers and Liang 1974; Letelier 1980) (see Raposo et al. 2019a for a recent fully covariant model). Like bosonic stars, anistropic stars can evade Buchdahl’s theorem due to their large anisotropies in the fluid.

A more ambitious first-principle model of ECO—aiming instead at replacing the classical horizon completely—emerges in the fuzzball proposal (Lunin and Mathur 2002a, b; Mathur 2005, 2009a). In the latter the classical horizon is replaced by smooth horizonless geometries with the same mass, charges, and angular momentum as the corresponding BH (Balasubramanian et al. 2008; Bena and Warner 2013, 2008; Mathur 2005; Myers 1997). These geometries represent some of the microstates in the low-energy (super)gravity description. For special classes of extremal, charged, BHs (Horowitz et al. 1996; Maldacena et al. 1997; Strominger

---

\(^3\) Often called simply boson stars in the scalar case and Proca stars in the vector case.
and Vafa 1996) one can precisely count the microstates that account for the BH entropy, thus providing a regular, horizonless, microscopic description of a classical horizon. In the fuzzball paradigm, all properties of a BH geometry emerge in a coarse-grained description which “averages” over the large number of coherent superposition of microstates, or as a ‘collective behavior’ of fuzzballs (Bena et al. 2019b, a; Bianchi et al. 2018, 2019, 2020b). Crucially, in this model quantum gravity effects are not confined close to the BH singularity, rather the entire interior of the BH is “filled” by fluctuating geometries, regardless of its curvature. It is worth noticing that, while being among the most motivated models for ECOs, fuzzballs anyway require beyond-GR physics confined at the horizon scale.

While microstate geometries emerge from a consistent low-energy truncation of string theory, other more phenomenological models sharing similar phenomenology have been proposed. For example, gravitational vacuum stars, or gravastars, are dark energy stars whose interior spacetime is supported by a negative-pressure fluid which is compensated by a thin shell of an ultrarelativistic positive-pressure fluid. Gravastars are not endowed with an event horizon and have a regular interior. Their model has been conceived in order to overcome the surprisingly huge BH entropy and to provide a model for a thermodynamically stable dark compact object (Mazur and Mottola 2004). The negative pressure might arise as a hydrodynamical description of one-loop QFT effects in curved spacetime, so gravastars do not necessarily require exotic new physics (Mottola and Vaulin 2006). In these models, the Buchdahl limit is evaded both because the internal effective fluid is anisotropic (Cattoen et al. 2005; Raposo et al. 2019a) and because the negative pressure violates some of the energy conditions (Mazur and Mottola 2015). Gravastars can also be obtained as the BH-limit of constant-density stars, past the Buchdahl limit (Posada and Chirenti 2019; Mazur and Mottola 2015). In this regime such configurations were found to be dynamically stable (Posada and Chirenti 2019).

Other models of ECOs include: wormholes (Damour and Solodukhin 2007; Lemos et al. 2003; Morris and Thorne 1988; Visser 1995), collapsed polymers (Brustein and Medved 2017; Brustein et al. 2017), nonlocal stars in the context of infinite derivative gravity (Buoninfante and Mazumdar 2019), dark stars (Barceló et al. 2009), naked singularities and superspinars (Gimon and Horava 2009), 2-2 holes (Holdom and Ren 2017), and quasi-BHs (Lemos and Weinberg 2004; Lemos and Zaslavskii 2008) (see Carballo-Rubio et al. 2018; Cardoso and Pani 2019 for some reviews on ECO models). Finally, it is worth mentioning that some of the existing proposals to solve or circumvent the breakdown of unitarity in BH evaporation involve changes in the BH structure, without doing away with the horizon. Some of the changes could involve “soft” modifications of the near-horizon region, such that the object still looks like a regular GR BH (Giddings 2013, 2017; Giddings et al. 2019), or drastic changes in the form of “hard” structures localized close to the horizon such as firewalls and other compact quantum objects (Almheiri et al. 2013; Giddings et al. 2019; Kaplan and Rajendran 2019). A BH surrounded by some hard structure—of quantum origin such as firewalls, or classical matter piled up close to the horizon—behaves for many purposes as an ECO.

Despite the wealth of models, ECOs are not without challenges. In addition to the lack of plausible concrete formation mechanisms in many models, there are other
generic problems. One issue is that spinning compact objects with an ergoregion but without an event horizon are prone to the ergoregion instability when spinning sufficiently fast (Friedman 1978; Yoshida and Eriguchi 1996). The endpoint of the instability could be a slowly spinning ECO (Brito et al. 2015b; Cardoso et al. 2008) or dissipation within the object could lead to a stable remnant (Maggio et al. 2017, 2019a). Another potential issue is that ultracompact ECOs which are topologically trivial have not one but at least a pair of light rings, one of which is stable, for physically reasonable matter sources (Cunha et al. 2017). Such stable light rings have been argued to source a spacetime instability at nonlinear level (Cardoso et al. 2014a; Keir 2016), whose timescale or endpoint, however, are unclear.

3.3 Observables and tests

3.3.1 Inspiral-based test with SMBH binaries, IMBH binaries, and EMRIs

a. Non-gravitational emission channels by extra fundamental fields: An obvious difference between BHs and certain models of ECOs is that the latter could be charged under some gauge fields, as in the case of current fuzzball microstate solutions, quasi-BHs, and potentially other models that arise in extended theories of gravity. These fields might not be electromagnetic and can therefore avoid current bounds on the charge of astrophysical compact objects coming from charge neutralization and other effects (Barausse et al. 2014; Cardoso et al. 2016c). In addition, their effective coupling might be suppressed, thus evading current constraints from the absence of dipole radiation in BBHs (see Sect. 2). A detailed confrontation of given charged ECO models with current constraints on dipolar radiation remains to be done.

b. Multipolar structure & Kerr bound: The multipole moments of a Kerr BH satisfy an elegant relation (Hansen 1974)\(^4\),

\[ M_{\ell}^\text{BH} + iS_{\ell}^\text{BH} = M^{\ell+1} (i\chi)^\ell, \tag{2} \]

where \( M_\ell \) (\( S_\ell \)) are the Geroch–Hansen mass (current) multipole moments (Geroch 1970; Hansen 1974), \( M = M_0 \) is the mass, \( \chi \equiv J/M^2 \) the dimensionless spin, and \( J = S_1 \) the angular momentum. The multipole moments of the Kerr BH are non-trivial, but Eq. (2) implies that they are completely determined by its mass and spin angular momentum. Thus, there is a multipolar structure, but not multipolar freedom (unlike, say, in stars).

Furthermore, introducing the dimensionless quantities \( \overline{M}_\ell \equiv M_\ell/M^{\ell+1} \) and \( \overline{S}_\ell \equiv S_\ell/M^{\ell+1} \), the only nonvanishing moments of a Kerr BH are

\[ \overline{M}_{2n}^{\text{BH}} = (-1)^n \chi^{2n}, \quad \overline{S}_{2n+1}^{\text{BH}} = (-1)^n \chi^{2n+1} \tag{3} \]

for \( n = 0, 1, 2, \ldots \). The fact that \( M_\ell = 0 \) (\( S_\ell = 0 \)) when \( \ell \) is odd (even) is a

\(^4\) For a generic spacetime the multipole moments of order \( \ell \) are rank-\( \ell \) tensors, \( M_{\ell m} \) and \( S_{\ell m} \), which reduce to scalar quantities, \( M_\ell \) and \( S_\ell \), in the axisymmetric case, see e.g., Bianchi et al. (2020a, 2021) for the general definitions.
consequence of the equatorial symmetry of the Kerr metric, whereas the fact that all multipoles with $\ell \geq 2$ are proportional to (powers of) the spin—as well as their specific spin dependence—is a peculiarity of the Kerr metric.

Non-Kerr compact objects (BHs or ECOs) will have, in general, a different multipolar structure. Differences will be model dependent, but can be considerable in some cases, e.g. for boson stars (Ryan 1997b) and BHs with synchronised scalar hair (Herdeiro and Radu 2014b). For ECOs, the tower of multipole moments is, in general, richer. The deformation of each multipole depends on the specific ECO’s structure, and in general vanishes in the high-compactness limit, approaching the Kerr value (Glampedakis and Pappas 2018; Pani 2015; Raposo and Pani 2020; Raposo et al. 2019b). In particular, a smoking gun of the “non-Kerminess” of an object would be the presence of moments that break the equatorial symmetry (e.g. the current quadrupole $S_2$ or the mass octopole $M_3$), or the axisymmetry (e.g. a generic mass quadrupole tensor $M_{2m}$ with three independent components ($m = 0, 1, 2$), as in the case of multipolar boson stars (Herdeiro et al. 2021) and of fuzzball microstate geometries (Bena and Mayerson 2020, 2021; Bianchi et al. 2020a, 2021).

The multipolar structure of an object leaves a footprint in the GW signal emitted during the coalescence of a binary system, modifying the PN structure of the waveform at different orders. The lowest order contribution, entering at 2PN order is given by the intrinsic (typically spin-induced) quadrupole moment (Barack and Cutler 2007). LISA can be able to detect deviations in the multipole moments from supermassive binaries for comparable and unequal mass systems. So far proposed tests of the Kerr nature have been based on constraints of the spin-induced quadrupole $M_2$ (Barack and Cutler 2007; Krishnendu et al. 2017), spin-induced octopole $S_3$ (Krishnendu and Yelikar 2019), current quadrupole $S_2$ (Fransen and Mayerson 2022).

GW signals emitted by EMRIs will provide accurate measurements of the spin-induced quadrupole at the level of one part in $10^4$ (Babak et al. 2017a; Barack and Cutler 2007), and of the equatorial symmetry breaking current quadrupole at the level of one part in $10^2$ (Fransen and Mayerson 2022). A rather generic attempt to constrain the multipole moments of an axisymmetric and equatorially symmetric central object with EMRIs has been done in Ryan (1995, 1997a) by mapping gauge-invariant geodesic quantities into multipole moments in a small-orbital velocity expansion. Constraining the radiative multipole moments of the entire binary system has been discussed in Kastha et al. (2018, 2019); these constrain deviations from the GR expectation of the binary system without explicitly parametrizing the compact objects’ multipole structure.

c. Tidal heating: The compact objects in the binary produce a tidal field on each other which grows as the bodies approach their final plunge and merger. If the bodies dissipate some amount of radiation, these tides backreact on the orbit, transferring rotational energy from their spin into the orbit. This effect is known as tidal heating. For BHs, energy and angular momentum absorption by the horizon is responsible for tidal heating. This effect is particularly significant for highly spinning BHs and
mostly important in the latest stages of the inspiral. Tidal heating can contribute to thousands of radians of accumulated orbital phase for EMRIs in the LISA band (Bernuzzi et al. 2012b; Datta and Bose 2019; Datta et al. 2020; Harms et al. 2014; Hughes 2001; Maggio et al. 2021; Taracchini et al. 2013). If at least one binary member is an ECO instead of a BH, dissipation is in general smaller than in the BH case or even negligible, therefore significantly reducing the contribution of tidal heating to the GW phase. This would allow to distinguish between BBHs and binary involving other compact objects. For LISA binaries, constraints of the amount of dissipation would be stronger for highly spinning objects and for binaries with large mass ratios (Datta et al. 2020; Maggio et al. 2021; Maselli et al. 2018). For EMRIs in the LISA band, this effect could be used to put a very stringent upper bound on the reflectivity of ECOs, at the level of 0.01% or better (Datta et al. 2020; Maggio et al. 2021). Absence of a horizon can also produce resonances that can be excited during EMRIs (Cardoso et al. 2019c; Macedo et al. 2013b; Maggio et al. 2021; Pani et al. 2010).

d. Tidal deformability: Tidal effects in compact binaries modify the dynamical evolution of the system, accelerating the coalescence. This modifies the orbital phase, and then in turn the GW emission (Hinderer et al. 2018; Poisson and Will 2014). The imprint on the waveform is encoded in a set of quantities which, as a first approximation, can be assumed to be constant during the coalescence (Hinderer et al. 2016; Maselli et al. 2012; Steinhoff et al. 2016): the tidal Love numbers (Binnington and Poisson 2009; Damour and Nagar 2009; Hinderer 2008). These numbers can be thought of as the specific multipole moment induced by an external tidal field, in a way akin to the electric susceptibility in electrodynamics. The main contribution in the GW signal from a binary is given by the quadrupolar term \( k_2 \), connected to the tidal deformability \( \tilde{\lambda} = \frac{2}{3}k_2R^5 \), or in its dimensionless form \( \tilde{\lambda} = \frac{2}{3}k_2C^5 \), where \( R \) and \( C \) are the object radius and compactness, respectively.

The tidal Love numbers depend on the internal composition of the central object. So far, they have been used to constrain the properties of the nuclear equation of state through GW observations of binary NSs (Abbott et al. 2018a). For a fixed equation of state, i.e. composition, the Love numbers depend on the object compactness only.

The tidal Love numbers of a BH in GR are precisely zero. This was shown explicitly for Schwarzschild BHs, for both small (Binnington and Poisson 2009; Damour and Nagar 2009) and large (Gürlebeck 2015) tidal fields. The same result was shown to be valid for slowly rotating BHs up to the second (linear) order in the spin for axisymmetric (generic) tidal fields (Landry and Poisson 2015; Pani et al. 2015a, b; Poisson 2015). Very recently, this result was extended to any tidal Love number of a Kerr BH with arbitrary spin (see Charalambous et al. 2021, 2021; Chia 2021; Hui et al. 2021; Le Tiec and Casals 2021; Le Tiec et al. 2021 for literature on this topic).

For ECOs, the tidal Love numbers are generically different from zero. In analogy with the NS case they depend on the ECO’s structure, and may be used to trace back the underlying properties of each model (Cardoso et al. 2017; Giddings et al. 2019; Herdeiro et al. 2020; Johnson-McDaniel et al. 2020; Maselli et al. 2018, 2019; Pani 2015; Porto 2016; Raposo et al. 2019a; Sennett et al. 2017; Uchikata et al. 2016). For
nonrotating BH mimickers, featuring corrections at the horizon scale and that approach the BH compactness, the Love numbers vanish in the limit $\mathcal{C}_{\text{ECO}} \to \mathcal{C}_{\text{BH}}$, often logarithmically (Cardoso et al. 2017).

LISA will be able to measure the tidal Love numbers of BH mimickers (Maselli et al. 2018), which are otherwise unmeasurable by current and future ground based detectors (Cardoso et al. 2017). In the comparable-mass case, this measurement requires highly-spinning supermassive ECO binaries up to 10 Gpc. LISA may also be able to perform model selection between different families of BH mimickers (Maselli et al. 2019), although this will in general require detection of golden binaries (i.e. binaries with a very large SNR) (Addazi et al. 2019). For a large class of slowly-rotating ECOs with compactness $\mathcal{C}_{\text{ECO}} \lesssim 0.3$, LISA can measure the Love numbers with very good accuracy below 1% (Cardoso et al. 2017). For (scalar) boson stars a recent study proposed a new data analysis strategy to consistently include several corrections (multipolar structure, tidal heating, tidal Love numbers) in the inspiral signal from boson star binaries, improving the accuracy on the measurement of the fundamental parameters of the theory by several orders of magnitude compared to the case in which the effects are considered independently (Pacilio et al. 2020).

Finally, EMRI observations can set even more stringent constraints, since the measurement errors on the Love number scale as $q^{1/2}$, where $q \ll 1$ is the mass ratio of the binary (Pani and Maselli 2019). A simplistic Newtonian estimate (that should be corroborated by a more sophisticated modelling and data analysis) suggests that in this case the tidal Love number of the central object can be constrained at the level of one part in $10^5$ (Pani and Maselli 2019).

e. Integrability/Chaos: One particular probe of extreme gravity that is tailor-made for EMRI signals relates to chaos. For Hamiltonian systems, chaos refers to the non-integrability of the equations of motion, i.e. the non-existence of a smooth analytic function that interpolates between orbits, and has nothing to do with a system being non-deterministic (Levin 2006).

EMRIs in GR can be approximated, to zeroth-order, as geodesics of the Kerr spacetime, and the latter has enough symmetries to guarantee that geodesics are completely integrable and thus non-chaotic. Beyond the zeroth-order approximation, however, other effects, such as the spin of the small compact object, could break the integrability of the systems even within GR (Zelenka et al. 2020). In some modified theories, even the geodesic orbital motion might not be integrable (Cárdenas-Avendaño et al. 2018; Lukes-Gerakopoulos et al. 2010). This is also true for some models of ECOs, such as spinning scalar boson stars and non-Kerr BHs in GR (Cunha et al. 2016). In this sense, the presence of large chaotic features in the GWs emitted by EMRIs could signal a departure from the SEP, a violation of the Kerr hypothesis, or an environmental effect.

Modifications to GR are expected to change the fundamental frequencies of the orbital motion of test particles, which will be then be imprinted on the GWs emitted by the system. A careful study of the evolution of these fundamental frequencies will allow us to understand the importance of chaos to GR and to the observations of GWs from EMRIs (Cárdenas-Avendaño et al. 2018; Destounis et al. 2021; Gair et al. 2008; Lukes-Gerakopoulos et al. 2010).
f. Motion within ECOs: If the ECO interior is made of weakly-interacting matter, a further discriminator of the absence of a horizon (or of a hard surface) would be the motion of test particles within the object and its peculiar GW signal, most notably as in the case of an EMRI moving inside a supermassive ECO. This motion can be driven by a combination of the self-gravity of the central object, accretion, and dynamical friction, etc. The study of geodesic motion inside solitonic boson stars was analyzed in e.g., Kesden et al. (2005). The effects of accretion and drag were included in Barausse et al. (2014), Barausse et al. (2015), Macedo et al. (2013a) and Macedo et al. (2013b). These effects are model independent to a certain extent, since they mostly depend on the density profile. For this reason they also share some similarities with environmental tests of dark matter (see Sect. 4). In general, they could be a smoking-gun signature for the existence of structures in supermassive ultracompact objects.

3.3.2 Ringdown tests

a. QNMs: Similarly to what was discussed in Sect. 2, measuring the ringdown modes in the post-merger signal of a binary coalescence provides a clean and robust way to test GR and the nature of the remnant. If the latter is a Kerr BH in GR, its (infinitely countable) QNM spectrum is entirely determined only in terms of its mass and spin. Thus, detecting several QNMs provides us with multiple independent null-hypothesis tests, and would allow us to perform GW spectroscopy (Berti et al. 2009; Kokkotas and Schmidt 1999). From a more theoretical perspective, the study of the QNMs of compact objects is crucial to assess their linear stability.

The ringdown waveform originates from the perturbed remnant object, and consists of a superposition of (complex) QNMs, whose amplitudes depend on the binary progenitors and on the underlying theory. As previously discussed, the fundamental QNM frequency and damping time have been measured by LIGO/Virgo only for a few events, providing an independent measurement of the mass and spin of the remnant which is in agreement with what inferred from the inspiral-merger phase (Abbott et al. 2019b, 2021b). Among the entire second GW transient catalogue (Abbott et al. 2021b) the first GW event, GW150914, remains among those for which the fundamental QNM of the remnant has been measured with the highest precision (roughly 3% and 7% for the frequency and damping time, respectively). More recently, the importance of overtones has attracted considerable attention, especially because they allow one to start the fitting of the ringdown signal closer to the peak of the signal, improving mass and spin measurements (Isi et al. 2019a). Overtones are particularly useful for tests of GR with equal-mass binaries (for which other angular modes can be suppressed) (Bhagwat et al. 2020; Jiménez Forteza et al. 2020; Ota and Chirenti 2020), but a detailed study for LISA remains to be done. Overall, tests of the no-hair theorem rely also on the ability to estimate the starting time of the ringdown when the signal is dominated by the QNMs of the remnant and on the modelling of higher modes (Baibhav et al. 2018; Bhagwat et al. 2018, 2020; Brito et al. 2018; Giesler et al. 2019; Jiménez Forteza et al. 2020; Ota and Chirenti 2020).

The large SNR expected in LISA for ringdown signals of SMBH coalescences provides a unique opportunity to perform BH spectroscopy (Dreyer et al. 2004) and
tests of the nature of the remnant. For a single “golden merger” up to redshift \( z = 10 \) several QNMs can be measured with unprecedented precision (Berti et al. 2016).

Besides introducing deformations in the QNM spectrum, if the remnant differs from a Kerr BH in GR, some further clear deviations in the prompt ringdown are: (i) possible presence of (or contamination from) other modes, e.g. fluid modes (Pani et al. 2009) in stars or extra degrees of freedom (e.g. scalar QNMs for boson stars; Macedo et al. 2013b), some of which—being at low frequency—could be resonantly excited during the inspiral (Cardoso et al. 2019c; Macedo et al. 2013b; Maggio et al. 2021; Pani et al. 2010); (ii) isospectrality breaking between modes that can be identified as even-parity and odd-parity in the zero-spin limit (Maggio et al. 2020). This produces a characteristic “mode doublet” in the ringdown. A generic framework to study the ringdown of a dark compact object was recently proposed in Maggio et al. (2020) by extending the BH membrane paradigm to ECOs.

b. Echoes: GW echoes (Cardoso et al. 2016a, b) in the post-merger signal of a compact binary coalescence might be a clear signature of near-horizon quantum structures (Abedi et al. 2017; Barceló et al. 2017; Cardoso et al. 2016a, b; Oshita and Afshordi 2019; Wang et al. 2020), ultracompact objects (Bueno et al. 2018; Cardoso et al. 2016a), exotic states of matter in ultracompact stars (Buoninfante and Mazumdar 2019; Ferrari and Kokkotas 2000; Pani and Ferrari 2018), and of modified theories of gravity (Buoninfante et al. 2019; Burgess et al. 2018; Delhom et al. 2019) (see Abedi et al. 2020; Cardoso and Pani 2017, 2017, 2019 for some recent reviews). Detecting echoes would give us the tantalizing prospect of probing the near-horizon structure of dark compact objects with the hope, in particular, to shed light on putative quantum properties of BHs (Ikeda et al. 2021).

If sufficiently compact, horizonless objects support quasi-bound modes trapped within their photon sphere (Cardoso et al. 2016a, b; Kokkotas 1995; Kokkotas and Schmidt 1999). For ultracompact objects the prompt ringdown is identical to that of a BH, since the signal is initially due only to the perturbation of the photon sphere, whereas the BH horizon is reached in infinite coordinate time (Cardoso et al. 2016a, b). At late times, a modulated train of GW echoes appears as a result of multiple reflections of the GWs between the object interior and the photon sphere, leaking out to infinity at each reflection. For the case of intermediate compactness, the prompt ringdown can show some differences with the BH case due to the interference with the first GW echoes (Maggio et al. 2020).

The delay time between echoes is related to the compactness of the object through a logarithmic dependence, which allows for tests of Planckian corrections at the horizon scale (Abedi et al. 2020; Cardoso and Pani 2017; Cardoso et al. 2016b; Oshita et al. 2020). The damping factor of subsequent echoes is related to the reflective properties of the compact object (Cardoso and Pani 2019; Maggio et al. 2019a, 2020; Price and Khanna 2017).

Several waveform templates for echo searches in LIGO/Virgo data have been developed, including: (i) templates in time domain based on standard IMR templates with additional parameters (Abedi et al. 2017; Nakano et al. 2017; Wang and
Afshordi 2018); (ii) superposition of sine-Gaussians with free parameters (Maselli et al. 2017; iii) frequency-domain templates based on the physical ECO parameters (Maggio et al. 2019b; Mark et al. 2017; Testa and Pani 2018). The former were developed for matched-filtered searches. In addition, unmodelled searches based on wavelets adapted from burst searches (Tsang et al. 2018, 2020) and on Fourier windows (Conklin and Holdom 2019; Conklin et al. 2018) have been proposed. For a review on modelling and echo searches, see Abedi et al. (2020).

LIGO/Virgo O1 and O2 events triggered some controversial claims on hints of GW echoes detection. Independent searches found evidence for GW echoes in the O1-O2 events (Abedi and Afshordi 2019; Abedi et al. 2017; Conklin et al. 2018). However, a low statistical significance of such events has been claimed (Nielsen et al. 2019; Westerweck et al. 2018), followed by more recent negative searches (Lo et al. 2019; Tsang et al. 2020; Uchikata et al. 2019). Very recently, using a simplistic template, a dedicated search for echoes has been performed by the LIGO/Virgo Collaboration using the second GW transient catalogue (GWTC-2) (Abbott et al. 2021b), finding no evidence for echoes. This is consistent with independent studies using physically motivated templates, suggesting that that models with almost perfect reflectivity can be excluded/detected with current instruments (5σ confidence level with SNR in the ringdown of ≈10), whereas probing values of the reflectivity smaller than 80% at 3σ or more confidence level requires SNRs of $\mathcal{O}(100)$ in the post-merger phase (Maggio et al. 2019b; Testa and Pani 2018). This makes LISA particularly well suited for echo searches.

c. GWs as messengers from the quantum world: The high sensitivity of LISA or of other advanced detectors will also allow to advance our understanding of quantum gravitational effects. One of the most long-standing ideas in this context is the proposal that the area of BH horizons is quantized in units of the Planck area $A = x\ell_p^2 N$, where $N$ is an integer characterizing the BH quantum state, $x$ is an $\mathcal{O}(1)$ dimensionless coefficient and $\ell_p$ is Planck’s length (Bekenstein 1974; Bekenstein and Mukhanov 1995; Mukhanov 1986). Transitions between states occur, surprisingly, at frequencies where ground- and space-based detectors operate. Thus, quantum BHs may have different tidal heating properties (Agullo et al. 2021; Cardoso et al. 2019a) or different tidal Love numbers (Brustein and Sherf 2022), and the coalescence of two BHs can lead to late-time echoes in the waveforms (Agullo et al. 2021; Cardoso et al. 2019a).

3.4 Burning questions and needed developments

Several issues discussed in this section need further detailed studies, among which we list the following:

- A better understanding of the consequences of superradiant instabilities triggered by e.g. real ultralight bosonic fields on the SGWB, on continuous GW sources, and on binaries, which could be detected by LISA.

---

5 Echo templates available at [http://www.darkgra.org](http://www.darkgra.org), [https://web.uniroma1.it/gmunu/](https://web.uniroma1.it/gmunu/)
A comprehensive study of possible deviations of multipole moments for SMBH binaries is important to have a good tool at disposal for testing the Kerr nature of the BHs.

First-principle models of (possibly stable) ultracompact objects should be developed to provide viable candidates for a BH mimicker.

In general for testing the different effects such as tidal heating, tidal deformability, multipole moments, resonances, and ringdown effects, and echoes there is need to develop quite accurate waveforms, with several parameters describing possible deviations, and to perform accurate statistical analysis (e.g. Bayesian model selection), especially in the case of a putative signature of new physics in the data. Clearly, this also requires to develop very accurate GR waveforms.

4 Dark matter and primordial BHs

The fundamental nature of dark matter (DM) is one of the major unresolved questions in physics and cosmology. Its presence is inferred from various observations, such as galaxy rotation curves, gravitational lensing, galaxy cluster dynamics and CMB data (see Bertone et al. 2005a; Bertone and Hooper 2018; Carr et al. 2016 for reviews). These observations constrain its behavior on larger scales: DM must broadly behave as non relativistic, collisionless matter, with an average density in galactic haloes of the order of \( \sim 0.1 \, M_\odot \, \text{pc}^{-3} \) (1 GeV cm\(^{-3}\)), low velocity dispersions of order \( \sim 100 \, \text{km s}^{-1} \), and it must be interacting weakly, if at all, with baryonic matter and with itself. However, the physical properties of its constituents, in particular their individual masses and spins, remain very poorly constrained.

This section is concerned with models where the DM consists of or is formed from (in the case of PBHs) some type of matter and not where the observed effects arise due to some modification to gravity, such as e.g., Modified Newtonian Dynamics (MOND) (Sanders and McGaugh 2002), which is the topic of Sect. 2. There are two broad categories of dark matter:

1. Particle dark matter, e.g. WIMPs or axions, where the problem of identification lies in determining the mass, spin and fundamental interactions of the particle (or particles).
2. Macroscopic objects, e.g. primordial BHs (PBHs) or exotic compact objects (ECOs), where the distributions of masses and spins, as well as the matter from which they formed, are key aspects to be understood.

Nothing prevents DM from being a combination of several different components drawn from either or both of these categories. Moreover, some models (e.g. axions), are naturally hybrids, where the DM may primarily consist of unbound particles, but also form macroscopic gravitationally bound structures such as boson stars in overdense regions. A further aspect of DM on which observations may shed light is in identifying possible formation mechanisms, e.g. signatures of phase transitions, which often motivate particular DM models.
Several previous reviews have considered the potential for GWs to shed light on the nature of dark matter (Bertone and Tait 2018; Bertone et al. 2020; Giudice 2017). Further, a complementary white paper by the LISA Cosmology WG (Auclair et al. 2022) assesses the prospects of GW observations of DM from a cosmological angle. In this section we focus on the potential for LISA specifically to identify and constrain the particle nature of dark matter and its interactions. We summarise what is known to be possible, highlight promising areas which require further investigation, and detail the work that should be done prior to launch to enable us to correctly interpret the data that we will obtain. Figure 2 summarises the key candidates and regions which may be probed by LISA observations.

Before we proceed we note two caveats. First, effects described in this section may be degenerate with environmental effects described in Sect. 7 and in the waveform modelling white paper. Breaking this degeneracy will be a key factor in our ability to accurately identify DM signatures. Second, we have in most cases been optimistic in assuming that LISA measurements will be optimized for the searches we describe and that models can allow for sufficiently precise predictions to proceed. Turning these caveats around, they could both be regarded as a starting point for action items for future work in order to ensure LISA’s science potential will be fully exploited.

We now present a number of interesting particle candidates, or rather categories of candidates, that have been explored so far, without making the above qualifications necessarily explicit in each section. We start by reviewing particle candidates, separated into low mass (sub eV) and high mass cases. This roughly divides cases where wave-like and particle-like behaviour respectively would dominate on astrophysically significant scales. We then turn to macroscopic candidates such as

![Fig. 2 Overview of key DM candidates (Bertone and Tait 2018; Bertone et al. 2020; Giudice 2017). Currently excluded ranges of particle masses and DM fractions $f_{DM}$ are indicated in white. Parameter regions where LISA observations may provide constraints are grey coloured. We see that the masses of DM particle candidates remain currently very much unconstrained and that DM particles of any allowed mass can make up any fraction ($f_{DM} \leq 1$) of the DM. Macroscopic DM candidates such as PBHs and ECOs are constrained by current observations but may still constitute 100% of the DM in the range $10^{-16} - 10^{-11} M_\odot$. Approximate regions where LISA can contribute to constraints are bracketed above and below the plot with a brief description—see the text for more details.](image)
PBHs, where we separate candidates that would clearly be of primordial origin (sub $M_\odot$) from those that could arise from stellar evolution. The last two sections review the potential for enhanced signatures from additional fundamental interactions.

### 4.1 Low mass particles $m < \text{eV}$

In the past few years, the possibility that dark matter could be composed of ultralight bosons, with masses ranging anywhere between $10^{-22} \text{eV} \lesssim 1 \text{ eV}$, has become a popular hypothesis (Ferreira 2021). A non-exhaustive list of theoretical models predicting the existence of such particles includes the well-known QCD axion proposed to solve the strong CP problem of QCD (Peccei and Quinn 1977; Wilczek 1978), axion-like particles arising in “string axiverse” scenarios (Arvanitaki et al. 2010), dark photons (Goodsell et al. 2009; Nelson and Scholtz 2011) and “fuzzy dark matter” (Hui et al. 2017).

Quite remarkably, GW observatories such as LISA could be used to constrain or find evidence for the existence of these particles. This stems from the fact that rotating BHs can become unstable against the production of light bosonic particles, if such particles exist in nature (Brito et al. 2013, 2020; Cardoso et al. 2018a; Damour et al. 1976; Detweiler 1980b; Dolan 2007, 2018; East 2017; Frolov et al. 2018; Moschidis 2016; Pani et al. 2012a, b; Shlapentokh-Rothman 2014; Witek et al. 2013), through a process known as BH superradiance (Brito et al. 2015b; Press and Teukolsky 1972; Zel’dovich 1971, 1972; Teukolsky and Press 1974). The instability spins the BH down, transferring up to a few percent of the BH’s mass and angular momentum to the boson field, forming a long-lived bosonic “cloud” outside the horizon (Arvanitaki and Dubovsky 2011; Arvanitaki et al. 2010, 2015, 2017; Baryakhtar et al. 2017; Brito et al. 2015a, b; Degollado et al. 2018; East and Pretorius 2017; Ficarra et al. 2019; Herdeiro and Radu 2014b). Superradiance is most effective when the boson’s Compton wavelength is comparable to the BH’s gravitational radius (Dolan 2007; Witek et al. 2013), meaning that observations of astrophysical BHs in the supermassive to the stellar mass range, allow us to probe bosons with masses in the range $10^{-21} - 10^{-10} \text{ eV}$.

For real bosonic fields, the cloud dissipates over long timescales through the emission of nearly-monochromatic GWs with a typical frequency $f \sim 2c^2 m_b/h$, where $m_b$ is the mass of the field (Arvanitaki and Dubovsky 2011; Arvanitaki et al. 2010, 2015, 2017; Brito et al. 2017a; Baryakhtar et al. 2017; Brito et al. 2017b; East 2018; Siemonsen and East 2020; Yoshino and Kodama 2014). These signals could be observable individually or as a very strong SGWB (Arvanitaki et al. 2015, 2017; Baryakhtar et al. 2017; Brito et al. 2017a, b, 2020; D’Antonio et al. 2018; Ghosh et al. 2019; Isi et al. 2019b; Sun et al. 2020; Tsukada et al. 2019, 2021; Zhu et al. 2020), and could thus be used to constrain the existence of light bosons in the absence of a detection. Current and future Earth-based detectors could detect GWs emitted by bosons in the range $m_b \sim 10^{-14} - 10^{-11} \text{ eV}$, while LISA could be sensitive to bosons of mass $m_b \sim 10^{-19} - 10^{-15} \text{ eV}$. It has also been proposed that LISA observations of BBHs with total masses in the range $\sim [100, 6000] M_\odot$ will provide sufficient information to enable targeted searches for these monochromatic GWs.
signals with future ground-based GW detectors for masses in the range $m_b \sim 10^{-14} - 10^{-12}$ eV (Ng et al. 2020).

Furthermore, since BHs affected by the instability would spin down in the process, accurate measurement of the spin of astrophysical BHs can be used to strongly constrain, or find evidence for, ultralight bosons (Arvanitaki and Dubovsky 2011; Arvanitaki et al. 2015, 2017; Baryakhtar et al. 2017; Brito et al. 2013, 2017a; Cardoso et al. 2018a; Ng et al. 2021; Pani et al. 2012a, b). Requiring that the instability acts on timescales shorter than known astrophysical processes, such as accretion and mergers, current measurements of SMBHs spin using continuum fitting or the Iron Kα (see e.g., Jovanovic 2012) can probe bosons in the mass range $m_b \sim 10^{-20} - 10^{-17}$ eV, whereas similar measurements for stellar-mass BHs probe the mass range is $m_b \sim 10^{-12} - 10^{-11}$ eV. LISA is expected to detect binaries and measure the spins of MBHs in the range $\sim 10^2 - 10^7 M_\odot$ and will therefore be able to constrain the existence of light bosons in the intermediate range $m_b \sim 10^{-16} - 10^{-13}$ eV (Brito et al. 2017a, 2020; Cardoso et al. 2018a).

When the BH-boson cloud system is part of a binary, one may also expect new phenomena that could leave imprints in the GW signal emitted by the binary system. For example, the presence of a companion can result in resonant transitions between energy levels of the cloud (Baumann et al. 2019) or complete tidal disruption (Cardoso et al. 2020), which could lead to a significant dephasing of the binary’s GW signal. In particular, for EMRIs, resonant transitions could lead to long-lived floating orbits (Baumann et al. 2020; Ferreira et al. 2017; Zhang and Yang 2019). The presence of a cloud would also be imprinted in the binary’s GW signal through its spin-induced multiple moment(s) and tidal Love number(s) (Baumann et al. 2019, 2020; De Luca and Pani 2021). For eccentric orbits, resonances can occur over a larger range of frequencies, extending the possibility of detection (Berti et al. 2019). In the case of EMRIs, dynamical friction and the gravitational pull of the cloud may also leave sizable imprints on the GW waveform (Ferreira et al. 2017; Hannuksela et al. 2019; Macedo et al. 2013a; Zhang and Yang 2020).

Ultralight scalars can also form self-gravitating structures. These “dark matter stars” are candidates to describe (dark) haloes in the central part of galaxies (Hui et al. 2017). In this context, it is crucial to understand how such structures respond to external excitations caused by compact objects (e.g. stars or BHs) living in these environments. Stability, proper modes, excitation of resonances, and depletion of the scalar through radiation are amongst the important issues that must be addressed here (Annulli et al. 2020a, b). There is also the potential for distinctive DM density enhancements due to accretion of DM onto BHs (Bamber et al. 2021; Clough et al. 2019; Hui et al. 2019), which is analogous to the case of dark matter spikes for high mass candidates (see Sect. 4.2 below). However, such density enhancements are estimated to be several orders of magnitude weaker than those from a superradiant build up, and concentrated near the BH horizon. They are therefore below LISA sensitivity in the inspiral phase, but may have an impact on merger, something which is yet to be tested using NR simulations. On the other hand, it is also essential to know how the compact objects themselves are affected by the surrounding dark matter. GW dephasing through dynamical friction and scalar radiation deserves a
proper and rigorous exploration (Annuli et al. 2020a, b; Hui et al. 2017). In particular, for sources in the LISA band, it was recently found that scalar radiation from these structures affects the gravitational waveform at leading -6 PN order with respect to the dominant quadrupolar term (Annuli et al. 2020a, b).

In summary, LISA has the potential to detect or constrain ultralight DM fields in regions of parameter space complementary to those covered by ground-based GW observations, although more work is needed to build waveform models that incorporate the effects of such fields and are sufficiently accurate to be used for data analysis purposes. It will also be important to explore in more detail BBHs with ultralight boson clouds with full NR simulations in order to understand how the presence of such boson clouds could be imprinted in the late stages of BBH mergers.

4.2 High mass particles \( m > eV \)

We now discuss what we could learn with LISA about more massive cold DM particles. If one or both the components of a binary system are surrounded by a dense dark matter overdensity, the binary’s inspiral will be driven by both dynamical friction (Chandrasekhar 1943) and GW emission. The densities required to produce an effect observable by LISA are typically higher than the characteristic DM densities in most galaxies (Barausse et al. 2014; Cardoso and Maselli 2020). There are however several mechanisms by which the DM density can be enhanced to such higher levels, notably when a smaller seed BH grows in mass adiabatically in a DM halo to produce a DM “spike” (Ferrer et al. 2017; Gondolo and Silk 1999; Quinlan et al. 1995; Sadeghian et al. 2013; Ullio et al. 2001). DM spikes around SMBHs at the center of galactic halos are prone to disruption by a number of processes, such as major mergers and gravitational interactions with stellar cusps (Bertone and Merritt 2005; Merritt et al. 2002). DM spikes around IMBHs might be more stable over cosmological timescales, and are arguably more promising targets for indirect and GW searches (Bertone et al. 2005b). The density profile of spikes depends on the nature of DM, with warm, self-interacting, or self-annihilating DM candidates leading to shallower profiles with respect to cold, collisionless DM (Alvarez and Yu 2021; Bertone and Merritt 2005; Gondolo and Silk 1999; Hannuksela et al. 2020; Shapiro and Shelton 2016).

The faster rate of inspiral in the presence of DM can be distinguished from the slower inspiral in vacuum, thereby allowing LISA to infer the presence of DM around a binary (Barausse et al. 2014; Eda et al. 2015; Macedo et al. 2013a). The prospects for detecting DM spikes through their influence on the GWs emitted by IMRIs were studied in (Eda et al. 2015), in which the dynamics of the IMRIs were treated in the leading Newtonian limit and the DM distribution was assumed to be static throughout the inspiral. Fisher-matrix forecasts predict that the properties of the DM spike could be mapped precisely for the higher overdensities (comparable to the errors on the detector-frame chirp mass) and to a few tens of a percent for lower overdensities.

Conservation of energy means, however, that the energy dissipated through dynamical friction must be balanced by a corresponding change in the energy of the dark matter distribution. For a range of different DM overdensities around different
IMRIs, the energy dissipated when the DM spike is assumed to be static turns out to be a substantial fraction of the binding energy of the spike, rendering the assumption of staticity questionable. In these cases, it is important to jointly evolve both the IMRIs’ orbits and their surrounding distributions of DM (Kavanagh et al. 2020). When this is done, the DM distribution is found to evolve non-trivially, significantly decreasing the amplitude of the dynamical friction force and hence the difference between an inspiral in vacuum and in a DM mini-spike (Kavanagh et al. 2020). Yet for larger DM overdensities, the amplitude of the effect of the DM is seen to remain significant in the coupled evolution of the IMRI and DM spike.

A Bayesian analysis of the detectability (signal-to-noise ratio), discoverability (discrimination against in-vacuum inspiral), and measurability (prospects for measuring the parameters) of dark matter environments shows that inspirals in presence of dark matter can be easily discriminated against in-vacuum inspirals. In case of detection, the DM halo’s density normalization can be distinguished from zero at high significance, with a 95% credible interval of order 10%. The DM halo’s slope can also be measured, although the posterior exhibits degeneracy with the chirp mass and the mass ratio. Interestingly, unlike in-vacuum inspirals, the mass ratio can be measured even in the Newtonian limit (Coogan et al. 2022).

For EMRIs and IMRIs in vacuum, there is good evidence that gravitational waveforms that are accurate to first post-adiabatic (1PA) order will be sufficient to detect and perform parameter estimation (van de Meent and Pfeiffer 2020). Such vacuum waveforms are under development and are expected to be ready for the beginning of the LISA mission. By contrast, the waveform modelling for IMRIs and EMRIs with surrounding DM spikes coupled to their evolution is only just beginning: the calculations in Kavanagh et al. (2020) assumed circular orbits and worked at leading Newtonian order. A more complete modelling of these systems will be necessary to produce waveforms that are suitable for describing the full range of possible orbits and DM distributions to cover the parameter space of these systems. The possibility of finding a simpler effective description of IMRIs and EMRIs in DM spikes should be investigated in greater detail, in order to decrease the size of the parameter space needed to look for signatures of cold DM substructures. At the other end of the scale, as in the light dark matter case, it may also be interesting to study whether there is any significant impact of these dark matter overdensities on the merger signal using NR simulations.

### 4.3 PBHs $m < M_{\odot}$

PBHs might originate in the early universe from the collapse of large density perturbations from an enhancement of the scalar curvature perturbation power spectrum generated during inflation (Ivanov 1998; Ivanov et al. 1994; Garcia-Bellido et al. 1996). PBHs may therefore span a wider range of masses compared to stellar BHs, which are expected to have a mass larger than the Chandrasekhar limit, around the solar mass, from stellar evolution.

For masses smaller than about $10^{-17}M_{\odot}$, PBHs would have evaporated by now due to the emission of Hawking radiation, and thus do not lead to late-time imprints.
For masses larger than this lower limit, while many observational bounds can be set on the PBH abundance in the Universe, these still leave some open windows for PBHs to constitute the totality of DM. A comprehensive review of the constraints on the PBH population can be found in Carr et al. (2021).

For PBHs with masses smaller than about the solar mass, the most stringent constraints on the PBH abundance arise from evaporation limits (Barnacka et al. 2012) and microlensing observations (Alcock et al. 2001; Allsman et al. 2001; Niikura et al. 2019a, b; Oguri et al. 2018; Smyth et al. 2020). Interestingly, there exists an open window for PBHs with masses in the range $(10^{-16} - 10^{-11}) \text{M}_\odot$ to account for the totality of the dark matter in the universe (see Carr et al. 2021; Green and Kavanagh 2021 for recent reviews). This range of masses is however notoriously difficult to probe with lensing observations. Indeed it was shown that both femtolensing and microlensing searches are not able to constrain this mass window once the extended nature of the source as well as wave optics effects are properly considered (Katz et al. 2018; Montero-Camacho et al. 2019; Smyth et al. 2020). Also, bounds coming from the survival of white dwarfs (WDs) and NSs after asteroidal mass PBHs capture were shown to be unable to constrain the abundance in this range (Montero-Camacho et al. 2019). Finally, the GWs from mergers of sub-solar mass PBHs, peaking at much higher frequencies compared to the ones testable by LISA, would not lead to an observable signal. However, LISA can search for SGWB signatures of the PBH formation in that interesting window, thus testing the possible nature of dark matter as asteroidal-mass PBHs, as we will describe in the following.

Curvature perturbations, responsible for the PBH production, would also lead to the emission of a second-order induced SGWB due to the intrinsic non-linear nature of gravity (Acquaviva et al. 2003; Mollerach et al. 2004), analysed in details in Ananda et al. (2007), Ando et al. (2018), Bartolo et al. (2018, 2019b, 2019c), Baumann et al. (2007), Bugaev and Klimai (2010), Chang et al. (2020), Clesse et al. (2018), De Luca et al. (2020a), Domènech (2020), Domènech et al. (2020), Espinosa et al. (2018), Fumagalli et al. (2021), García-Bellido et al. (2016, 2017), Inomata and Terada (2020), Kohri and Terada (2018), Wang et al. (2019), Pi and Sasaki (2020), Saito and Yokoyama (2010) and Yuan et al. (2020). Since the emission mostly takes place when the length scale of the perturbation crosses the cosmological horizon, one can relate the peak of the GW frequency spectrum to the dominant PBH mass $M_{\text{PBH}}$ as (Saito and Yokoyama 2010)

$$f_\star \simeq 3 \text{ mHz} \left( \frac{M_{\text{PBH}}}{10^{-12} \text{M}_\odot} \right)^{-1/2}.$$ (4)

There are several current and future experiments searching for this SGWB in various frequency ranges. In the ultra-low frequency range (around nHz), the null detections at Pulsar Timing Array experiments like PPTA (Shannon et al. 2015), NANOGrav (Arzoumanian et al. 2018) and EPTA (Lentati et al. 2015), give rise to stringent

---

6 Note that even perturbations that do not create PBHs contribute to the GWs produced, and it is the total of all these contributions that must be considered (De Luca et al. 2019b).
constraints on the abundance of GWs (although NANOGrav shows some evidence for a background, as discussed further below). Future experiments like SKA (Dewdney et al. 2009) (see also Moore et al. 2015) will significantly improve the detection sensitivity. These constraints can also be translated into a bound on the amplitude of the comoving curvature perturbations at the corresponding scales (Inomata and Nakama 2019; Ünal et al. 2021). Recently, the Earth-Moon system has been suggested as a detector of SGWB at µHz frequencies (Blas and Jenkins 2022a, b), providing a complementary probe to LISA for signals in this range.

In the intermediate frequency range (around mHz) the most relevant experiment is LISA. LISA will be able to probe the wavenumbers of the primordial curvature perturbation power spectrum on scales in the range \((10^{10} - 10^{15}) \text{Mpc}^{-1}\) (García-Bellido et al. 2016). Moreover, the production of PBHs with masses around \(M_{\text{PBH}} \sim \mathcal{O}(10^{-15} - 10^{-8})M_{\odot}\) would be associated to GW signals with peak frequencies which fall within the LISA sensitivity band. LISA would be therefore able to search for DM candidates in the form of PBHs within this range of masses (Bartolo et al. 2019b; García-Bellido et al. 2016, 2017). Notice that, due to the exponential sensitivity of the PBH abundance to the amplitude of the curvature power spectrum, null detection of a SGWB at LISA would imply the PBH abundance in the corresponding mass range to be negligible.

Physical SGWB spectra would typically have a white-noise \((\propto f^3)\) behaviour at low frequencies (Cai et al. 2020; Espinosa et al. 2018), a peak corresponding to the frequency provided in Eq. (4) and a high frequencies tail which goes like the squared curvature power spectrum, \(\Omega_{\text{GW}}(f \gg f_*) \sim \mathcal{P}_S^2(f)\). PBH formation from multi-field inflationary models (Fumagalli et al. 2020; Palma et al. 2020), and more generally large particle production mechanisms involving a sharp feature, lead to oscillatory power spectra that result in a 10% oscillatory modulation in the peak of the frequency profile of the SGWB (Braglia et al. 2021; Fumagalli et al. 2021), thus representing a potentially observable characteristic of these models.

Recently the NANOGrav Collaboration has published a 12.5 yrs analysis of pulsar timing data reporting strong evidence for a signal whose interpretation in terms of a stochastic common-spectrum process is strongly preferred against independent red-noise signals (Arzoumanian et al. 2020). At the moment, the NANOGrav Collaboration does not yet claim a detection of a SGWB due to the absence of evidence for the characteristic quadrupole correlations. There are several models proposed so far to explain the origin of this SGWB signal. One possibility is provided by the SGWB associated to PBH formation (De Luca et al. 2021b; Domenech and Pi 2020; Kohri and Terada 2021; Vaskonen and Veermäe 2021). The relative PBH abundance in the mass range probed by NANOGrav would be subdominant (Vaskonen and Veermäe 2021). However, the model proposed in De Luca et al. (2021b), where a flat power spectrum of the curvature perturbation (Biagetti et al. 2018; Leach and Liddle 2001; Leach et al. 2001; Wands 1999) leads to a mass function peaked in the asteroidal mass range (De Luca et al. 2020e), a total abundance of DM in the forms of PBHs allowed by observational bounds and hinted by the lensing event candidate in the HSC data (Kusenko et al. 2020; Sugiyama et al. 2021).
This scenario predicts a flat SGWB spectrum which would be detectable and tested by LISA (De Luca et al. 2021b).

As discussed in more detail in the complementary white paper of the Cosmology WG (Auclair et al. 2022), LISA will also be able to search for signatures of primordial non-Gaussianity (Biagetti et al. 2018, 2021; Ezquiaga and García-Bellido 2018; Ezquiaga et al. 2020a; Franciolini et al. 2018; Pattison et al. 2021) at small scales. Primordial non-Gaussianity on small scales would have an impact on the shape of the SGWB, leading to potentially detectable signatures in the spectrum of frequencies of the observed monopole signal (Cai et al. 2019a; Unal 2019). Possible small deformations smearing the GWs spectrum can also arise from similar effects (Domcke et al. 2020). Non-Gaussian signatures in the tensor three-point function would in principle be significant, but get washed out by (Bartolo et al. 2019b, c) time-delay effects originating during the propagation of the GWs in the perturbed universe, see Bartolo et al. (2018); Margalit et al. (2020). Non-Gaussian signatures may be detectable by searching for angular anisotropies in the SGWB which, in the absence of non-Gaussianity, are generally undetectable at LISA (Bartolo et al. 2020) given its angular resolution of $\ell \lesssim 15$ (Baker et al. 2021; Contaldi et al. 2020). However, local scale-invariant non-Gaussianity, constrained by the Planck Collaboration to be $-11.1 \leq f_{\text{NL}} \leq 9.3$ at 95% C.L. (Planck Collaboration et al. 2020c), would correlate short and long scales and potentially lead to an enhancement of the SGWB, see Cai et al. (2019a, 2019b); Unal (2019); for details, and possible large-scale anisotropies at detection (Bartolo et al. 2020). Now if PBHs constitute a large fraction of the DM, one expects a highly isotropic and Gaussian SGWB due to the strong constraints by CMB observations (Planck Collaboration et al. 2020b) on the isocurvature modes in the DM density fluid, associated to the non-Gaussianity (Young and Byrnes 2015). On the other hand, the detection of a large amount of anisotropy in the signal would imply that only a small fraction of the DM can be accounted by PBHs (Bartolo et al. 2020). Moreover, the propagation of GWs across disconnected regions of the universe may lead to large-scale anisotropies at detection, see e.g., Alba and Maldacena (2016), Bartolo et al. (2019a), Bartolo et al. (2020), Bertacca et al. (2020), Contaldi (2017), Cusin et al. (2017, 2019), Jenkins and Sakellariadou (2018) and Renzini and Contaldi (2019).

In this range of masses, the non detection of a SGWB will rule out the possibility of PBHs being the dark matter in the asteroidal mass range. However, this assumes the standard scenario where GWs are produced at PBH formation and the PBH abundance is exponentially sensitive to the size of the fluctuations. One should investigate whether non standard scenarios would change this conclusion, e.g. the potential role of non-Gaussianity or of other PBH formation mechanisms.

### 4.4 PBHs $m > M_{\odot}$

For PBHs with masses larger than about a solar mass, LISA should be able to detect merger events of resolved sources, and unresolved signals in the form of a SGWB. The most relevant constraints come from CMB temperature and polarization anisotropies, which are impacted by the emission of ionizing radiation from PBHs
accreting gas in the early universe at redshift between \( z \sim (300 - 600) \), for which the PBH abundance is constrained to be below \( f_{\text{PBH}}(M) \lesssim (M/10M_\odot)^{-4} \) for masses smaller than \( 10^4 M_\odot \) (Ali-Haïmoud and Kamionkowski 2017; Carr et al. 2021; Green and Kavanagh 2021; Serpico et al. 2020). At larger masses, constraints coming from CMB distortions become even more stringent (Nakama et al. 2018).

Late-time constraints can be set from the comparison of emitted EM signals from accreting PBHs with observations of galactic radio and X-ray isolated sources (Gaggero et al. 2017; Manshanden et al. 2019) and X-ray binaries (Inoue and Kusenko 2017), and from Dwarf Galaxy Heating observations (Lu et al. 2021). The totality of these bounds constrain the PBH abundance \( f_{\text{PBH}} \) (the fraction of the dark matter composed of PBHs) to be below \( f_{\text{PBH}} \lesssim 10^{-2} \) for masses larger than few tens of \( M_\odot \). A comparable constraint is set by the merger rates observed by the LIGO/Virgo Collaboration (Abbott et al. 2019c). Indeed, for masses around \( 30 M_\odot \), a PBH abundance larger than \( f_{\text{PBH}} \approx 10^{-3} \) would give an expected number of events larger than the one observed. This has been shown in De Luca et al. (2020c), Hall et al. (2020) and Raidal et al. (2019); using the GWTC-1 data and recently in Wong et al. (2021) with the GWTC-2 catalog. Even though PBHs are expected to follow a Poisson distribution at formation in the absence of non-Gaussianities (Ali-Haïmoud 2018; Ballesteros et al. 2018; Desjacques and Riotto 2018; Moradinezhad Dizgah et al. 2019), they may start forming clusters even before the matter-radiation equality for large enough \( f_{\text{PBH}} \) (Inman and Ali-Haïmoud 2019). Clustering has an impact on the PBH binary formation and subsequent disruption rates (Jedamzik 2020, 2021; Raidal et al. 2019; Trashorras et al. 2021). While ineffective for abundances smaller than \( f_{\text{PBH}} \lesssim 10^{-2} \), it is expected to reduce the merger rate prediction for larger abundances by an amount which is not enough to evade the LIGO/Virgo bound if one takes into account the fraction of PBHs in dense substructures (De Luca et al. 2020; Tkachev et al. 2020; Vaskonen and Veermäe 2020). Overall, we stress that a complete description of PBH clustering up to very low redshift along with its impact on the GW signals coming from a PBH population still represents one of the main theoretical challenges, which deserves a thorough investigation.

Even if PBHs within this mass range can only account for a small fraction of the dark matter in the universe, they could still both form binaries observable by LISA and act as progenitors of SMBHs, observed at high redshift as large as \( z \gtrsim 6 \), through accretion of baryonic particles during the cosmological evolution (Clesse and García-Bellido 2015; Serpico et al. 2020). The presence of a strong phase of accretion through the cosmological history before the reionization epoch has the effect of shifting the corresponding late-time mass function to larger masses relaxing the constraints (De Luca et al. 2020b), even though large uncertainties are present in modelling the accretion for massive compact objects. This is a key challenge to overcome in order to assess the PBH contribution to merging binaries observable by LISA. Though difficult to conclusively associate with PBHs, the observation of events within this range of masses at high redshift by experiments like LISA would help in understanding the physical origin of SMBHs (Volonteri 2010).

Given that PBHs may potentially contribute to current GW data (Franciolini et al. 2021), within this range of masses they need to be distinguished from astrophysical
BHs formed after the collapse of stars in the late universe. Various observables could however help in discriminating the two populations. For instance, even though PBHs are expected to form with small spins (De Luca et al. 2019a; Mirbabayi et al. 2020), for PBH binaries in the mass range observable by LISA, the baryonic mass accretion leads to spin growth, potentially increasing their spins for the binary components at the merger time depending on the accretion strength before the reionization epoch (De Luca et al. 2020d), which is still however affected by many uncertainties. It is of crucial importance to investigate in details the correlation between PBH masses and spins induced by baryonic accretion as a discriminant with the prediction for astrophysical BHs. Another possibility is to look for high redshift events (i.e. \( z \gtrsim 30 \)) where PBHs can contribute without any astrophysical contamination (Chen and Huang 2020; De Luca et al. 2021; Kousshiappas and Loeb 2017).

Furthermore, the massive BBH merger population currently observed at LIGO/Virgo is expected to give rise to a SGWB of unresolved sources with a tail at low frequencies which can be detected by LISA (Chen et al. 2019; Clesse and García-Bellido 2017). The primordial scenario is expected to give a stronger contribution to the SGWB with respect to the astrophysical one, due to the additional contribution given by PBH mergers happening at higher redshift (Raidal et al. 2019). In particular, the PBHs merger rate is expected to increase with redshift, while the one of stellar BHs first increases and peaks at redshift around \( z \sim (1 - 2) \), and then rapidly decreases (Chen et al. 2019). Considering both observations of resolved and unresolved signals may help to break the degeneracy within the two scenarios (De Luca et al. 2021a; Mukherjee and Silk 2021; Mukherjee et al. 2022; Wang and Kohri 2021).

LISA will also be able to detect the inspiral phase of binaries with masses of at least a few tens \( M_{\odot} \) with a high precision. This mass range, falling in the large mass portion of the window observable by the LIGO/Virgo Collaboration, can be filled by a PBH population. The presence of the known mass gap in the astrophysical BH spectrum and the recent event GW190521 (Abbott et al. 2020c) make the primordial scenario investigated in Clesse and García-Bellido (2020); De Luca et al. (2021); Kritos et al. (2021) potentially relevant (see however Abbott et al. 2020e for different scenarios).

Finally, the detection of GWs from the coalescence of a BBH system including a compact object lighter than a solar mass could indicate primordial origin. Experiments like LISA, which target low frequency regimes, may detect GWs emitted from EMRIs of a SMBH with a light compact PBH, and GW constraints can be set on the PBH abundance from the expected PBH-SMBH merger rate (Guo et al. 2019; Kühnel et al. 2017; Kuhnel et al. 2020). A null-detection during a 5-year operation of the experiment would constrain the PBH abundance to values smaller than \( f_{\text{PBH}} \lesssim 3 \times 10^{-4} \) for PBHs with masses in the range \((10^{-2} - 1) M_{\odot}\) (Guo et al. 2019).

Another interesting LISA observable with relevance to the DM PBH scenario arises from GW bursts from hyperbolic encounters of PBHs (García-Bellido and Nesseris 2018). This would be particularly relevant for the clustered and broad mass spectrum PBH scenario, where the BHs cover a wide range of masses with a peak
\( \sim 1 \, M_\odot \) (Carr et al. 2021; Ezquiaga et al. 2018; García-Bellido 2017; García-Bellido and Ruiz Morales 2017), and have a clustered distribution which allows them to evade some, but not all, of the constraints discussed above, see García-Bellido (2018, 2019) for a review.

To summarize, LISA would be sensitive to many different GW signals coming from a PBH population in this mass range. Given its sensitivity to BH mergers at high redshifts and large masses, even reaching the supermassive range, it will complement the reach of ground-based detectors to search for PBHs and inform models of PBH accretion. On the theoretical side, a future goal is to sharpen the predictions for the impact of accretion and clustering on the evolution of PBH binaries and the generation of SMBHs from PBH seeds (Serpico et al. 2020), in order to differentiate between BHs with a primordial origin and those formed from stellar evolution.

### 4.5 Exotic compact objects

If Exotic Compact Objects (ECOs) exist, they could compose some fraction of the dark matter. ECOs have already been discussed in Sec. 3.2.2, and for those with compactness similar to BHs, the opportunities for LISA to observe and constrain them will be similar to the PBH cases described in Sec. 4.3 and Sec. 4.4 above.

### 4.6 Signatures from self-interactions

Self-interactions in DM were proposed as a way to solve tensions on small scales between N-body simulations and DM observations (Spergel and Steinhardt 2000), but for hard interactions these are constrained to be of order \( \sigma/m < 1 \, \text{cm}^2 \, \text{g}^{-1} \sim 10^{-24} \, \text{cm}^2 \, \text{GeV}^{-1} \) to maintain consistency with structure formation and observations of the Bullet Cluster (Randall et al. 2008). At the upper end of this range, they have a non-negligible effect on structures at typical DM halo densities and velocity dispersions (see Tulin and Yu 2018 for a review). Smaller cross sections would have an impact only in more dense regions, e.g. in DM spikes around BHs, and so self-interacting dark matter (SIDM) could potentially have interesting consequences. However, we have yet to see a model with sufficient capture to realistically manifest these, at least where the SIDM composes the entirety of the DM.

It has been proposed that a small amount of SIDM, composing less than 10% of the DM, with cross section \( \sigma/m \gg 1 \, \text{cm}^2/\text{g} \), could have seeded the initial growth of SMBHs (Pollack et al. 2015), which are a key LISA target. However, it seems unlikely that such a formation mechanism could be identified from LISA observations. SIDM could also undergo dissipative collapse to form compact dark disks (Fan et al. 2013), which could potentially be detectable via their consequent density enhancements in the same way as the superradiant clouds or DM spikes described above.

In the case of light bosonic DM, self-interactions may play an important role in superradiant instabilities and, moreover, are an essential part of some models such as...
axions, where the massive potential is an approximation to a more general periodic function. Studies of superradiance have mostly treated self-interactions in the bosonic field as negligible, but in Yoshino and Kodama (2012, 2014, 2015) attractive self-interactions were shown to destabilise the superradiant build up, resulting in a so-called “bosenova”. Such collapses have been conjectured to result in repeating instabilities, where the fall of the cloud back into the BH prior to saturation results in it being “spun up” again, and thus the superradiant growth resumes. This could weaken constraints that have been derived in the purely massive case, or lead to potentially observable GW signals from the bosenova explosions. The frequency of the bosenova signal would be \( f \sim 1/\Delta t \) where \( \Delta t \) is the duration of the bosenova. This is in the LISA band for SMBHs, but would be below detection sensitivity unless it occurs in our own or a nearby galaxy (Yoshino and Kodama 2012). More work is required to simulate a complete build up and dissipation of the bosonic cloud, to further quantify the GW signals produced and to confirm that the superradiant build up is able to recover. Due to the long timescales involved, such simulations are very challenging, but feasible with current NR tools.

4.7 Multi-messenger signatures

If dark matter particles interact at all with standard model particles, the interaction must be weak, given the constraints that have been imposed by terrestrial direct detection experiments and indirect astrophysical observations (see e.g., Bertone et al. 2005a; Irastorza and Redondo 2018; Kahlhoefer 2017 for reviews). However, some dark matter candidates, such as the QCD axion, have a well defined coupling to standard model matter. In this case, coincident GW and EM signatures, for example, during SMBH mergers or in EMRIs, may provide distinctive signatures. This has been explored in the case of the QCD axion in Edwards et al. (2020), on the basis of LISA observations of dephasing in the GW signal of a NS inspiral around an IMBH embedded within a DM spike. Such observations could identify axions in the mass range \( 10^{-7} - 10^{-5} \) eV by matching the signal to radio wave emission from axion-photon conversion. Similar multi-messenger effects could be relevant in other cases, such as where GWs from superradiant clouds or ECOs are detected, see e.g., Amin and Mou (2021); Blas and Witte (2020a, 2020b); Cardoso et al. (2021); Ikeda et al. (2019); for recent work in this direction.

4.8 Burning questions

- The effects of dark matter on GW signals may often be degenerate with ‘environmental’ astrophysical effects (cf. Sect. 7). It will be absolutely crucial for the success of LISA to devote continual effort to disentangle these as much as possible.
- LISA has the potential to detect or constrain the presence of both ultralight and heavier dark matter fields in regions of parameter space that are complementary to those covered by ground-based GW observations. However, the waveform
modelling for IMRIs and EMRIs embedded in DM halos is but in its infancy. More work is needed to build waveform models that incorporate the effects of DM fields in mergers, covering a significant parameter space, and that are sufficiently accurate for data analysis purposes.

- Beyond this, it will be important to study BBH dynamics with ultralight boson clouds with full NR simulations in order to understand how the presence, and the dynamics, of such boson clouds could be imprinted in the late stages of BBH mergers.
- To properly interpret the observables related to PBHs and identify their origin, the impact of non-Gaussianity or novel PBH formation mechanisms in the small mass range, and clustering and accretion in the mass range above $M_C$ up to SMBHs, should be further characterised.

5 Tests of the $\Lambda$CDM model and dark energy

The past 20 years have seen the emergence of a standard model of the Universe based on a handful of parameters: the so-called $\Lambda$CDM model. This framework can explain coherently a large number of observations but the nature of one of its main components, dark matter, is still unknown, and the observed value of the cosmological constant $\Lambda$, a crucial element of the model, is still mysterious.

In particular, many cosmological observations indicate that the Universe is currently undergoing accelerated expansion. To explain this observation assuming GR, the $\Lambda$CDM model requires that the energy density of the vacuum is comparable to the total energy density of the Universe. However, quantum vacuum fluctuations of all particles contribute to the vacuum energy by an amount which is expected to be many orders of magnitudes larger than its measured value. This mismatch is called the old cosmological constant problem (see e.g., Burgess 2015; Padilla 2015) and it is one of the most disconcerting puzzles in fundamental physics (see e.g., Alberte et al. 2016; Bousso 2008; Charmousis et al. 2012; Kaloper and Padilla 2014).

This situation constitutes a theoretical motivation to consider alternatives to $\Lambda$CDM in which one assumes that for some unknown reason the value of $\Lambda$ is simply zero—dubbed the new cosmological constant problem—and attempts to explain the accelerated expansion by different mechanisms. This has spurred the exploration of variations of the $\Lambda$CDM model in which the energy budget of the Universe is dominated by some dynamical dark energy, distinct from a constant $\Lambda$, or where the observed acceleration is explained by a modification of gravity (see Amendola et al. 2018b; Clifton et al. 2012 for a review). This field of investigation has been rich of interesting theoretical developments and is currently energized by the prospect that a number of forthcoming galaxy surveys will test and tightly constrain these models in the near future (see e.g., Frusciante and Perenon 2020). A second theoretical motivation for considering modified gravity extensions of $\Lambda$CDM is the so-called coincidence problem (Zlatev et al. 1999), the observation that the current vacuum energy density is comparable to the matter density, even though the former remains constant in time while the latter dilutes with the expansion.
Quite apart from these theoretical considerations, several tensions have emerged in the ΛCDM model from recent observations. The most notable one is the discrepancy between the values of the Hubble constant measured through CMB (Planck Collaboration et al. 2020a) and large-scale structure (D’Amico et al. 2021; Philcox et al. 2020) observations on the one hand, and those measured by late-time probes (Huang et al. 2020; Jee et al. 2019; Pesce et al. 2020; Riess et al. 2019; Verde et al. 2019) on the other hand. Other milder tensions include the difference between the amplitude of matter fluctuations measured by Planck versus that inferred from weak lensing observations (Heymans and other 2021; Tröster et al. 2020), as well as the lensing excess in the CMB temperature Planck data (Planck Collaboration et al. 2014, 2016, 2020a). While it is possible that these tensions could be accounted for by some yet unknown systematics, several proposals have attempted to explain these discrepancies by new physics.

In this section we summarize how GW astronomy, in particular the LISA mission, can inform theoretical physics in the cosmology area, in particular by probing the fundamental physics underlying the ΛCDM model and its extensions involving dark energy and modified gravity. To this extent this section is complementary to the Sects. 2 and 4 of the White Paper of the LISA Cosmology WG (Auclair et al. 2022), whose main goal is to develop the appropriate methods to probe cosmology with GW observations. In the next subsection we present an overview of the dark energy and modified gravity models of interest for LISA. We then discuss the theoretical implications and the constraints on GWs propagating in a homogeneous universe in Sect. 5.2. We extend this analysis to an inhomogeneous universe in Sect. 5.3, by discussing cosmological tests that involve the cross-correlation of GWs with the large-scale structure of the Universe and the use of the GW lensing to probe the distribution of structures. Future prospects are discussed in Sect. 5.4.

5.1 Dark energy and modified gravity

Dark energy and modified gravity are extensions of the ΛCDM scenario. Although a strict distinction between the two does not exist (Joyce et al. 2016), the former usually refers to adding to the field equations the stress-energy tensor of a fluid responsible for the acceleration, leaving the Einstein tensor unchanged, while the latter refers to either changing the gravitational part of the field equations or adding a non-minimal coupling to matter. Since at low-energy GR is the unique Lorentz-invariant theory of a massless helicity-2 field (Weinberg 1964), in most cases (see e.g., Lin and Mukohyama 2017 for an exception) one must add additional degrees of freedom to those present in GR.

5.1.1 Effective field theory of dark energy

In the simplest case, dark energy and modified gravity rely on the presence of a scalar field that spontaneously breaks time parametrizations, inducing a preferred slicing. Since in cosmology we are interested in studying fluctuations around FLRW solutions, a particularly convenient way to parametrize the action in this case is to use the so-called Effective Field Theory of Dark Energy (Bloomfield et al. 2013;
Creminelli et al. 2009; Gleyzes et al. 2013; Gubitosi et al. 2013), which consists in writing the action in terms of all possible operators in this foliation (Cheung et al. 2008; Creminelli et al. 2006). This allows one to describe a large class of models, including $f(R)$ theory (see Sotiriou and Faraoni 2010 for a review), Horndeski theories (Deffayet et al. 2009; Horndeski 1974) and beyond (Ben Achour et al. 2016; Gleyzes et al. 2015; Langlois and Noui 2016; Zumalacárregui and García-Bellido 2014), and Lorentz violating theories (Horava 2009a), in terms of a fewer number of parameters, corresponding to time-dependent functions. More generally, requirements such as locality, causality, unitarity and stability can be automatically enforced on the action so that the predicted signatures are consistent with well-established principles of physics. The generality of this approach has been successfully applied to the implementation of Einstein–Boltzmann solvers such as EFTCAMB (Hu et al. 2014; Raveri et al. 2014) and hi_class (Zumalacárregui et al. 2017) and the derivation of observational constraints and forecasts in a model-independent fashion (see Frusciante and Perenon 2020 for a recent review).

In these extensions of $\Lambda$CDM, we can identify two main classes of modifications to the propagation of GWs with observational implications. The first is a scale-independent additional friction term in the propagation equation, discussed in Sect. 5.2.2, which can be traced to the fact that the effective Planck mass that canonically normalizes the graviton can be time-dependent in models of modified gravity. In this case, the effect amounts simply to a rescaling of the graviton amplitude between the time of emission and the one of observation. The second class are scale-dependent modifications such as an anomalous speed of propagation, effects in the dispersion relation, GW decay, etc. These are thoroughly discussed in Secs. 2.2.2 and 6.3.2 and will not be treated here.

5.1.2 Non-local model

Another possibility, leading to scale-independent modifications that have been much investigated recently, is that the long-distance behavior of gravity is modified by quantum effects. The presence of infrared (IR) divergences in space-times of cosmological interest, such as de Sitter (Antoniadis and Mottola 1991; Taylor and Veneziano 1990; Tsamis and Woodard 1995), suggests that, even without adding new degrees of freedom or modifying the fundamental action of the theory, the long-distance dynamics of gravity could be different from that derived from the classical Einstein–Hilbert action. When quantum effects enter into play the relevant quantity, rather than the fundamental action, is the quantum effective action. While the former, according to basic principles of quantum field theory, is local, the latter unavoidably develops non-local terms whenever the spectrum of the theory contains massless particles, such as the graviton.

An extensive series of investigations (see Belgacem et al. 2018b, 2020 for reviews) have eventually led to identify a unique candidate model (the so-called ‘RT’ model originally proposed in Maggiore 2014). Physically, this model corresponds to assuming that IR effects generate a non-local mass term for the conformal mode of the metric. At the conceptual level, this is appealing because of various arguments that identify the conformal mode as the main candidate for producing strong IR
quantum effects (Antoniadis and Mottola 1991, 1992). Tentative evidence for the generation of this specific non-local term in the quantum effective action has also emerged from lattice gravity simulations (Knorr and Saueressig 2018). At the phenomenological level, the model passes Solar System tests (Kehagias and Maggiore 2014) and limits on the time variation of Newton’s constant (Belgacem et al. 2019b), it generates a dark energy dynamically (Maggiore 2014), and its cosmological perturbations, in the scalar sector, are well-behaved and quite close to those of $\Lambda$CDM (Dirian et al. 2014). The model has been compared to CMB, BAO, SNe and structure formation data, and fits them at the same level as $\Lambda$CDM (Belgacem et al. 2018b; Dirian et al. 2016). A main surprise then came from the study of tensor perturbations (Belgacem et al. 2018b, 2019a): The model displays the phenomenon of modified GW propagation that will be discussed below, whose amplitude can be significant. At the redshifts relevant for LISA, the deviations with respect to GR can even be of order 80% (depending on a free parameter of the model, related to the initial conditions). A complete and up-to-date review of the conceptual and phenomenological aspects of the model is given in Belgacem et al. (2020).

Apart from its specific features, this model also provides an explicit example of a modification of GR that is very close to $\Lambda$CDM, at the percent level, for the background cosmological evolution and for the scalar perturbations (therefore complying with existing limits), while at the same time producing very large deviations from GR in the sector of tensor perturbations. In the context of scalar-tensor theories described by the effective field theory of dark energy, a similar case is discussed in Lombriser and Taylor (2016). This shows how the study of GWs on cosmological scales is a genuinely new window, accessible to LISA, that can produce significant surprises.

5.1.3 Early and interacting dark energy

Other interesting models to test with LISA are the so-called early dark energy (Wetterich 2004) and interacting dark energy (Amendola 2000; Wetterich 1995) scenarios. The former consists of a redshift-dependent dark energy component which gives a non-negligible cosmological contribution at early times, in contrast to the more standard dark energy used to explain the late-time cosmic acceleration. The presence of early dark energy in the cosmological energy budget affects the CMB anisotropies and polarization and the inferred constraints can change by varying the epoch at which early dark energy starts to contribute (Pettorino et al. 2013). In the context of LISA, early dark energy can be tested by measuring the distance-redshift relation at high redshift (see Sect. 5.2.2). As such, LISA can be complementary to other cosmological probes, such as the CMB, in helping to resolve the $H_0$-tension for which such early dark energy models are candidate solutions (Caprini and Tamanini 2016).

Interacting dark energy is characterized by a direct coupling between dark energy and dark matter, while ordinary matter remains uncoupled from dark energy and gravity is described by the ordinary Einstein–Hilbert action (i.e., in the so-called Einstein frame). It can be motivated, for instance, by the fact that it can alleviate the coincidence problem mentioned in the introduction (Amendola 2000; Wetterich...
In this scenario, the propagation of GWs is modified through a change in the Hubble friction arising from an altered cosmic expansion history (Caprini and Tamanini 2016; Dalang and Lombriser 2019). As for early dark energy, interacting dark energy can be tested by LISA by accurately probing the background expansion history and the luminosity distance-redshift relation.

5.1.4 Screening

Modifications of gravity that aim to explain the cosmic acceleration often entail extra light fields which, besides changing the cosmic expansion, manifest themselves at short distances, mediating a *fifth force*. To pass Solar System and binary-pulsar tests, screening mechanisms are therefore necessary to hide, or strongly suppress, local deviations from GR (see e.g., Joyce et al. 2015). Different types of screening mechanisms have been proposed over the years, all relying on non-linear physics becoming important close to matter sources. Here we discuss only modified gravity models based on a scalar field. In chameleon and symmetron screening (Hinterbichler and Khoury 2010; Khoury and Weltman 2004), non-trivial self-interactions and non-minimal couplings to matter dynamically suppress deviations from GR in high-density environments (see Burrage and Sakstein 2018 for a review), however these theories were shown to have a negligible effect on cosmological scales (Wang et al. 2012). Self-interactions involving one or two derivatives of the field are at the origin of, respectively, $k$-mouflage (Babichev et al. 2009) and Vainshtein Babichev and Deffayet (2013); Vainshtein (1972) screening. These kinetic types of screening allow to reconcile theories such as Horndeski and DHOST, or massive (bi)gravity, with local tests of gravity. However, it is worth noticing that, until recently, kinetic screening had been thoroughly tested in static/quasi-static configurations, while it was less studied in dynamical settings, e.g. at GW generation (Chu and Trodden 2013; Dar et al. 2019; de Rham et al. 2013b, a). Only very recently, numerical relativity simulations have been performed, which point at a partial breakdown of the screening in black-hole collapse (Bezares et al. 2021; ter Haar et al. 2021) and in the late inspiral and merger of binary neutron stars (Bezares et al. 2022). In more detail, stellar collapse seems (quite surprisingly) to produce a very low frequency signal potentially detectable by LISA, while waveforms from binary neutron stars seem to deviate from their GR counterparts at the quadrupole (but not dipole) multipole order.

In general, when screening is efficient, one can define three *gravitational constants*. The one experienced by GWs, $G_{gw}$, the one felt by matter appearing in the Poisson equation, $G_{dyn}$; the one felt by light intervening in null geodesics, $G_{light}$. These three constants are generally different and independent from each other (Dalang and Lombriser 2019; Lombriser and Taylor 2016; Tsujikawa 2019; Wolf and Lagos 2020), even when submitted to the constraints coming from Solar System tests, such as lunar-laser ranging (Williams et al. 2004), which tests $G_{dyn}$, and the Cassini bound (Bertotti et al. 2003), which tests a combination of $G_{dyn}$ and $G_{light}$. However, GW constraints can provide independent relations among them. For instance, in Horndeski theory the combination of Solar System tests and the fact that

---

7 This is proportional to the inverse square power of the effective Planck mass appearing in (12) below.
$c_T = 1$ imposes that $G_{gw}$ varies very little with the redshift, implying that observational signatures in distance tests could be challenging (Dalang and Lombriser 2019; Dalang et al. 2020; Lagos and Zhu 2020; Wolf and Lagos 2020) (see Eq. (12) below).

This conclusion, however, is evaded in theories that extend Horndeski gravity, such as GLPV and DHOST theories, where additional parameters can lift the degeneracy. In these theories, even when imposing $c_T = 1$, the three constants remain independent (Crisostomi and Koyama 2018b; Crisostomi et al. 2019; Dima and Vernizzi 2018; Langlois et al. 2018; Lombriser and Taylor 2016) and there can be effects in the luminosity distance. The recovered degeneracy can, however, be broken again if other constraints are imposed, such as those from the instability of GWs (Creminelli et al. 2020). Another example where these constants remain independent is massive gravity (de Rham et al. 2011) or bigravity (Hassan and Rosen 2012), see Babichev and Crisostomi (2013); Koyama et al. (2011). Finally, non-local models do not rely on screening and hence are not subject to these limitations.

### 5.2 Homogeneous cosmology

In this subsection we consider GWs propagating on a homogeneous universe; we postpone the discussion of perturbations to the next subsection. We assume a spatially-flat background space-time described by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric,

$$ds^2 = -dt^2 + a^2(t)\delta_{ij}dx^i dx^j,$$

(5)

where $a(t)$ is the cosmological scale factor. We describe GWs as perturbations of the spatial part of the above metric, $g_{ij}(t, x) = a^2(t)\left[\delta_{ij} + h_{ij}(t, x)\right]$, using the traceless and transverse gauge $h_{ij} = \partial_i h_{ij} = 0$. Moreover, we decompose $h_{ij}$ in its two polarisation modes, $h_{ij} = h_+ \epsilon^{\gamma}_{ij} + h_\times \epsilon^\gamma_{ij}$, where $\epsilon^{\gamma}_{ij}$ are the polarization tensors.

#### 5.2.1 Standard sirens

Standard sirens are GW sources at cosmological distances which provide a direct measurement of the luminosity distance. Standard sirens which come with an associated EM counterpart that enables one to determine their redshift can be used to infer the luminosity distance-redshift relation, and thus trace the evolution of the Universe independently of other distance ladders.

For inspiraling binaries, at leading order in the PN approximation, the GW strain reads, for the two polarizations (see e.g., Maggiore 2007),

$$h_+ = \frac{2(1 + \cos^2 t)}{d_L(z)} (GM_c)^{5/3} (\pi f)^{2/3} \cos \Phi,$$

(6)
\[ h_x = \frac{4 \cos i}{d_L(z)} \left( G M_c \right)^{5/3} \left( \frac{\pi f}{2} \right)^{2/3} \sin \Phi, \]  

(7)

where \( \Phi \) is the phase of the wave, \( M_c \equiv (1+z)\mu^{3/5}m^{2/5} \) is the redshifted chirp mass (\( \mu \) is the reduced mass and \( m \) the sum of the masses of the two bodies), \( f \) is the observed GW frequency, \( i \) is the inclination angle of the normal to the orbit with respect to the line of sight and \( d_L \) is the luminosity distance of the source, which for a spatially-flat universe reads \( d_L(z) \equiv (1+z) \int_0^{z} \frac{c}{H(z')} dz' \), where \( H \equiv \dot{a}/a \) is the Hubble parameter. One factor that hinders the determination of the luminosity distance is its degeneracy with the inclination angle. This degeneracy can be broken if higher order modes are included or if the binary has a precessing spin, in which case the characteristic modulation of the amplitude can disentangle the inclination angle.

Standard sirens with LISA will provide a formidable probe of the cosmic-expansion history, from low and intermediate redshifts of \( 0 \leq z \leq 1 \) with stellar-mass BBHs and EMRIs, to deep redshifts of \( z \leq 10 \) for MBBHs. This will naturally place stringent constraints on a great amount of dark energy models (see e.g., Belgacem et al. 2019; Tamanini et al. 2016). While the low-redshift data will be of particular interest for cosmic acceleration (Lombriser and Taylor 2016), the high-redshift data will provide invaluable constraints on early dark energy (Caprini and Tamanini 2016) in a regime that is not accessible to other observational methods such as large-scale structure surveys.

For this scope, redshift measurements are of primary importance. These can be obtained through a detection of an EM counterpart of the GW event. In the context of LISA only MBHBs are expected to produce counterparts at cosmological distances. There are non-standard scenarios to produce EM counterparts for SOBHBs and IMBHs, which could be used as multi-band standard sirens.

When an EM counterpart is not available, one can still obtain the redshift information in a statistical way. One method relies on the use of cross-correlation of each GW event with a galaxy catalogue, which provides a set of potential hosts inside the GW localization volume times its redshift uncertainty, computed from the GW distance measurements by taking into account the full prior range on all cosmological parameters (Abbott et al. 2021; Chen et al. 2018; Schutz 1986; Del Pozzo 2012; Fishbach et al. 2019; Gray et al. 2020; MacLeod and Hogan 2008; Palmese et al. 2020; Soares-Santos et al. 2019; Vasylyev and Filippenko 2020). Combination of multiple events can significantly increase the constraining power of this method, as results from spurious galaxies will eventually average out. On the other hand, one has to properly take into account accurate modelling of the redshift uncertainties of galaxy catalogues, completeness issues—which can be relevant for sources at high redshift—and GW selection effects. The ideal candidates in this case are events with good localization and distance measurement, so that the potential hosts are in limited number (or even unique; Borhanian et al. 2020), at redshifts such that the available catalogues have a sufficient completeness (Del Pozzo et al. 2018).

Alternatively, one can exploit the full three-dimensional spatial cross-correlation between GW sources and the galaxy distribution (Bera et al. 2020; Mukherjee et al. 2021). These techniques are the only possibility for LISA sources that are not
expected to produce an EM counterpart, such as EMRIs, SOBHBs, and IMBHBs. Although the cosmological potential of these sources is limited to low and intermediate redshift by the completeness of galaxy catalogues, they will anyway yield interesting constraints on \( H_0 \), reaching the percent level in the most optimistic scenarios (Del Pozzo et al. 2018; Kyutoku and Seto 2017; MacLeod and Hogan 2008).

As an alternative to the use of a galaxy catalogue, the redshift information can be obtained from the fact that the GW waveform gives the masses of the binaries redshifted by cosmic expansion. Information about the mass distribution of GW sources can then break this mass-redshift degeneracy and allow a test of the luminosity distance-redshift relation. For stellar-mass BHs, this can be done by exploiting the presence of a mass scale in the population (Farr et al. 2019; Ezquiaga and Holz 2021; You et al. 2021). For binary NSs, it is possible to jointly constrain the parameters of the NS mass function and the cosmological parameters (Taylor and Gair 2012; Taylor et al. 2012).

5.2.2 Luminosity distance

In GR, the observed gravitational strain is inversely proportional to the luminosity distance to the GW source, see eqs. (6) and (7). This can be understood by considering the free propagation of GWs, which in Fourier space is governed by the equation

\[
\mathcal{h}'_\lambda(k, \eta) + 2\mathcal{H}\mathcal{h}'_\lambda(k, \eta) + k^2\mathcal{h}_\lambda(k, \eta) = 0,
\]

where \( \lambda = \pm, \times \) is the polarization index (see above), a prime denotes the derivative with respect to conformal time \( \eta \equiv \int \frac{dt}{a(t)} \) and \( \mathcal{H} \equiv \dot{a}/a \) is the conformal Hubble parameter. This equation can be rewritten by defining \( \chi_\lambda \equiv a\mathcal{h}_\lambda \), so that it becomes \( \chi''_\lambda + (k^2 - a''/a)\chi_\lambda = 0 \). For modes well inside the horizon, such as those considered for LISA, the term \( a''/a \) can be neglected and we obtain a frictionless wave equation for \( \chi \). For a spherical wave, the amplitude evolves as \( \chi_\lambda \propto 1/r \), where \( r \) denotes the comoving distance to the source, so that \( h_\lambda \propto 1/(ar) = 1/d_L \).

In alternative gravity models, however, the strain is inversely proportional to a “GW luminosity distance” which is generally different from the one measured with EM signals. To distinguish between the two, in the following we denote the former by \( d_L^{GW} \) and the latter by \( d_L^{EM} \). Indeed, in non-GR theories the propagation equation (8) is generally modified. Here we consider only scale-independent modifications (see Sects. 2.2.2 and 6.3.2 for scale-dependent ones) and thus one can have (Amendola et al. 2018a; Arai and Nishizawa 2018; Belgacem et al. 2018a, c; Gleyzes et al. 2014; Lombriser and Taylor 2016; Nishizawa 2018; Saltas et al. 2014)

\[
\mathcal{h}'_\lambda(k, \eta) + \mathcal{H}[1 - \delta(\eta)]\mathcal{h}'_\lambda(k, \eta) + k^2\mathcal{h}_\lambda(k, \eta) = 0,
\]

for some function \( \delta(\eta) \). In this case, to eliminate from the equation the term

---

8 Strictly speaking, interferometers measure the strain of GWs, not their luminosity, although the two are related. In modified gravity this relation is generally different from the GR one.

---
proportional to $h'_i$, we may introduce $\chi_i \equiv \overline{a} h_i$ with $\overline{a}' / \overline{a} = \mathcal{H} [1 - \delta(\eta)]$. As before, $\chi_i \propto 1/r$ for a spherical wave and hence $h_i \propto 1/(\overline{a} \rho r)$. In other words, the GW luminosity distance is now related to the EM one by

$$d_L^{gw}(z) = d_L^{em}(z) \exp \left\{ - \int_0^z \frac{dz'}{1 + z'} \delta(z') \right\}. \quad (10)$$

Non-trivial background cosmological solutions such as self-accelerating ones (Crisostomi and Koyama 2018a; Lombriser and Lima 2017; Lombriser and Taylor 2016) are of particular interest since they are responsible for the peculiar behavior of the friction term, which can change the ratio $d_L^{gw}/d_L^{em}$. For other works studying the damping of GWs see e.g., Deffayet and Menou (2007); Amendola et al. (2018a); Calabrese et al. (2016); Frusciante (2021); Pardo et al. (2018); Visinelli et al. (2018). A deviation in the friction term has potential signature in LISA (Belgacem et al. 2019c) or third-generation ground based standard sirens observations. GWs detectors are thus a complementary means with other cosmological surveys for constraining the friction term.

In general, the reconstruction of a full function of redshift from the data is challenging and it is convenient to have a useful parametrization (analogous, for instance, to the $(w_0, w_a)$ parametrization of the dark energy equation of state). For modified GW propagation, a convenient parametrization can be given in terms of two parameters $(\Xi_0, n)$ (Belgacem et al. 2018a),

$$\frac{d_L^{gw}(z)}{d_L^{em}(z)} = \Xi_0 + \frac{1 - \Xi_0}{(1 + z)^n}. \quad (11)$$

This parametrization reproduces the fact that, as $z \to 0$, $d_L^{gw}/d_L^{em} \to 1$ since, as the redshift of the source goes to zero, there can be no effect from modified propagation. In the limit of large redshifts, in (11) $d_L^{gw}/d_L^{em}$ goes to a constant value $\Xi_0$. This is motivated by the fact that in typical DE models associated with the cosmic acceleration the deviations from GR only appear in the recent cosmological epoch, so $\delta(z)$ goes to zero at large redshift and, from (10), $d_L^{gw}(z)/d_L^{em}(z)$ saturates to a constant.\footnote{In a recent LISA Cosmology WG paper (Belgacem et al. 2019c), the predictions of some of the best-studied modified gravity models (such as several examples of Horndeski and DHOST theories, non-local infrared modifications of gravity, or bigravity theories) have been worked out, and it has been found that all these models predict modified GW propagation of the form (9), and the corresponding ratio $d_L^{gw}(z)/d_L^{em}(z)$ is well reproduced by the functional form (11), except in bigravity, where the interaction between the two metric leads to oscillations in $d_L^{gw}(z)/d_L^{em}(z)$.}

For Horndeski and beyond-Horndeski theories the above discussion simplifies. In this case the damping in Eq. (9) comes from the time dependence of the effective Planck mass $M_*$ that canonically normalizes the graviton (Gleyzes et al. 2013). In particular, the quadratic action for $h_{ik}$ reads, using the conformal time, $S \propto \int d^3x d\eta (aM_*)^2 \left[ h_{ik}^2 - (\partial_k h_{ik})^2 \right]$, which shows that in this case the quantity that satisfies a standard wave equation inside the horizon is $\chi \equiv aM_* h_{ik}$, with an extra $M_*(\eta)$ in the normalization. This implies that
\[
\frac{d_{\text{gw}}^L(z)}{d_{\text{em}}^L(z)} = \frac{M_e(0)}{M_e(z)},
\]

i.e., the ratio between the GW and EM luminosity distance simply reflects the ratio between the value of the effective Planck mass at the observer (at redshift 0) and the one at the source at redshift \( z \). If these are equal, for instance as a consequence of other constraints (see Sect. 5.1.4 above), then there is no effect (Dalang and Lombriser 2019; Dalang et al. 2020). Another way to see this is by noting that the damping term in Eq. (9) is \( \delta = -d \ln M_e/d \ln a \) (see e.g., Bellini and Sawicki 2014; Gleyzes et al. 2014); Eq. (12) then follows from Eq. (10).

In the past, studies of dark energy at the next generation of GW detectors have focused on the DE equation of state. However, while current EM observations show that in modified gravity theories deviations from the \( \Lambda \)CDM cannot be much larger than a few percent, modified GW propagation is a phenomenon accessible only to GW detectors, and therefore currently basically unconstrained. For instance, there are explicit examples of phenomenologically viable modified gravity models that predict a very large effect. In particular, the RT non-local model mentioned above predicts a value of \( \Xi_0 \) that, depending on a free parameter of the model (related to the initial conditions) can be as large as \( \Xi_0 \approx 1.8 \) (Belgacem et al. 2018b, 2019a), corresponding to a 80% deviation from the GR value \( \Xi_0 = 1 \). Therefore, as observed in Belgacem et al. (2018a, 2018c) modified GW propagation provides a complementary and more promising observable than those inferred from EM observations.

Constraints on a modified GW luminosity distance were placed after the multimessenger event GW170817 (Arai and Nishizawa 2018; Belgacem et al. 2018a; Lagos et al. 2019). The first limits on \( \Xi_0 \), using dark sirens from the O1-O2-O3a LIGO/Virgo runs have been recently presented in Finke et al. (2021), while constraints using the binary black hole mass distribution were obtained in Ezquiaga (2021). Forecasts on the sensitivity of LISA to \( \Xi_0 \) (together with other parameters, such as the parameter \( w_0 \) that enters in the parametrization of the DE equation of state) have been performed in Belgacem et al. (2019c), using the coalescence of SMBH binaries. These are expected to produce powerful EM counterparts, since they are believed to merge in a gas rich environment, and are therefore potential standard candles for LISA. The results, using the astrophysical models of Barausse (2012); Klein et al. (2016), which make a range of plausible assumptions for the BH seeds and for the merger delay times, and performing full MCMC exploration of the parameter spaces, confirm the strong potential of LISA for constraining, or detecting, modified GW propagation. As an example, Fig. 3 shows the forecast in the \( (\Xi_0, w_0) \) plane, for two different astrophysical scenarios.

To end this discussion, let us remark that in theories beyond GR that feature additional degrees of freedom, be it scalar, vector, or tensor, GWs are expected to exhibit new polarisations. In general, these additional polarisations will interact with the standard ones and such interactions may affect the amplitudes of the standard polarization modes, whose energy could leak into the new ones, thereby changing the measured gravitational distance \( d^\text{gw}_L \). Over cosmological backgrounds GW mixings only appear when additional tensor modes are present (Caldwell et al. 2016; Jiménez...
et al. 2020; Max et al. 2017) or cubic interactions are taken into account (Creminelli et al. 2018, 2019). Over general space-times, quadratic mixing with scalars and vector modes are also possible (Dalang et al. 2021; Ezquiaga and Zumalacárregui 2020) (see also discussion in Sect. 2.2.2).

5.2.3 Hubble constant

The present value of the Hubble parameter $H_0$ is one of the fundamental parameters of the ΛCDM model. Over the last few years, a disagreement has persisted between local measurements obtained from the late-time distance ladder (such as the type Ia supernovae calibrated by cepheids, Riess et al. 2018) and the value inferred from early-time observables such as the CMB (Planck Collaboration et al. 2020b) and the large-scale structure (D’Amico et al. 2021; Philcox et al. 2020) (see e.g., Shah et al. 2021 for a recent review on the Hubble constant tension). In the next decade, GWs will play a crucial role in constraining $H_0$. The luminosity distance of a BBH and a BNS can be inferred and calibrated using an empirically constructed distance ladder at various distance scales. If combined with the known redshift of the host galaxy, the present rate of $H_0$ can then be estimated (Schutz 1986), as it is done using type Ia supernovae via the local distance ladder (Riess et al. 2016).

The first standard siren-based $H_0$ measurement came from the neutron-star merger event, GW170817, with the optical counterpart identification of the host galaxy, NGC 4993. While currently the observations of this merger set a loose constraint of $H_0 = 70^{+12}_{-8}$ at 1σ (Abbott et al. 2017a), future detections of additional GW standard sirens, obtained with ground-based second-generation detectors at full sensitivity, have the power to constrain the Hubble parameter to an accuracy of $\sim 1\%$ with an error that depends on the number of detected events (Chen et al. 2018; Nissanke et al. 2013) and redshift distribution (Dalal et al. 2006). Notably, for nearby sources, the

![Fig. 3](image-url)
dominant uncertainty comes from the error on the peculiar velocities when correcting
the measurement of the observed redshift (Boruah et al. 2021; Nicolaou et al. 2020). A recent work by the LISA Cosmology WG (Belgacem et al. 2019c) forecast $H_0$ within the $\Lambda$CDM scenario using MBH and found the relative error to be 3.8% in the optimistic case for the accuracy of redshift measurement and delensing and 7.7% in the more realistic case (see also Tamanini et al. 2016).

Furthermore, the rate of dark GW events without detection of EM counterpart, such as EMRIs, SOBHBs, or IMBHBs, should be much higher than those with EM counterpart (Lyutikov 2016). To these dark sirens are also added all events whose EM counterpart will be difficult to identify, such as possible well-localised SMBHs at relatively low-redshift (Petiteau et al. 2011). The strategy adopted by GW detectors then consists in estimating the Hubble parameter for each potential host galaxy of the GW, resulting in an average value of $H_0$ (Chen et al. 2018; Fishbach et al. 2019; Schutz 1986), as done for the GW170814 event (Soares-Santos et al. 2019). In this context, the accurate sky location provided by LISA for the identification of the host galaxy of the GW event is definitely a major advantage (Kyutoku and Seto 2017) and the use of redshift catalogs from large-scale spectroscopic galaxy surveys such as the Dark Energy Spectroscopic Instrument (DESI) (Aghamousa et al. 2016), 4MOST (de Jong et al. 2012) or Euclid (Amendola et al. 2018b) will therefore be required to improve the Hubble constant measurement. Notably, LISA should be able to observe up to $\sim 100$ stellar-mass BBHs within 100 Mpc, providing a few % measurement of $H_0$ (Del Pozzo et al. 2018; Kyutoku and Seto 2016). This number however strongly depends on the performance of LISA at the high end of its frequency band, which is not guaranteed to provide the required sensitivity (Moore et al. 2019). A more promising population of dark sirens may be provided by EMRIs, which depending on the actual rate at which they will be detected, will yield constraints on $H_0$ at the few to one % level (Laghi et al. 2021).

The current forecasts mentioned here above do not indicate that LISA will be able to contribute to solving the Hubble tension, which by the 2030s is expected to be largely solved by other observations. However they show that LISA will provide strong complementary constraints on $H_0$, which will be used to corroborate results obtained with other measurements and different techniques. This kind of cross-validation will be extremely important to reinforce our confidence on the actual value of the Hubble constant, especially if hints of deviations from $\Lambda$CDM will have been observed.

5.3 Large-scale structure

The launch of LISA will be preceded by several large-scale structure surveys, such as DESI and the Vera C. Rubin Observatory from ground, Euclid and Roman from space, and SKA in the radio band. These instruments will map a large part of the observable universe with unprecedented accuracy. Here we discuss how LISA observations can be combined with these surveys’ data and how the intervening large-scale structure will affect the observed GW signals.
5.3.1 Cross-correlation with large-scale structure

Another possible way to test cosmological models is represented by the cross correlation of GW detections with the large-scale structure (LSS) of the Universe. This was pioneered initially to test dark energy and dark matter models and the distance-redshift relation in Camera and Nishizawa (2013); Oguri (2016); Raccanelli et al. (2016). More recently, this probe became very popular and several different approaches have been suggested and investigated (see e.g., Calore et al. 2020; Cañas Herrera et al. 2020; Raccanelli 2017; Scelfo et al. 2018), looking at both correlations with resolved sources and the SGWB (Alonso et al. 2020; Bertacca et al. 2020; Cusin et al. 2018; Jenkins and Sakellariadou 2018; Jenkins et al. 2019; Mukherjee et al. 2020b; Yang et al. 2021).

For the resolved-event case, the GW/C2/LSS correlation can be used for a variety of tests. By measuring the effective bias of the hosts of compact-object mergers, one can get information about the presence and abundance of PBHs. This happens because binary PBHs would preferentially merge in halos with low velocity dispersion, and hence halos with very little or absent star formation, which have a galaxy bias \( b < 1 \). On the other hand, mergers of compact objects that are the endpoint of stellar evolution naturally happen for the vast majority within star forming-rich halos, which have larger galaxy bias values, \( b > 1 \), where in the linear bias scheme, \( b \) quantifies the mismatch between the distribution of matter and that of galaxies, \( \delta_c = b \delta_m \).

Therefore, by measuring the amplitude of the angular cross correlation of galaxy maps with catalogues of compact-object mergers, which directly depends on the bias of the mergers’ hosts, one can statistically probe the abundance of PBH mergers. This information can also be used to discriminate between different astrophysical models (see e.g., Scelfo et al. 2020)

When EM counterparts are available (or the redshift determination of BBHs is available through other methods), the GW/C2/LSS correlation can be used to test the distance-redshift relation (Oguri 2016) and models of dark energy and modified gravity (Camera and Nishizawa 2013; Mukherjee et al. 2020a; Raccanelli 2017; Raccanelli et al. 2018).

In the case of the SGWB, the cross correlation with the LSS will be fundamental to understand the observed signal. GWs will experience projection effects that will need to be accounted for in order to observe the signal in the appropriate frame (Bellomo et al. 2021; Bertacca et al. 2020). Moreover, the cross correlation with the LSS can allow us to disentangle the astrophysical SGWB from the primordial one coming from inflation.

5.3.2 Gravitational lensing of GWs

Similarly to EM radiation, GWs experience the gravitational potential of massive objects while traveling across the Universe. This opens the possibility of both probing the distribution of structures in the cosmos and the underlying gravitational interactions. LISA will offer a unique perspective since it will detect high-redshift
GWs in a lower frequency band compared to ground-based detectors, increasing the lensing probabilities and the detectability of diffraction effects.

The probability that a GW is lensed depends on the redshift of the source and the distribution of lenses. For high-redshift GWs, intervening matter between the source and the observer can magnify the GW signal and introduce a systematic error in the determination of the luminosity distance. Correcting for this weak-lensing effect will be relevant when inferring cosmological and population parameters from GW standard sirens (Cusin and Tamanini 2021; Fleury et al. 2017; Hirata et al. 2010; Hughes and Holz 2003). A fraction of these GW events will pass close enough to the lens so that lensing will produce multiple images of the same event. LISA could detect a few strongly-lensed massive black-hole binaries during its mission (Sereno et al. 2010). The detection of multiple lensing events could provide new means for cosmography with LISA (Sereno et al. 2011).

In the strong-lensing regime the time delay between the images increases linearly with the mass of the lens, reaching delays of months for a $10^{12} M_\odot$ galaxy lens. Each of the lensed images will have a different magnification and they will acquire a fixed phase shift depending on how many times the image has crossed a lens caustic (Schneider et al. 1992). The precise measurement by LISA of the phase of long duration signals could be key in distinguishing different types of lensed images (Ezquiaga et al. 2021a).

When time delays between lensed images become shorter than the signal’s duration, interference effects among the images become relevant. This wave-optics regime is achieved when the wavelength of the GW is comparable to the Schwarzschild radius of the lens. Since LISA will observe lower frequencies, diffraction effects will be more relevant than for ground-based detectors (Takahashi and Nakamura 2003). In the diffraction limit the GW waveform can be distorted and one could use these features to identify the event as lensed (Dai et al. 2018). In addition, these distortions may make the GW phase to appear to arrive earlier than an unlensed EM signal (Morita and Soda 2019; Takahashi 2017), although this is only an apparent superluminality (Ezquiaga et al. 2020b; Suyama 2020). This effect has to be taken into account (Ezquiaga et al. 2020b) when for example inferring constraints on the speed of gravity from the possible pre-merger modulated EM brightness of a super-massive BBH (Haiman 2017). If these lensing features distorting the GW waveform are not properly identified, they could be misinterpreted as modified gravity, see e.g., Cusin and Lagos (2020); Ezquiaga et al. (2021a, 2021b); Toubiana et al. (2021).

### 5.4 Burning questions

We have reviewed several ways in which LISA will contribute to test the $\Lambda$CDM model, and to probe the fundamental physics nature of its extensions involving dark energy and modified gravity. There are many open directions.

- Screening is likely to play an important role, yet largely unexplored. In particular, it would be relevant to understand what happens when a GW propagates from, to, and through screened regions of the Universe. The considerations of Sect. 5.2.2
were made in the context of an idealised homogeneous and isotropic FLRW background space-time and in the absence of extra polarization modes. It is thus natural to wonder which of its conclusions hold as GWs propagate through the actual inhomogeneous Universe or in the presence of extra modes. Theoretical analyses along these lines were initiated in Dalang and Lombriser (2019); Dalang et al. (2020); Lagos and Zhu (2020); Wolf and Lagos (2020) for Horndeski theories.

- Another direction is to assess how other standard candles such as SMBH binaries and EMRIs will improve, using statistical methods, our measurements of $H_0$, of the equation of state of dark energy and of the effects on the GW propagation.

- Moreover, it would be relevant to explore, in terms of forecasts, how the synergy between LISA and large-scale structure surveys will help to improve the redshift estimate of GW sources and the inferred cosmological parameters.

- Another possible synergy (Piro et al. 2021) is with the Athena observatory (Nandra et al. 2013). Measuring the X-ray signal and optical follow-up will make it possible to identify the host of an individual GW source. The distance-redshift relationship can then be measured to high accuracy (1%) to probe the Hubble rate.

- Finally, we have discussed several effects of lensing of the GWs but their potential to constrain dark energy and modified gravity remains largely unexplored.

6 Model-independent tests

Although one can test a given model of gravity by comparing its predicted GW templates against data, a different approach is to perform model-agnostic tests. There are two main types of such tests. The first one is a consistency test of GR, where assuming that GR is correct one compares the GR waveform with GW data. The second one aims at constraining modified theories of gravity by performing a certain model-independent test and then mapping the outcome to specific modified gravity models. We will discuss these tests in some detail in what follows.

6.1 Consistency tests of GR

Let us first discuss tests to check the consistency of GR predictions with the detected GW signals. We begin with the residual test, which is the most generic, model-independent test that could, in principle, pick up arbitrary departures of the theory from the data. We will then describe a test looking for consistency in the intrinsic parameters of the source (e.g. the merger remnant’s mass and spin) as determined from the early inspiral phase and the late IMR phase of the coalescence event. Finally, we discuss a generalization of the BH no-hair test using the entire IMR waveform, wherein one compares the intrinsic parameters of the source determined from the different multipole modes of the signal.

It is worth noting that consistency tests are also possible by comparing the parameters obtained from the GW signal with those from other techniques in a multi-
messenger scenario (Baker et al. 2019). There are already systems observed in GW detectors (with LIGO/Virgo) and EM telescopes (Abbott et al. 2017d). This will become much more prevalent with LISA (Chen et al. 2020; Korol et al. 2017; McGee et al. 2020; Mukherjee et al. 2020a; Wyithe and Loeb 2003). In particular, joint observations of GW systems (Congedo and Taylor 2019; Edwards et al. 2020; McGee et al. 2020), will provide independent estimates of the BH parameters assuming GR, which can be compared for consistency.

6.1.1 Residual tests

A Bayesian inference algorithm can be used to estimate the parameters of a signal/source in the data (Abbott et al. 2020a). Such an algorithm would essentially maximize the likelihood $P(d|h)$, where $d$ is the data and $h$ is the expected signal. The parameters $\lambda$, that maximize the likelihood, are deemed to be the best estimate of the signal parameters. If GR gives the correct description of GWs, then the data would be consistent with the waveform predicted by GR. In that case the residual, constructed by subtracting the best-fit waveform from the data, would be consistent with background noise. Failure of GR to accurately describe the data would lead to a residual that is statistically inconsistent with background noise. Such tests have been applied to the existing GW events by the LIGO/Virgo Collaboration (Abbott et al. 2016b, 2019b).

For the test to be effective, it is critical to characterize the statistical properties of the residual. The data from GW detectors is often contaminated by non-stationary and non-Gaussian background. Thus, one cannot simply ask if the residual is consistent with a Gaussian distribution but one must ask if it is consistent with the detector noise at times when no GW signals are known to be present. One can deploy statistical tools such as Anderson-Darling or Kolmogorov–Smirnov tests to compare the residual with the data at other times. Alternatively, one can use a transient detection algorithm, e.g. Bayeswave (Cornish and Littenberg 2015), to estimate the coherent SNR in the residual data at a certain statistical significance and ask what is the probability that one gets an SNR as large as for data sets that contain only noise.

By construction, the residual test is the most generic model-independent method that one can construct. This is because, the test does not require any non-GR waveforms, does not use additional parameters in the waveform to look for deviations (unlike, e.g., in parameterized tests of GR described in Sect. 6.2 below) and is sensitive to departures from GR outside the region where GR waveform has most of its support, (e.g. echoes). Its drawback, however, is that subtle departures from GR cannot be easily identified since the test is not phase coherent and the purely statistical nature will make it difficult to identify the origin of the failure of the theory.

More work is needed to determine how to implement such a residual test in LISA. An approach similar to LIGO/Virgo’s may not work because, unlike LIGO/Virgo currently, LISA will always have a background of sources that are on during the entire observation period. Therefore, removing only the loudest signal from the data will not necessarily lead to a residual that is consistent with noise, i.e. there could be other signals hidden below the loudest one. Work has began to determine how to deal
with multiple simultaneous sources in the data for parameter estimation reasons, but similar studies should be undertaken to generalize LIGO/Virgo’s residual tests to LISA.

### 6.1.2 Inspiral-merger-ringdown (IMR) tests

The orbital dynamics of a BBH essentially consists of three phases: (i) a long *adiabatic inspiral phase* during which the companion objects slowly spiral in towards each other as they lose energy to GWs, (ii) a *rapid plunge-merger phase* during which the system cannot be described by orbital dynamics as the horizons of the two BHs merge to form a single horizon, and (iii) a final *ringdown phase* during which the merger remnant quickly settles down to its final stationary state. An IMR consistency test compares the parameters determined using the inspiral phase of the signal only to those obtained using the late-time merger-ringdown phase (Ghosh et al. 2018, 2016b). If GR correctly describes both the inspiral and the plunge-merger regimes then the parameters determined from these two phases would be consistent with each other, within statistical uncertainties associated with the measurement; any discrepancy would then be a hint of a failure of the theory.

The IMR test currently implemented by the LIGO/Virgo Collaboration (Abbott et al. 2019b, 2016b) translates the intrinsic parameters (the masses and spins) measured in the inspiral phase to the mass $M_f$ and the dimensionless spin magnitude $\chi_f$ of the merger remnant using a fitting formula derived from NR simulations. One then plots contours of the (two-dimensional) posterior probability density in the $M_f$-$\chi_f$ plane at different credible intervals. An IMR test therefore checks if the 90% credible contours corresponding to the two different ways of measuring the parameters overlap with each other. Non-overlapping contours could indicate that one of the two phases is not consistent with the GR model. Large BH spins and small mass ratios are potentially more effective in such tests as they break the symmetries that could hide failure of GR in equal-mass systems with non-spinning BH companions. The IMR consistency test will strengthen its power when inspiral observations with LISA is combined with merger-ringdown observations with ground-based detectors through multiband GW observations (Carson and Yagi 2020a, b; Tso et al. 2019).

One can go beyond the consistency check of GR and apply the IMR test to probe specific non-GR theories and spacetimes. Unlike the parameterized tests to be discussed in Sect. 6.2, there is no simple known mapping to convert the information of the theory-agnostic IMR consistency test to bounds on specific theories or spacetimes, and hence the application needs to be done on a case-by-case basis. This has been demonstrated for a string-inspired quadratic gravity theory (Carson and Yagi 2020d) and generic BH spacetimes beyond Kerr (Carson and Yagi 2020c). The IMR tests with multiband GW observations will be able to probe such a theory and spacetimes more accurately than the existing GW and EM observations by a few orders of magnitude.

One problem that must be overcome to carry out such IMR tests with LISA is to develop a more thorough understanding of the mass and spin of the remnant in...
mergers of intermediate mass ratio systems and systems with non-negligible eccentricity and double-spin precession. Indeed, unlike LIGO/Virgo sources, many LISA binary sources are expected to be eccentric and spin-precessing. Waveform models for such systems are beginning to be developed, but detailed numerical simulations are currently lacking. We refer the interested reader to the waveform modeling white paper (in prep.) for further details.

6.1.3 Multipolar gravitational wave tests

GWs at leading-order in a post-Minkowskian expansion is quadrupolar in GR. However, non-quadrupole (higher order) modes make appreciable contribution to the GWs from BBHs with large mass ratios and misaligned spins—which was indeed the case for GW190412 (Abbott et al. 2020b) and GW190814 (Abbott et al. 2020d). The multipolar structure of the GWs is fully determined by the intrinsic parameters of a binary, e.g. masses and spin angular momenta of the companion BHs in a quasi-circular orbit. One can formulate multiple ways of testing the consistency of the observed GW signal with the expected multipolar structure of GWs from BBHs in GR. This is called a “no-hair” test of BBHs (Dhanpal et al. 2019; Islam et al. 2020) as it is similar to testing the “no-hair” theorems for isolated BHs through mutual consistency of the quasi-normal mode spectrum.

The main idea of this test is to look for consistency of the source parameters determined independently from the quadrupole and higher order modes. We typically use GR waveforms when performing consistency tests, so we do not consider lower-order multipole modes, such as scalar dipole modes discussed in Sect. 2.2.3. In spirit, this idea is similar to checking the consistency of cosmological parameters estimated from the low- and high multipoles of the CMB radiation. There are multiple ways in which the test could be implemented. For example, one could use different intrinsic parameters to describe the phase evolution of the quadrupole and higher-order modes and, as in the IMR consistency test described earlier, check to see if the parameters determined from these modes are consistent with one another. Alternatively, one could introduce additional parameters to describe the amplitude of the higher modes while keeping the amplitude of the quadrupole mode as in GR. Such tests could reveal any modification from GR of the multipolar structure of the binary that is imprinted in the GWs observed at detectors.

An advance that would be useful in this context is the development of a mapping between such generic higher-mode deviations from GR and specific modified theories. Most studies of the dynamics of binary systems outside of GR have been only done to leading PN order in the inspiral. Going beyond leading order would be essential to find the higher-modes discussed above as predicted in a given theory.

6.2 Parametrized tests

We now discuss model-independent tests that introduce generic parameters capturing non-GR effects independent of the underlying gravitational theory. We first present a generic framework, followed by a few examples on how one can use such
parameterized tests to probe specific non-GR theories and BH spacetimes beyond Kerr.

Many modified gravity theories have been proposed over the years, and sufficiently accurate gravitational waveform models are only available for a small subset of these theories, when their calculation is even possible at all (Berti et al. 2015). Therefore it is appealing to develop generic “null” tests of Einstein’s theory, following the approach pursued with Solar System tests, where the parametrized PN (PPN) framework proposed by Will and Nordtvedt (Nordtvedt 1968; Nordtvedt and Will 1972; Will 1971; Will and Nordtvedt 1972) has provided a unifying scheme for tests of GR that has been in use for over 50 years.

### 6.2.1 Inspiral tests

In GW data analysis, a natural generalization of the PPN approach consists of verifying the PN structure of the waveform phase (Arun et al. 2006b). The idea is to decompose the Fourier-domain waveform model into a frequency-dependent amplitude and a frequency-dependent phase, and to then rewrite the phase (schematically, and ignoring logarithmic terms) as $\Psi_{\text{GR}}(f) = \sum_{n=0}^{7} a_n v(f)^{5+n}$. In GR the PN coefficients $a_n$ are known functions of the parameters of the binary (the individual masses $m_1$ and $m_2$ for non-spinning BH binaries of total mass $m = m_1 + m_2$), and $v(f) = (\pi mf)^{1/3}$ is the orbital velocity. The idea is to treat all of these coefficients (or just a subset; Arun et al. 2006a) as independent, and find their best-fit values by comparing the above template waveform with the data. One can then check the consistency of the measured masses from each of the above coefficients. This procedure resembles binary pulsar tests in the parameterized post-Keplerian formalism (Damour and Taylor 1992; Stairs 2003), but it has some limitations. There are known modified theories of gravity for which the Fourier phase does not have a leading-order term $\propto v^{-5}$; these include, for example, theories with dipole emission ($\propto v^{-7}$) and variability of the fundamental constants like the gravitational constant $G$ ($\propto v^{-13}$). Second, some modified gravity theories may modify the GW amplitude more than the phase: one example is gravitational birefringence (Alexander et al. 2008; Yagi and Yang 2018; Yunes et al. 2010).

One proposal to address these problems is the so-called ppE approach (Yunes and Pretorius 2009b). In this framework, one extends the GR waveform model as follows:

$$\tilde{h}(f) = \tilde{A}_{\text{GR}}(f) \left[ 1 + \alpha_{\text{ppE}} v(f)^a \right] e^{\Psi_{\text{GR}}(f) + i\beta_{\text{ppE}} v(f)^b}.$$  \hspace{1cm} (13)

Here $\tilde{A}_{\text{GR}}(f)$ and $\Psi_{\text{GR}}(f)$ represent the most accurate GR models for the Fourier amplitude and phase. The quantities $(\alpha_{\text{ppE}}, \beta_{\text{ppE}})$ are ppE constants that control the magnitude of deviations from GR, while $(a, b)$ are real numbers that determine the type of deviation that is being constrained. The ppE approach is similar in spirit to the parametric analysis performed by the LIGO/Virgo Collaboration (Abbott et al. 2019a, 2016b). Different variants and extensions of this idea have been proposed (Cornish et al. 2011; Chatziioannou et al. 2012; Sampson et al. 2013, 2014).
Constraints on ppE parameters can be mapped to constraints on specific extensions of GR (Yunes and Siemens 2013; Yunes et al. 2016). Astrophysical bounds on ppE parameters have also been derived using observations of relativistic binary pulsars (Yunes and Hughes 2010), and parametrized tests using the ppE formalism (in particular the amplitude corrections) can also be used to detect deviations from GR in the GW background emitted by BBHs (Maselli et al. 2016; Saffer and Yagi 2020).

In most data analysis applications (including recent tests of GR by the LIGO/Virgo Collaboration, Abbott et al. 2019a, 2016b), one considers variations of a single ppE term. This is often convenient, and it can generally be accurate enough because higher-order terms will not significantly alter the bounds coming from leading-order corrections (see e.g., Chatziioannou et al. 2012; Yunes et al. 2016). However we can expect modifications appearing at different orders to be correlated, and in general it is important to take into account these correlations, because in any specific theory, all modifications of GR will depend on a single coupling constant (or on a finite number of coupling constants).

These parametrized tests of inspiral waveforms can be thought of as generic null tests of GR, although they can be translated into bounds on specific beyond-GR theories on a case-by-case basis (see Sect. 6.2.3). Interestingly, they can be applied to several source classes, including SOBHs, EMRIs and SMBHs. Each source class has distinct advantages and disadvantages: SMBHs are expected to have large SNR, EMRIs should compensate for their lower SNR with the large number ($\sim 10^5$) of observable inspiral cycles, and—as we discuss below—SOBHs can be used to test GR modifications at large negative PN orders through multiband observations in conjunction with 3G detectors (such as the Einstein Telescope and Cosmic Explorer) that should be operational by the time LISA flies. Each source class can produce tight constraints on different ppE modifications (and therefore on different modifications of GR), depending on astrophysical event rates and uncertain details on the future network of ground-based interferometers. One of the main challenges facing theorists, astrophysicists and experimentalists is to figure out how this complex interplay between astrophysical event rates, beyond-GR waveform models, ground-based detector developments and data analysis methods will affect our ability to do fundamental physics with LISA.

Multiband observations of stellar-mass BH binaries can play an important role in constraining modified gravity via parametrized tests. This is because “multibanding” allows us to combine the information from the early inspiral dynamics using LISA with late inspiral, merger and ringdown observations using third-generation ground-based detectors. For example, it has been shown that the bounds on dipole GW emission from multibanding can be several orders of magnitude better than bounds coming from either of the bands (Barausse et al. 2016; Perkins et al. 2021; Toubiana et al. 2020). Single-parameter tests at positive PN orders are also expected to improve tremendously, as shown in Carson and Yagi (2020a, 2020b); Gnocchi et al. (2019). The impact of multibanding on multiparameter tests is equally profound: Gupta et al. (2020) shows that multibanding observations may be the only way of carrying out accurate multiparameter tests of GR, where simultaneous measurement of most of the known PN coefficients could be achieved with percentage-level precision.
One important challenge of these parameterized inspiral tests is to extend the mapping between generic deviations and specific theories; we will present two examples below (in Sect. 6.2.3). Indeed, the more such mappings we possess, the more inferences about fundamental physics we will be able to make from parameterized inspiral tests. Another challenge is to determine how to handle parameterically modifications of GR that do not admit a simple PN expansion in the inspiral. Indeed, certain theories of gravity, such as those that predict the sudden activation of dipole radiation at a specific length scale, cannot be exactly modeled as described. The above parameterization may be approximate enough to still be useful, but this needs further study in the LISA context.

6.2.2 Ringdown tests

Any parametrization requires analytical control of the “unperturbed” GR waveform, therefore it is very difficult (if not impossible) to construct parametrizations of the nonlinear merger of compact binaries. However one can attempt to construct parametrized ringdown waveforms based on perturbative treatments of the ringdown in GR.

Such parametrizations are useful because a detailed, ab-initio description of the BH ringdown phase beyond GR is very challenging, especially for rotating BHs. Despite years of efforts and significant advances in the field (Barausse et al. 2014; Cardoso et al. 2019b; Franciolini et al. 2019; Glampedakis and Pappas 2018; Glampedakis and Silva 2019; McManus et al. 2019; Tattersall et al. 2018), the vast majority of theoretical proposals to parametrize ringdown waveforms are currently plagued by certain approximations and assumptions that make them of limited utility in GW data analysis. Modified ringdown waveforms should be built upon the existence of BH solutions different from those predicted by GR (at least in theories where no-hair theorems do not apply), and in addition they must take into account that the perturbations will have different dynamics even for theories which admit the same solutions as GR (Barausse and Sotiriou 2008; Molina et al. 2010; Tattersall et al. 2018). These restrictions reflect the complexity of the problem, and they are of a varied nature. Some calculations were developed in specific modifications of GR, but most of them assume nonrotating BH backgrounds. Most follow-up works that attempt to relax the nonrotating approximation either deal only with scalar (rather than gravitational) perturbations, use the so-called geometric optics approximation, which is only valid in the eikonal (high-$\ell$) limit and in the absence of coupling between the metric and other degrees of freedom (but see Silva and Glampedakis 2020 for an attempt to include such a coupling with a small BH rotation), or work in the slow-rotation approximation (Wagle et al. 2021).

One of the most difficult open problems is the modelling of rotating BHs beyond GR (see also Sect. 2.2.1). This is important, because NR and recent GW observations indicate that the spin of the merger remnant is generally large (Abbott et al. 2019c; Berti and Volonteri 2008; Berti et al. 2007; Buonanno et al. 2007; Hofmann et al. 2016). Unfortunately most rotating BH solutions beyond GR are only known either perturbatively, as small-spin expansion around non-spinning backgrounds (Ayzenberg and Yunes 2014; Cano and Ruipérez 2019; Cardoso et al. 2018b; Maselli et al. 2014).
or they are given in the form of fully numerical solutions of the field equations (Cunha et al. 2019; Herdeiro et al. 2016; Kleihaus et al. 2011; Sullivan et al. 2020, 2021), which makes the application of perturbation theory difficult.

The “miraculous” separability of the master equations for the perturbations in GR (Teukolsky 1972, 1973) is in general absent in modified theories of gravity, so the calculation of QNMs must rely on fits of time-domain waveforms computed from numerical simulations in selected classes of theories beyond GR, which are now possible (at least in the small-coupling limit and for selected theories; Okounkova 2019; Okounkova et al. 2017, 2019b; Witek et al. 2019) but technically challenging.

Rather than attacking the problem “from the ground up”, various groups have considered a data-analysis-first approach, where one parametrizes deviations in the QNM frequencies and damping times (Carullo et al. 2018; Gossan et al. 2012; Maselli et al. 2020b; Meidam et al. 2014). In this framework, beyond-GR QNMs are described in terms of generic shifts from the corresponding Kerr QNM frequencies. Of course, once we set constraints on these shifts, we still need to map the frequency shifts to the new degrees of freedom of the underlying theory of gravity that produced them.

In conclusion, parametrized ringdown is an active research field which is still in its infancy.

6.2.3 Mapping to non-GR theories

Parametrized tests will show its power when bounds on generic, non-GR parameters can be mapped to those on theoretical constants in specific theories that represent violations of fundamental pillars in GR. In this subsection, we discuss two specific example theories, namely scalar-tensor theories and quadratic gravity. Mappings to other modified theories of gravity for both the phase and amplitude corrections can be found, e.g., in Tables I and II of (Tahura and Yagi 2018).

One of the simple examples to which we can relate the parametrized tests of GR concerns scalar-tensor theories. In these theories the leading modification to GR is the presence of a dipolar radiation. Such a radiation is generated by the (dimensionless) scalar charge $a_A$ of each body that violates the SEP. The parametrized dipolar test on the GW luminosity of $\dot{E}_{\text{GW}} = \dot{E}_{\text{GR}} \left( 1 + B \left( \frac{Gm}{rc} \right)^{-1} \right)$ gives a bound on the parameter $B$ where $m$ is the total mass, $r$ is the binary separation and $B$ for scalar-tensor theories is $B_{\text{ST}} = \frac{5}{96} (x_1 - x_2)^2$ (Barausse et al. 2016). Scalar charges are typically zero for BHs in scalar-tensor theories and $B_{\text{ST}} = 0$ for BBHs unless e.g. the scalar field is coupled to the Gauss–Bonnet invariant or some other forms of higher curvature invariant (see Sect. 2.2.1). However, $B_{\text{ST}} \neq 0$ for EMRIs if the secondary objects are NSs. Currently the best constraint on $B$ is given by the observation of binary pulsars, $|B| \lesssim 10^{-7}$ (Shao et al. 2017; Yunes and Hughes 2010). This can be related with the ppE framework in Eq. (13) with $b_{\text{ST}} = -7/3$ and $\beta = -\frac{3}{224} \eta^{2/5} B$. Such a correction enters at $-1\text{PN}$ order to the waveform phase.
Another important example theory that can be tested within the ppE framework is quadratic gravity that introduces quadratic-curvature corrections (coupled to a scalar field) to the Einstein–Hilbert action. Under the small-coupling approximation, the leading-order modification to the Fourier phase $\Psi(f)$ takes on the ppE form (Yunes and Pretorius 2009b) in Eq. (13), where $b_{dCS} = -1/3$ in dCS gravity (a 2PN correction) and $b_{EdGB} = -7/3$ in Einstein-dilaton Gauss–Bonnet gravity gravity (a $-1\text{PN}$ correction). The specific mapping to these two well-motivated theories can be found in Nair et al. (2019); Yagi et al. (2012a) for EdGB, and in Nair et al. (2019); Yagi et al. (2012b) for dCS. From the data presented in first LIGO/Virgo catalogue, GWTC-1 (Abbott et al. 2019c), EdGB’s coupling constant was constrained to $\alpha_{EdGB}^{1/2} \lesssim 5.6$ km (Nair et al. 2019) from single events and $\alpha_{EdGB}^{1/2} \lesssim 1.7$ km (Perkins et al. 2021; Yamada et al. 2019) from combined multiple events, and the latter is better than previous bounds, such as those from BH low-mass x-ray binaries (Yagi 2012). On the other hand, dCS gravity still remains unconstrained from the leading PN order correction to the waveform phase, although recent consistency constraints from the combination of LIGO/Virgo and NICER data require $\alpha_{dCS}^{1/2} < 8.5$ km (Silva et al. 2021a). Work on other modified theories within the broad class of quadratic gravity models (Yagi et al. 2016) remains to be done.

Apart from testing specific non-GR theories, one can use the parameterized tests to probe beyond-Kerr BH spacetime in a generic way. For example, it is possible to map the ppE constraints from the inspiral regime to constraints on parametric deviations from the GR BH metric (Cardenas-Avendano et al. 2020; Carson and Yagi 2020c). The modified ringdown frequencies can also be used to place constraints on deviations away from the GR BH metric (Konoplya and Zhidenko 2020; Völkel and Barausse 2020). Similar to the IMR consistency tests discussed in Sect. 6.1.2, one can perform consistency checks between the above BH spacetime tests from the inspiral and ringdown. The two sets of constraints can be compared for discrepancies (which might be an indicator of new physics) or improving the overall constraints (Carson and Yagi 2020c).

### 6.3 Other model-independent tests

There are other model-independent tests one can perform on GW observations with LISA (some of them have already been applied to the existing GW events by the LIGO/Virgo Collaboration). In this section, we describe three such model-independent tests. First, one can look for additional GW (scalar and vector) polarization modes that are absent in GR. Next, one can study how the modified propagation of GWs affect the amplitude and phase of gravitational waveforms and how one can use multi-messenger observations or the arrival time difference of GWs at different detectors to probe such modified GW propagation. Lastly, one can use the SGWB of either astrophysical or cosmological (primordial) origin to test GR.
6.3.1 Polarization tests

A model-independent test of GR is based on the possibility of detecting additional polarizations, not present in GR. Metric theories of gravity can have up to six polarization modes—two tensorial, two scalar and two vector modes. While in GR only the tensorial polarization modes are present, this is not the case for most of the viable alternative theories of gravity (Will 2014). Such additional polarizations can be detected with the GW observations of inspiraling compact objects (Chatziioannou et al. 2012; Pang et al. 2020; Takeda et al. 2018, 2019). Indeed, ground based detectors have already put an upper limit on such additional polarization modes in a theory-agnostic way (Abbott et al. 2017b, 2019a, b, 2021b; Takeda et al. 2021), including the case with mixed polarizations of standard tensor modes with additional non-GR ones (Chatziioannou et al. 2021; Takeda et al. 2022).

LISA offers new possibilities for testing the non-GR polarizations with an increased sensitivity (Tinto and da Silva Alves 2010). For frequencies larger than roughly $6 \times 10^{-2}$ Hz, the sensitivity of LISA for scalar-longitudinal and vector polarization modes can be up to ten times larger compared to the tensorial or scalar-transverse modes, while in the lower frequency, the sensitivity is of the same order. Therefore, it is expected that LISA will be able to assess the polarization of the detected GWs and thus put strong constraints on modified gravity.

The SGWBs (to be discussed in more detail in Sect. 6.3.3) can also be a powerful test of non-GR polarizations (Callister et al. 2017; Nishizawa et al. 2009). The sensitivities of LISA to circular-polarization modes in this context were addressed in Seto (2006, 2007); Smith and Caldwell (2017). Such modes can appear for example in modifications of GR that lead to parity violation (Alexander and Yunes 2009).

Similar to the parameterized tests in Sect. 6.2, we now discuss how GW polarization tests can be used to probe specific non-GR theories. Below, we focus on a few examples.

In Horndeski theories, which form the most general class of non-degenerate scalar-tensor models leading to second-order equations of motion (Horndeski 1974), three degrees of freedom may propagate under the form of GWs. The first two are the usual plus and cross tensor polarisations (or equivalently the left- and right-handed helicities). Since the observation of GW170817/GRB 170817 (Abbott et al. 2017c, e), these tensor modes are constrained to propagate at the speed of light (see Sect. 6.3.2), which dramatically reduced the parameter space of viable Horndeski models (Baker et al. 2017; Creminelli and Vernizzi 2017; Ezquiaga and Zumalacárregui 2017; Langlois et al. 2018; Lombriser and Taylor 2016; Sakstein and Jain 2017). Stringent constraints may also be obtained from birefrigence effects in lensed GWs (Ezquiaga and Zumalacárregui 2020). However, even in such reduced Horndeski theories, the third degree of freedom, namely scalar waves, may propagate at subluminal speeds. While the tensor (plus and cross) modes produce shearing tidal forces that are transverse to the direction of propagation, scalar waves generally produce triaxial tidal forces. Specifically, in the rest frame of some inertial observer, there exists an orthonormal basis such that the curvature perturbation caused by a scalar wave reads (Dalang et al. 2021)
\( (\delta R^S_{00ij}) = \begin{pmatrix} \mathcal{R}_1 & 0 & 0 \\ 0 & \mathcal{R}_2 & 0 \\ 0 & 0 & \mathcal{R}_3 \end{pmatrix} \cos \Phi(t) \), \hspace{1cm} (14)

where \( \mathcal{R}_{1,2,3} \) are three amplitudes and \( \Phi(t) \) is the phase of the scalar wave. Note that the third direction does not necessarily coincide with the direction of propagation of the wave. However, for luminal scalar waves, which only happens for conformally coupled quintessence, then \( (\delta R^S_{00ij}) \) reduces to a transverse “breathing” mode, i.e. \( \mathcal{R}_1 = \mathcal{R}_2 \), and \( \mathcal{R}_3 = 0 \) in the direction of propagation.

Lorentz-violating gravity can have both scalar and vector modes. For example, Einstein-Æther theory has two vector modes and one scalar mode (Foster 2006, 2007; Yagi et al. 2014), while HoVision Res.ava gravity has one scalar mode with no vector modes (Blas and Sanctuary 2011; Yagi et al. 2014). GWs from compact binary inspirals have been derived in Hansen et al. (2015); Zhang et al. (2020), including the additional scalar and vector modes.

\( f(T) \) gravity (with \( T \) representing a torsion) is a popular modified gravity theory (Cai et al. 2016; Capozziello and De Laurentis 2011; Ferraro and Fiorini 2007) based on the teleparallel equivalent of general relativity (TEGR), see, e.g., Maluf (2013) or Hayashi and Shirafuji (1979) for New General Relativity, which is a variant of the teleparallel formulation. Some \( f(T) \) gravity violates local Lorentz invariance (Li et al. 2011a). When a certain boundary term \( (B = 2\nabla_{\mu}T^\mu \text{ for a torsion vector } T^\mu) \) is taken into account (Bahamonde et al. 2015) one can formulate a theory called \( f(T, B) \) gravity which naturally links to \( f(R) \) gravity. In the context of GWs these teleparallel theories are interesting models as they contain additional propagating degrees of freedom. While there is some controversy on the number of extra degrees of freedom in \( f(T) \) gravity, see Blagojević and Nester (2020); Ferraro and Guzmán (2018a, 2018b); Li et al. (2011b), there is agreement that the theory possesses at least one and up to three additional degrees of freedom. Consequently, these extra degrees of freedom should manifest themselves as extra polarizations. At first order in perturbation theory, extra polarizations only appear in \( f(T, B) \) gravity and not in \( f(T) \) gravity (Farrugia et al. 2018). Likewise, when considering New General Relativity it was found that up to six possible polarizations may exist, depending on the various parameter choices (Hohmann et al. 2018).

Finally, we comment on probing extra dimensions through polarization tests. Within GR, the structure of the geodesic deviation equation in a general spacetime was first analyzed by Szekeres (1965) who decomposed the Riemann tensor into its canonical components, namely the transverse and traceless component, the longitudinal component, and the Newtonian-type tidal component. This analysis can be extended to a metric theory of gravity in any dimension \( D \) (Podolsky and Svarc 2012). In such a case, the gravitational wave in (macroscopic) higher dimension \( D > 4 \) is naturally identified as the transverse-traceless component with \( D(D - 3)/2 \) independent polarization states. Higher-dimensional gravity also brings new effects. In particular, the presence of extra spatial dimensions into which the GWs could possibly propagate could be (in principle) observable as a violation of the traceless property of the gravitational wave measured by the detector located in our...
\(D = 4\) spacetime. Such an anomalous behavior could possibly serve as a sign of the existence of higher dimensions. Although this effect has already been described in Podolsky and Svarc (2012) and analyzed in the model of exact plane GWs (and their generalizations within the Kundt family of geometries) (Podolský and Švarc 2013), the specific representation of the violation of the transverse-traceless property in the triangular LISA configuration remains an open problem.

The development of polarization tests in the LISA context is still not mature enough for deployment in a data analysis pipeline. Few studies have been performed to determine how well LISA can constrain the extra polarizations of specific modified theories, given current constraints on the speed of tensor modes. The idea of the construction of null channels has been suggested for LIGO/Virgo (Chatziioannou et al. 2012), but can also be extended for LISA in the ppE framework. The mapping of generic polarization tests to specific modified theories has also not yet been developed.

### 6.3.2 GW propagation tests

As discussed in Secs. 2.2.2 and 5.2.2, GW propagation over cosmological distances for generalised theories of gravity and other extensions of \(\Lambda\)CDM can be parametrised by the wave equation (Belgacem et al. 2018c; Gleyzes et al. 2014; Lombriser 2018; Lombriser and Taylor 2016; Nishizawa 2018; Saltas et al. 2014)

\[
h_{ij}'' + \left(3 + \frac{H'}{H} + \nu\right)h_{ij}' + \left[c_T^2 \left(\frac{k}{aH}\right)^2 + \frac{\mu^2}{H^2}\right]h_{ij} = \frac{\Gamma}{H^2} \gamma_{ij}
\]  

(15)

where \(h_{ij}\) is the linear traceless spatial tensor perturbation and primes indicate derivatives with respect to \(\ln a\). In addition to the Hubble friction, the damping term may receive a contribution from \(\nu\), which can for instance parametrise an effective Planck mass evolution rate or the impact of extra dimensions. The wave propagation may further be modified by a deviation in its speed \(c_T\) and the mass of the graviton \(\mu\). Finally, there can be a source term \(\Gamma \gamma_{ij}\), which may for instance arise in bimetric theories or from anisotropic stress. The modifications can generally be time and scale, or frequency, dependent, where GR/\(\Lambda\)CDM is recovered in the limit of \(\nu = \mu = \Gamma = 0\) and \(c_T = c\). The modifications in Eq. (15) can also be cast into the ppE framework (Yunes and Pretorius 2009b) in Sect. 6.2.1, where the ppE amplitude parameter \(\alpha\) can be expressed as an integral over \(\nu\) while the ppE phase parameter \(\beta\) may be expressed as an integral involving \(c_T\) and \(\mu\) (Mirshekari et al. 2012; Nishizawa 2018).

Late-time constraints on the modified damping term have been forecasted in Amendola et al. (2018a); Belgacem et al. (2018c, 2019c); Lombriser and Taylor (2016) based on standard sirens tests (Holz and Hughes 2005; Schutz 1986), and early-time modifications can be constrained by CMB B-modes (Amendola et al. 2014). So far there has not been much exploration of a possible frequency dependence of \(\nu\) (see, however, Belgacem et al. 2019c for forecasts on oscillations in the GW amplitude).
The GW speed $c_T$ is constrained by a variety of measurements. The detection of ultra high energy cosmic rays implies a strong constraint on gravitational Cherenkov radiation from a subluminal propagation of the waves as otherwise the radiation would decay away at a rate proportional to the square of their energy $O(10^{11} \text{ GeV})$ before reaching us (Caves 1980; Kimura and Yamamoto 2012; Moore and Nelson 2001). For galactic $O(10 \text{ kpc})$ or cosmological $O(1 \text{ Gpc})$ origin, the relative deviation in $c_T$ is constrained to be smaller than $O(10^{-15})$ or $O(10^{-19})$, respectively. This bound, however, only applies for subluminal propagation, redshifts of $z \lesssim 0.1$, and modifications in the high-energy regime. Another constraint on $c_T$ at the subpercent level can be inferred from the energy loss in binary pulsar systems (Brax et al. 2016; Beltran Jimenez et al. 2016). A stringent and prominent direct constraint on deviations of $c_T/c = 1$ of $\lesssim O(10^{-15})$ was obtained from the arrival times of the GW from the LIGO/Virgo event GW170817 (Abbott et al. 2017c, e) and its EM counterparts. Photons have been assumed to be massless given the stringent experimental upper bounds on their mass, though we note that some standard-model extensions imply Lorentz-violations that naturally give rise to massive photons (Bonetti et al. 2018; Colladay and Kostelecký 1997, 1998; Spallicci et al. 2021). As anticipated, the measurement left a strong impact across a wide range of cosmic acceleration models (Baker et al. 2017; Battye et al. 2018; Creminelli and Vernizzi 2017; Ezquiaga and Zumalacáregui 2017; Lombriser and Lima 2017; Lombriser and Taylor 2016; McManus et al. 2016; de Rham and Melville 2018; Sakstein and Jain 2017). Importantly, however, the constraint only applies to low redshifts of $z \lesssim 0.01$ and the LIGO/Virgo frequency range (Battye et al. 2018; de Rham and Melville 2018). In particular, it was argued in de Rham and Melville (2018) that UV completion terms for modified gravity theories naturally recover a luminal speed of gravity in the high-energy limit tested by GW170817 while allowing deviations at lower energies relevant to modifications that could drive cosmic acceleration. Weak bounds on $c_T$ can be inferred without counterpart emissions from BH mergers (Cornish et al. 2017), from the comparison of the GW arrival times between the terrestrial detectors (Blas et al. 2016), or the CMB B-mode power spectrum (Raveri et al. 2015) for early-time modifications. Constraints on $c_T$ have also been discussed as forecasts for potential arrival time comparisons with nearby supernovae emissions (Brax et al. 2016; Lombriser and Taylor 2016; Nishizawa and Nakamura 2014; Saltas et al. 2014) (which are however very rare) and LISA eclipsing binary systems (Bettoni et al. 2017).

LISA will provide a twofold improvement over the current GW170817 bound. With the detection of GWs from massive BBHs up to $z \sim 8$ and their EM counterparts (Tamanini et al. 2016), $c_T$ will be tested across much larger distances, tightening the current constraint. Furthermore, LISA tests the frequency range below the expected UV transition of modified gravity models (de Rham and Melville 2018). Besides providing a new test of GR at different energy scales, a measurement of the speed of GWs with LISA is thus of particular relevance to cosmic acceleration models.

Frequency dependent modifications in the velocity term of Eq. (15) can also more generally be parametrised with the group velocity
(\(v_g/c\) \(\simeq\) \(1 + (\alpha - 1)A_2E^{\alpha - 2} + \mathcal{O}(A_2^3)\), where \(\alpha\) and \(A_2\) may parametrise quantum gravity effects, extra dimensions, or Lorentz invariance violations (Mirshekari et al. 2012). The parametrisation recovers for instance the effect of a massive graviton for \(\alpha = 0\) with \(A_0 = m_g^2c^4\) and \(m_g = \mu h/c^2 \neq 0\), where the group velocity becomes \((v_g/c) \simeq 1 - (m_g^2c^2/E)^2\). The combination of current LIGO/Virgo sources yields a constraint of \(m_g \leq 4.7 \times 10^{-23}\) eV/c\(^2\) (Abbott et al. 2019b). More distant and more massive sources are generally more effective in constraining the dispersion relation and the lower energy range of LISA will be favourable for tightening the bounds on \(m_g\) (Berti et al. 2005, 2011; Stavridis and Will 2009; Toubiana et al. 2020; Will 1998; Yagi and Tanaka 2010). Currently, the strongest bound on the mass of the graviton with \(m_g \leq 7 \times 10^{-32}\) eV/c\(^2\) is inferred from weak gravitational lensing (Choudhury et al. 2004).

Finally, the presence of a source term \(\Gamma \neq 0\) further modifies the GW amplitude with an oscillatory correction of order \(\Gamma \gamma_{ij}/(Hk)\) (Alexander and Yunes 2009; Nishizawa 2018). The effect can be constrained with LISA standard sirens (Belgacem et al. 2019c).

### 6.3.3 Stochastic gravitational wave background tests

Astrophysical sources of the SGWB can come from a variety of phenomena (Caprini et al. 2019; Schneider et al. 2010). The first of these sources is the unresolved evolution of stellar BBH, BNS, and BWD systems (Abbott et al. 2016a, 2018b; Farmer and Phinney 2003; Meacher et al. 2015; Zhao and Lu 2021; Zhu et al. 2011b, 2013). Superradiant instabilities are another astrophysical source of the SGWB (Barausse et al. 2018; Brito et al. 2017a, b; Cardoso et al. 2018a; Huang et al. 2019; Yoshino and Kodama 2014). The presence of a population of BH-bosonic condensates can form a GW background, which depend on the formation rate, number density, and spin. Superradiant instabilities occur when scalar field spinning around a BH transfer rotational energy and lead to the formation of a bosonic condensate. Other astrophysical SGWB sources that are less relevant for LISA include r-mode instabilities within NSs (Andersson 1998; Friedman and Morsink 1998; Ferrari et al. 1999; Zhu et al. 2011a), stellar core collapse (Buonanno et al. 2005; Crocker et al. 2017; Kuroda et al. 2014; Mueller et al. 2013; Ott et al. 2013), and magnetars (Marassi et al. 2011; Regimbau and de Freitas Pacheco 2006).

Together, these sources make up a signal whose SNR is too small to be detected as individually-resolved detections, and thus combine to form a SGWB (Allen and Romano 1999; Romano and Cornish 2017). The SGWB can be characterized by the spectral energy density as

\[
\Omega_{\text{GWB}} = \frac{1}{\rho_c} \frac{d\rho_{\text{GW}}}{d\ln f},
\]

where \(\rho_c = 3H_0^2/(8\pi)\) is the critical density necessary for a closed universe with Hubble constant \(H_0\), and \(\rho_{\text{GW}}\) is the energy density of all background GWs as a
function of frequency. This energy density may be approximated as a power law (Thrane and Romano 2013) given by

\[ \Omega_{\text{GWB}} = \Omega_\beta \left( \frac{f}{f_\ast} \right)^\beta, \quad (17) \]

where \( \Omega_\beta \) is the GW spectral energy density at a reference frequency \( f_\ast \), and the various factors of \( \beta \) correspond to different sources for the SGWB. These can be given for binary systems (\( \beta = 2/3 \)), r-mode instabilities (\( \beta = 2 \)), stellar core collapse (\( \beta \) is model dependent), superradiant instabilities (\( \beta = 1 - 7 \)), and magnetars (\( \beta = 3 \)) (Kuroyanagi et al. 2018). One can use such SGWB spectrum to test GR in a theory agnostic way, in particular the amplitude corrections in the ppE framework (Saffer and Yagi 2020).

A SGWB can also originate from the primordial tensor perturbations generated in the early universe during inflation or via alternative mechanisms such as thermal production. Given a primordial tensor spectrum \( P_t(k) \), one can compute the SGWB. The spectrum can be given directly by the model under study or, if allowed, in a parametrized form via the tensor spectral index \( n_t \) and its running \( \xi_t \):

\[ P_t(k) = P_t(k_0) \exp \left[ n_t(k_0) \ln \frac{k}{k_0} + \frac{\xi_t(k_0)}{2} \left( \ln \frac{k}{k_0} \right)^2 \right], \quad P_t(k_0) = r(k_0) P_s(k_0), \quad (18) \]

where \( n_t, \xi_t \) and the tensor-to-scalar ratio \( r \) are calculated at the pivot scale \( k_0 \), while \( P_s(k_0) \) is the measured amplitude of the scalar perturbations. During inflation, the running term does not play an important role because \( \xi_t \ll 1 \) and \( \left[ \ln(k/k_0) \right]^2 \ll 1 \). However, at higher frequencies \( \left[ \ln(k/k_0) \right]^2 \) increases and the running term can be large enough to affect the spectrum, an effect consistent with the above parametrization as long as \( \xi_t \) is small. The relation between the primordial tensor spectrum and the SGWB observed today is

\[ \Omega_{\text{gw}} = \frac{\pi^2 f^2}{3a_0^2 H_0^2} P_t(f) T^2(f), \quad (19) \]

where the transfer function \( T \), which we do not write here, depends on the history of the universe and on the details of reheating (Kuroyanagi et al. 2015; Turner et al. 1993; Watanabe and Komatsu 2006). When the primordial tensor spectrum is blue-tilted, the amplitude of the SGWB at high frequencies can increase up to the sensitivity threshold of LISA and other interferometers. In general relativity with standard inflation, the tensor spectrum is red-tilted and therefore its associated SGWB is unobservable. However, models beyond the standard paradigm in the matter or gravity sector could have a blue tilt and a large-enough amplitude (Bartolo et al. 2016; Calcagni and Kuroyanagi 2021; Kuroyanagi et al. 2018).

Methods for analyzing the LISA detection of an underlying SGWB signal have been developed in a source agnostic way (Caprini et al. 2019; Flauger et al. 2021; Karnesis et al. 2020; Pieroni and Barausse 2020). For example, a recent work
(Pieroni and Barausse 2020) shows that the SGWB from stellar origin BH and NS mergers may be observable with LISA with an SNR of $\sim 50$. This acts as a foreground noise to other backgrounds. One can employ a principal component analysis to subtract this astrophysical foreground, which enables LISA to detect SGWB that is 10 times weaker than the foreground. Given these methods, and a better understanding of source modelling which may occur by launch, the prospects for detecting and analyzing a SGWB with LISA are promising.

A strategy to separate the astrophysical from the cosmological background in the context of LISA has been proposed in Boileau et al. (2021) using a Bayesian strategy based on an Adaptive Markov chain Monte-Carlo algorithm. This method has been later used in the case of a SGWB produced by a network of cosmic string loops (Boileau et al. 2022), updating a previous study where only the instrument noise was taken into account (Auclair et al. 2020). In particular, it has been shown that given the ability of LISA to simultaneously detect a large number of galactic double white dwarf binaries and a large number of compact binaries, a cosmic string tension (Newton’s constant is denoted by $G$ and the string linear mass density by $\mu$) in the $G\mu \approx 10^{-16}$ (for loop distribution model; Kibble 1985) to $G\mu \approx 10^{-15}$ (for loop distribution model; Blanco-Pillado et al. 2014; Lorenz et al. 2010; Ringeval et al. 2007) range or bigger could be measured by LISA, with the galactic foreground affecting this limit more than the astrophysical one.

### 6.4 Burning questions and needed developments

Throughout this section, we have described in passing a few of the many challenges we still face when developing and carrying out model-independent tests. We therefore end this section by presenting a brief summary of the main open problems.

- Residual tests, as well as other tests, need to be more carefully studied in the context of LISA, due to the expected abundance of sources at multiple frequencies. This is because the LISA data analysis challenge is a global one, in which multiple signals must be estimated simultaneously.

- Regarding the IMR consistency tests (and for other tests), the challenge is to ascertain that the waveform models used in the measurement process are accurate enough so as not to cause systematics, especially at the large SNRs expected for BBH mergers in LISA. This can affect both the extraction of parameters in the inspiral, as well as the inference of remnant masses and spins in the merger. It is therefore important to control modelling systematics to the required level. Work must begin now to address these issues to either demonstrate the readiness of the waveform models or to establish the required accuracy.

- We currently lack complete IMR waveform model in non-GR theories, even with simple extension to GR, which prevents us from performing a full mapping between parametrized tests and specific theories. This is partially because we currently lack sufficient NR simulations performed in theories beyond GR. Furthermore, many of non-GR corrections to the inspiral portion have focused on the leading PN contribution, and therefore they neglect among other features,
higher harmonics. Higher PN order terms may become important as the binary separation becomes smaller, especially at LISA SNRs.

- Parameterized ringdown tests need further developments, such as including the BH spin without the slow-rotation approximation and finding more mappings to specific non-GR theories. This also requires that we construct new rotating BH solutions in theories beyond GR at arbitrary rotation.

- One needs to reveal how the complex interplay between astrophysical event rates, non-GR waveform models, ground-based detector developments and data analysis methods will affect our ability to extract fundamental physics information with LISA.

- An interesting open problem regarding GW propagation is the mixture of propagation and source effects for extended theories of gravity and the role of screening mechanisms. A PN expansion for the source emission in screened regimes can be performed using a scaling relation (McManus et al. 2017; Perkins and Yunes 2019; Renevey et al. 2020). For some GR extensions the relevant modification in \( \nu \) are in fact determined by the screened environments of emitter and observer rather than the cosmological background, whereas for \( c_T \) screening effects may safely be neglected.

- Tests that aim to constrain the existence of non-GR polarizations need further development, both from the standpoint of generic (null channel type) tests, as well as from the standpoint of specific examples in given modified theories.

- The tests described above, for the most part, require a specific or generic model for the signal, but LISA could have access to GWs that are not properly covered by such models. Efforts to develop methods to extract unexpected signals should also be pursued.

- More detailed analyses are needed to see how well one can remove the astrophysical background from stellar origin BH and neutron star mergers and how much its residuals affect tests of GR.

7 Astrophysical and waveform systematics

The discovery of GWs has provided a powerful new tool to probe gravity in its most extreme regime and to search for signatures of new physics. To do so, however, requires the development of highly accurate waveform templates—our “filters” to identify and interpret gravitational signals. Before being able to claim a discovery of new physics we need to

1. provide more accurate waveforms within GR to reduce systematic errors (and potential misinterpretation) due to modelling systematics;
2. understand astrophysical and environmental effects to avoid misinterpreting deviations from vacuum GR; and
3. construct theory-specific waveforms for targeted tests of (classes of) beyond-GR theories.
A detailed account of the status and challenges of gravitational waveform modelling and their systematics can be found in the LISA Waveform WG white paper (in prep.). Here we give a summary with focus on its relevance for the fundamental physics WG.

### 7.1 Astrophysics: environmental effects

LISA sources, whether supermassive or stellar-mass, are likely to be surrounded by matter that could limit our ability to place constraints on fundamental physics. In the case of supermassive BBHs the impact of gaseous accretion raises questions regarding our ability to probe fundamental physics, especially in the case of EMRIs. Stellar mass BHs can also be sources for LISA and in this case if they are inside dense stellar clusters, then they could be in the form of hierarchical triplets.

#### 7.1.1 Binary and disk interactions

When a BBH is embedded in a magnetized gaseous disk, binary-disk interactions can induce change in the evolution of the GW phase. Typically the evolution of a binary disk-system is broadly composed of two phases: i) the pre-decoupling phase where the disk is coupled to the binary, and tracks the inspiral of the binary, because the GW timescale is very long compared to the effective disk viscous timescale which controls the inflow velocity of the disk; ii) the post-decoupling phase where the GW timescale becomes shorter than the disk effective viscous timescale and the binary begins to run away from the disk which can no longer track the inspiral of the binary.

**a. Comparable mass binaries:** Equating the GW timescale with the disk viscous timescale (Milosavljević and Phinney 2005), one finds that the decoupling radius is given by (Gold et al. 2014)

\[
\frac{a_d}{GM/c^2} \approx 140 \left( \frac{a}{0.01} \right)^{-2/5} \left( \frac{H/R}{0.05} \right)^{-4/5} \left( \frac{\eta}{1} \right),
\]

where \(M\) is the binary’s total mass, \(H/R\) is the disk aspect ratio, with \(H\) the disk scale height (or thickness), \(a\) is the Shakura–Sunyaev viscous parameter (Shakura and Sunyaev 1973), and \(\eta = 4q/(q + 1)^2\) is four times the symmetric mass ratio, with \(q\) the binary mass ratio (see also Armitage and Natarajan 2002). Notice that for extreme mass-ratio and intermediate mass-ratio inspirals the decoupling radius can be comparable to the gravitational radius of the binary, and thus affect its evolution if the disk is massive enough (Kocsis et al. 2011; Yunes et al. 2011b). A good understanding of binary-disk interactions is important to better understand the binary evolution into the LISA band. For a binary at the decoupling radius, the GW frequency is given by:

\[
f_{GW} \approx 10^{-3} \text{Hz} \left( \frac{a_d}{140GM/c^2} \right)^{-3/2} \left( \frac{M}{10^6 M_\odot} \right)^{-1},
\]

which suggests that for typical values of the viscosity parameter, and if the disks are
slim (Abramowicz and Fragile 2013), decoupling could occur inside the LISA band, and the effects of binary-disk could be imprinted on the GWs. For example, the binary residual eccentricity when it enters the LISA band is determined at binary-disk decoupling (Roedig et al. 2011). For disk thickness, e.g., $H/R \sim 0.2$, for a comparable mass the decoupling radius becomes $a_d \sim 40GM/c^2$ at $f_{GW} \sim 0.01Hz$, i.e., well within the LISA band. The work of (Roedig et al. 2011) suggests that residual eccentricities of up to $10^{-2}$ in the LISA band are possible. How these effects could affect precision tests of GR has not been studied, yet. Note however that at least for SOBHBs, environmental effects can be significant in gas-rich systems (such as AGN disks) (Caputo et al. 2020; Toubiana et al. 2021). Because these effects typically enter at low frequencies, they are expected to compete with (or even dominate over) non-GR effects such as vacuum dipole emission.

b. EMRIs: EMRIs are thought to form in dense galactic nuclei where a passing by stellar compact object (SCO), or a nascent SCO formed in the accretion disk via fragmentation and/or coagulation of density enhancements (Goodman and Tan 2004; Levin 2007; Milosavljevic and Loeb 2004), can be captured by the central SMBH. It is expected that only EMRIs involving compact enough objects, such as WDs, NSs, stellar-mass BHs or probably the Helium cores of giant stars, can survive to low redshifts ($z \lesssim 1$) where EMRIs may be observed (Amaro-Seoane et al. 2007; Barack and Cutler 2004).

A typical EMRI event will be emitting GWs with frequencies ranging between $10^{-4}$ to 1 Hz, which are expected to be in the LISA frequency band for many years, allowing for accurate measurements of the parameters of the system. In particular, a GW event detection with a SNR of $\sim 30$ will allow us to determine the mass of the SMBH with a precision of $\sim 10^{-4}$ and the distance to the source with a precision of $\sim 3\%$ (Gair et al. 2010) (see also Babak et al. 2017a). Moreover, these events can be used to probe cosmological parameters, test GR in the strong field regime (e.g., EMRI may provide a constraint on deviations from the Kerr metric at a level of 0.01–1%; Babak et al. 2017a; Gair et al. 2017), etc. It is, therefore, necessary to assess whether accretion or other SCOs in the vicinity of the EMRI can induce observable changes in the GW frequency. This is especially important for sources in active galaxies with BHs accreting near the Eddington limit, for which the disk mass can be comparable to the secondary object.

Tidal interaction of the SCO with the accretion disk may induce mutual angular momentum exchange that modifies both the disk and the orbit of the SCO (Ward 1997). The nature of this interaction depends basically on the ratio of their masses $q_{CO,disk} \equiv M_{SCO}/M_{disk}$. If $q_{CO,disk} > 1$, tidal torques exerted by the SCO are strong enough to carve a cavity in the disk. Radial inflow of the gas disk then pushes the SCO inward causing the binary to harden on a time-scale comparable to the viscous timescale $t_{visc}$ (Type-II migration) (Gould and Rix 2000; Ward 1997). Furthermore, if the mass of the SCO exceeds the mass of the nearby gas disc, the migration slows down as the matter cannot remove angular momentum away from the binary at a rate on which the gas flows (Ivanov et al. 1999; Syer and Clarke 1995). Analytic work has predicted that the “pile-up” of gas disc around the orbit of the SCO mitigates the slow down, and the hardening timescale is $t_{hard} \approx q_{2,disk}^2 t_{visc}$ with a relatively shallow
scaling of $z \approx 0.5$ in a steady-state limit (Ivanov et al. 1999; Syer and Clarke 1995). However, numerical simulations have found no evidence for this pile-up, and instead suggest a simple scaling with $z \ll 1$ (Duffell et al. 2014; Dürmann and Kley 2017; Kanagawa et al. 2018). On the other hand, if the $q_{\text{CO,disk}} \lesssim 1$, the SCO is not massive enough to carve the cavity, it will migrate inwards by exciting density waves in the disc (Goldreich and Tremaine 1980; Tanaka et al. 2002) (Type-I migration). However, this mechanism is very sensitive to the temperature and opacity of the disk (Jang-Condell and Sasselov 2005; Menou and Goodman 2004). If the radiative cooling in the dense inner regions surrounding the SCO is inefficient, torques exerted by the disc will induce outward migration, instead of the usual inward migration found in locally isothermal disks, up to regions when the disc opacity is low. As the accretion proceeds, these regions move inward, and hence the SCO’s migration will continue on a timescale much longer than the viscous timescale (Paardekooper and Mellema 2006). A slower migration mechanism has been recently identified (Kocsis et al. 2012). In this case, the gas piles-up significantly outside the orbit of the SCO, but the viscosity increases to the point that in steady-state gas can reach the SCO’s Hill sphere, and is able to flow across its orbit (Type-1.5).

The effect of such interactions on the waveform might be degenerate with the effects due to modifications to GR (Barausse et al. 2014). In such a scenario, the precision with which modifications to GR can be measured with LISA will be significantly affected, since, in order to be detectable, the effect induced by GR modifications should be larger than the environmental effect. Effectively, this induces a lower bound on GR modifications observable with LISA. Development of waveforms beyond GR is in early stages, and the role of astrophysical systematics has not been studied in any detail.

The inspiral part of the waveform will be largely unaffected by the astrophysical interactions as long as the BH resides in a thick disk environment, where densities are low. On the contrary, in thin-disk environments the waveform can be significantly perturbed by certain phenomena. Planetary migration can have a strong influence on the inspiral (Kocsis et al. 2011; Yunes et al. 2011b), while dynamical friction (Barausse and Rezzolla 2008; Derdzinski et al. 2021, 2019) and accretion (Barausse and Rezzolla 2008; Barausse et al. 2014; Cardoso and Maselli 2020) have a weaker but non-negligible effect. The effect of the disk’s gravity can safely be neglected. All this can have a large enough effect to cause systematic errors when performing precision tests of gravity (Barausse et al. 2014).

c. Ringdown: A remarkable feature of the ringdown of GR BHs is the isospectrality of the polar and axial modes (Kocsis et al. 2011; Pani et al. 2013a; Yunes et al. 2011b; Wagle et al. 2021). Any deviation between the axial and polar modes could indicate new physics, but matter distribution around the BH also breaks the isospectrality (Barausse et al. 2014), making it non-trivial to identify GR modifications this way. While a smoking-gun type test is not possible due to this degeneracy, the impact of typical matter distributions around a BH on the ringdown frequencies is very weak. Thus, the lower bounds introduced by the degeneracy is low enough for ringdown-based strong-fields tests of GR to be meaningful (Barausse et al. 2014).
d. Plasma effects: Even in the absence of a massive accretion disk around the binary, recent work suggests that for binaries embedded in a plasma environment, it may be challenging to perform fundamental physics tests that probe plasma-driven superradiant instabilities, EM emission from charged BH binaries and electromagnetically driven secondary modes in the GW ringdown signal (Cardoso et al. 2021). The latter two effects can be relevant also in the case of massive fields propagating in vacuum, and provide an obstacle to tests of modified gravity with massive degrees of freedom. In addition, matter effects can change the tidal-deformability of a vacuum BBH (Cardoso and Duque 2020), which could potentially limit our ability to test for dark compact objects that are not BHs.

7.1.2 Interaction with stellar environments and hierarchical triplets

In recent years interaction of BBHs with a third BH in stellar environments or near a SMBH have revealed new ways to think about LISA sources that exhibit effects which are absent in a standard isolated binary evolution. SMBH triplets can potentially induce binary eccentricities as high as 0.9 (Bonetti et al. 2019) upon entrance in the LISA band, and could be as high as 0.1 at merger. Apart from eccentricity tidal effects induced on the binary by the companion BHs could contribute significantly to the binary waveform (Yang and Casals 2017) by e.g. shifting the innermost stable orbit, and hence the onset of the plunge phase. BH binaries formed through such channels not only would require suitable templates for parameter estimation, but also necessitates an investigation on how eccentricity may affect tests of fundamental physics.

In the case of stellar-mass BBHs, a non-zero graviton mass and dipole emission can be probed (Toubiana et al. 2020) among other pieces of fundamental physics. However, if the binary is in a dense stellar environment or near a SMBH 3 body effects can become important (Blaes et al. 2002; Miller and Hamilton 2002; Naoz et al. 2013; Randall and Xianyu 2018, 2019; Yunes et al. 2011a; Wen 2003), e.g., through Kozai-Lidov mechanism (Kozai 1962; Lidov 1962) oscillations. The extent to which such 3-body effects can provide an obstacle to probing fundamental physics with stellar mass BH binaries has not been investigated, yet.

7.2 Waveform modelling and systematics for EMRIs

The most promising avenue for GW modelling of EMRIs is perturbation theory (gravitational self-force), see e.g., Barack and Pound (2019) for a recent review. The full calculation of the EMRI with perturbation theory is a formidable problem as it requires first-order metric perturbation for the gravitational self-force, second-order metric perturbation for the dissipative part, the evolution of the inspiral and the final calculation of the waveform. The small parameter for the perturbation is the mass ratio $q$. According to Hughes (2016) reliable templates will require that one keeps track of at least $O(q^2)$ terms. The state-of-the-art in the calculation of the self-force is the first-order calculation of van de Meent (2018) for generic orbits with spinning BHs, and the second-order calculation for circular orbits and non-spinning BHs.
(Pound et al. 2020), and including the spin of the secondary (Akcay et al. 2020). The state-of-the-art in fast calculation of EMRI GWs are the augmented kludge waveforms of Chua et al. (2017), the two-timescale expansion approach (Hinderer and Flanagan 2008; Miller and Pound 2021), near-identity transformations (Van De Meent and Warburton 2018), and the self-consistent evolution of Diener et al. (2012).

Apart from not yet having the full self-force calculation to second-order for generic primary BHs that include the spin of the secondary on generic orbits, potential systematics in the calculation involve the internal structure of the secondary (Isoyama and Poisson 2012; Witzany 2019) and non-vacuum backgrounds (Zimmerman and Poisson 2014), e.g., the case where the EMRI is embedded in a dark matter bosonic condensate.

It is worth noting that several of the issues faced by EMRI modelling are also challenges in the case of intermediate mass-ratio inspirals.

7.3 Waveform modelling in GR and systematics for comparable mass binaries

Modelling the gravitational waveform of massive binaries covering their inspiral, merger and ringdown requires a combination of perturbative techniques and NR. In particular, the inspiral is modelled by PN theory, the late inspiral and merger where nonlinear effects of gravity become relevant require full NR, and the ringing down of the final BH is described within black-hole perturbation theory. These different pieces are combined into full IMR waveforms using effective-one-body (EoB) or phenomenological models.

7.3.1 PN waveforms

PN methods are analytic techniques based on a series expansion in weak fields and small velocities that describe the dynamics of a binary during the early inspiral phase. The accuracy of the description is controlled by the PN order, which identifies the relative order to which an expansion has been taken in small velocities and weak fields. While waveforms of non-spinning, quasi-circular binaries are currently known up to 4PN order (Blanchet 2014), those of eccentric, spinning or precessing binaries are currently only available up to 3PN and 3.5PN order, e.g., Moore and Yunes (2019, 2020); Tanay et al. (2021). Efforts to go beyond the current state of the art are under way, see, e.g., Bini et al. (2019, 2020a); Foffa et al. (2019). Although for ground-based GW observatories going to high PN order will be beneficial, a recent analysis of requirements for LISA non-eccentric stellar mass BBHs finds that 3PN order is sufficient for all sources, while for ~90% of those sources, waveforms at PN order ≤ 2 are sufficiently accurate for an unbiased recovery of the source parameters (Mangiagli et al. 2019). However, at this time it is unclear what PN accuracy is required for SMBH binaries. This will be crucial to determine as the high SNR ratio at which LISA is going to detect these binaries implies that the statistical errors are going to be smaller than the systematic errors in the theoretical waveforms. These systematic errors could hinder our ability to perform precision tests of fundamental physics with LISA.
7.3.2 NR waveforms

NR techniques are numerical methods for solving the equations of relativistic gravitation that are particularly suited in regimes where perturbation theory and PN methods fail (see e.g., Alcubierre 2008; Baumgarte and Shapiro 2010a; Bona et al. 2009; Gourgoulhon 2012; Shibata 2015 and references therein). In the case of BBH, NR is the only reliable method for capturing the late inspiral and merger of the binary from first principles. NR methods are used both in the case of GR and in modified gravity, on which we expand further below.

Despite the tremendous strides made by NR since the breakthrough simulations of Baker et al. (2006); Campanelli et al. (2006); Pretorius (2005), NR still faces several challenges: current methods become prohibitively expensive for high mass ratios \(q \gtrsim 10\) and near extremal BH spins. As a result, calibration of phenomenological waveforms such as Phenom or EOB (see next section) based on NR methods have systematic errors in the more extreme parts of the parameter space. Such systematic errors can pose challenges when performing precision tests of fundamental physics with LISA. Furthermore, it is known that in some cases numerically generated waveforms exhibit non-convergence at very high grid resolution (Zlochower et al. 2012), which makes it challenging to assess the accuracy of the waveforms. Resolving these non-convergence problems may require new gauge conditions (see e.g., Etienne et al. 2014) or next generation codes. Finally, comparisons in the context of the NRAR Collaboration (Hinder et al. 2014) showed that even when adopting the same methods at the analytic level, different NR codes (i.e. different implementations) do not entirely agree at the finite resolutions typically adopted to generate high-quality NR waveforms. Efforts like the NRAR Collaboration and code comparisons are important to assess the accuracy requirements for LISA. Initial work toward assessing these requirements (Ferguson et al. 2021) (see also Pürrer and Haster 2020) suggests that NR will require a substantial increase in accuracy compared to today’s most accurate waveforms; for example it is found that reaching with NR templates SNRs above 1000, finite-difference NR codes would have to efficiently scale to resolutions of at least \(\Delta x < M/700\) (\(M\) being the binary gravitational mass). Reaching resolutions for template indistinguishability is particularly important because residuals resulting from using lower resolution templates can be comparable to those resulting from ignoring higher modes entirely.

7.3.3 Phenom and EoB waveforms

Taracchini et al. (2012, 2014) are the current state-of-the-art waveforms used in GW data analysis. Using two different waveform models is essential to assess the size of systematic uncertainties in the measurements. The Phenom and EOB modelling approaches are both based on combining information from analytical and NR. For binaries of BHs (and NS-BHs where tidal disruption does not occur; Matas et al. 2020; Thompson et al. 2020), they describe the entire signal from the early inspiral to the ringdown of the final remnant. The Phenom models provide a closed-form description of the frequency-domain signals, although see Estellés et al. (2021) for a time-domain approach. They are focused primarily on efficiency and based on a modular assembly of combined theoretical and numerical insights into the main important features of the waveforms. The EOB models describe both the binary dynamics and waveforms in the time-domain. They are based on incorporating theoretical inputs from a variety of approximation methods into the structure of the model before introducing calibrations to NR. The efficiency of the EOB waveforms is then improved, e.g., by developing reduced-order frequency-domain models (Bohé et al. 2017; Canizares et al. 2015; Field et al. 2014; Lackey et al. 2017; Pürrer 2014, 2016) or using post-adiabatic approximations (Gamba et al. 2021; Nagar and Rettegno 2019); see also García-Quirós et al. (2021) for recent applications to Phenom models.

There is ongoing rapid progress on improving the accuracy and physical realism of the waveform models within GR, as systematic differences are already starting to become noticeable for LIGO/Virgo/KAGRA detections such as Abbott et al. (2020b). For instance, the baseline Phenom model has recently been significantly upgraded for the dominant mode of aligned spin configurations (Pratten et al. 2020). Current models include several physical effects for circular orbits: a set of higher harmonic modes for aligned spins (Cotesta et al. 2018; García-Quirós et al. 2020; London et al. 2018; Nagar et al. 2020b), spin precession (Hannam et al. 2014; Khan et al. 2019; Ramos-Buades et al. 2020b; Schmidt et al. 2015), and spin precession with higher harmonics (Pratten et al. 2021; Khan et al. 2020; Ossokine et al. 2020). Models for nonprecessing binaries with mildly eccentric orbits have also recently been developed (Cao and Han 2017; Chiaramello and Nagar 2020; Liu et al. 2020) and explored (Hinder et al. 2018; Hinderer and Babak 2017; Ramos-Buades et al. 2020a).

The Phenom and EOB models have been widely tested against NR simulations mainly in the regime of comparable masses, i.e. mass ratios below four, and moderately high spins below eighty percent of the maximum, aside from a few more extreme cases. Ongoing efforts to improve the structure of the models aim to include more information from a mix of perturbative approaches involving PN calculations combined with results from scattering calculations within the post-Minkowski approximation and gravitational self-force results (Antonelli et al. 2019, 2020; Bini et al. 2019, 2020b; Khalil et al. 2020). Significant further work is required to turn these theoretical explorations into full, calibrated models, and deploy them for data analysis. While all of these developments mark important advances, waveforms for LISA will require models to account for yet more physics content and have higher accuracy than current waveform models. MBH binary mergers and intermediate mass ratio systems will generically involve eccentricity, arbitrary spins, substantially...
different masses, and SNR of up to several thousands. The resulting waveforms will have a very rich structure characterized by Fourier harmonics with several different frequencies. Further advances in both analytical and NR will be required as inputs for developing adequate Phenom and EOB models for these sources.

For objects beyond BHs, Phenom and EOB models for the inspiral signals are also available. In this regime, a number of generic spin and tidal effects lead to characteristic GW signatures that depend on the objects’ internal structure. Current models include the matter effects from spin-induced quadrupole moments and tidal deformability (Akcay et al. 2019; Bernuzzi et al. 2012a, 2015b; Bini and Damour 2014; Bini et al. 2012; Damour and Nagar 2010; Damour et al. 2012; Dietrich et al. 2017, 2019a, b; Flanagan and Hinderer 2008; Kawaguchi et al. 2018; Nagar and Rettemegno 2019; Nagar et al. 2019, 2018; Vines et al. 2011), and the tidal excitation of fundamental quasi-normal modes (Hinderer et al. 2016; Schmidt and Hinderer 2019; Steinhoff et al. 2016). For the time-domain EOB models, the matter effects are based solely on theoretical results without any additional calibrations to NR. The frequency-domain models can either employ PN results for matter effects on top of the BH baseline, or, as has become standard for current data analysis, a calibrated tidal model based on NR simulations for binary NSs. In all of these models, matter signatures are described in a parametric form, where the characteristic coefficients such as tidal and rotational Love numbers, and quasi-normal mode frequencies encode the fundamental information on the matter. As these matter waveforms describe only the inspiral, they are tapered to zero in practical applications at the predicted merger frequency of double NS systems as determined from fits to NR simulations (Bernuzzi et al. 2015a), or when the fundamental mode resonance is reached. Future work remains on going beyond these restrictions and on enlarging the physics content of the models with matter phenomena. Another effect that has not yet been included are the objects’ absorption coefficients. All of this remaining work presents no fundamental obstacles but will require time and effort to develop the full models. Environmental effects such as from dynamical friction or energy-level transitions in clouds of ultralight fields are not yet included in the Phenom and EOB models. This is in part due to the fact that at present, studies of these phenomena are mainly exploratory. Developing a deeper understanding and more accurate descriptions of the backreaction and evolution of the system during an inspiral will be useful before including these phenomena in the full waveform models.

As discussed in other parts of this white paper, there are many ways to test GR, but let us here focus on three broad classes: parameterized deviations of the waveforms from GR during all epochs, IMR consistency tests, and spectroscopy of the final remnant. For the Phenom models, where the signals are described by closed-form algebraic expressions, it is straightforward to add extra beyond-GR parameters (Abbott et al. 2019a). For the EOB model, a similar approach is used with a slightly different form of parameterized deformations. These are added to reduced-order models of frequency-domain EOB waveforms (Sennett et al. 2020). Current data analysis for these theory-agnostic tests constrain only one deviation parameter at a time, as it is unfeasible to obtain useful information on a very large number of extra parameters in the waveforms. Tests of the consistency of the inspiral and the ringdown with the EOB and Phenom models (Abbott et al. 2019b; Ghosh et al.
2016b) rely on measuring the remnant properties (Hughes and Menou 2005) and comparing with the final mass and spin predicted from the progenitor parameters by fits to NR data (Bohé et al. 2017; Haegel and Husa 2020; Jiménez-Forteza et al. 2017). For spectroscopy using the ringdown of the remnant from a BBH merger, parameterized deformations of the quasi-normal modes, full waveform models can improve the constraints (Brito et al. 2018). Recent work has also started to develop foundations for theory-specific models as examples of non-GR waveforms. For instance, the EOB dynamics for nonspinning binaries have been calculated in scalar-tensor theories of gravity (Julié 2018; Julié and Deruelle 2017), Einstein-Maxwell-dilaton theories (Khalil et al. 2018) and Einstein-scalar-Gauss–Bonnet gravity (Julié and Berti 2019). Significant further work and inputs from NR will be required to develop complete waveforms, and to incorporate the expected physical effects such as spins, mass ratio, and eccentricity.

7.4 Waveform modelling beyond GR and systematics for comparable mass binaries

7.4.1 NR waveforms

Let us now discuss the current systematic errors of NR BBH merger waveforms beyond GR, and how to address these errors in future simulations. In particular, we will focus on recent work in beyond-GR effective field theories of gravity, including dynamical Chem–Simons (dCS) gravity, Einstein-dilaton-Gauss–Bonnet (EdGB), Einstein-scalar-Gauss–Bonnet gravity (EsGB), Einstein-dilaton-Maxwell (EdM), and a flavor of a tensor-vector-scalar gravity as these are the only beyond-GR theories (aside from massive scalar-tensor gravity, which is identical to GR under ordinary initial and boundary conditions; Berti et al. 2015) for which we have non-linear BBH evolutions (Bozzola and Paschalidis 2021; East and Ripley 2021; Hirschmann et al. 2018; Okounkova 2020; Okounkova et al. 2019b, 2020; Witek et al. 2019).

For NR BBH simulations in GR, the greatest source of systematic error is finite numerical resolution (cf. Baumgarte and Shapiro 2010b). To quantify numerical resolution errors, simulations are performed for a series of increasing numerical resolutions, checking for convergence of the resulting gravitational waveforms. The error for a given waveform is reported as the mismatch between the waveform and the waveform from the next-lowest resolution simulation (see, e.g., the methods in Boyle et al. 2019 for technical details). For gravity theories which admit a well-posed initial value problem no approximation to the theory is necessary, and hence numerical accuracy is still the biggest source of systematic errors. For example, this is the case in the recent evolutions in EdM (Hirschmann et al. 2018), which admits a well-posed initial value problem, and in EsGB (East and Ripley 2021), where a well-posed modified Harmonic formulation of the theory is adopted.

BBH simulations beyond-GR, such as in dCS gravity (Okounkova et al. 2020) and Einstein dilaton Gauss–Bonnet gravity (Okounkova 2020) have numerical resolution errors comparable to those of GR (see the convergence analyses in Okounkova et al. 2019a, 2020; Witek et al. 2019), and thus the codes to generate these waveforms must improve in the same way as GR codes in order to be ready for LISA. However,
in addition to numerical resolution errors (comparable to those of GR), present beyond-GR NR simulations have additional sources of systematic errors. In order to ensure a well-posed initial value problem (which may not exist in EdGB; Papallo 2017; Papallo and Reall 2017; Ripley and Pretorius 2019b, a, and dCS; Delsate et al. 2015), BBH mergers in quadratic gravity are performed in an order-reduction scheme, where the spacetime metric is expanded about a GR BBH merger spacetime, and one computes the leading-order beyond-GR correction to the metric and gravitational waveform perturbatively. One then obtains the ‘full’ gravitational waveform by combining the background GR metric with the leading-order correction, for a suitable choice of EdGB or dCS coupling constant. The order-reduction scheme operates under the assumption that corrections to GR are small, and that higher-order corrections (such as the second-order correction to the spacetime metric) are negligible. This assumption governs an instantaneous regime of validity, where there is a maximum allowed value of the coupling constant that is allowed in order for the corrections to the spacetime metric to be smaller than the background metric (see Okounkova et al. 2019b for technical details). Thus, the NR simulations performed in this scheme are not valid for all possible physical values of coupling constant.

Additionally, the order-reduction scheme introduces systematic errors through secular dephasing. Since the location of the BHs in the perturbative scheme is governed by GR (with no back-reaction from the beyond-GR fields), the phase of the BBH inspiral in the order-reduction scheme is different from that of the full theory, hence leading to secular dephasing. This is a common feature of perturbative approximations to dynamical systems (Bender and Orszag 1978), and in particular, PN approximations (Will and Maitra 2017) and self-force BH perturbation theory (Pound et al. 2005) encounter these challenges when applying perturbative approaches to long-duration inspirals. Hence, the present beyond-GR waveforms focus primarily on the merger (the phase that is crucial to numerically simulate for tests of GR; Yunes et al. 2016) and ringdown phases, ensuring that the beyond-GR simulation starts late enough so that the secular systematic errors are negligible (Okounkova 2020; Okounkova et al. 2020). Finally, the extraction of GWs themselves is another potential source of systematic errors. For example, it is unclear whether the Newman–Penrose scalars that are commonly adopted in GR, have the same physical interpretation in modified gravity.

Another modified theory where some work has been done is Moffat’s tensor-vector-scalar theory (Moffat 2015). The theory admits a well-posed initial value problem, but for sufficiently small scales, it becomes mathematically equivalent to Einstein–Maxwell theory in vacuum, where BHs carry a “gravitational” (not electric) charge, that is determined by the theory’s coupling constant. First simulations in this context have been performed by Bozzola and Paschalidis (2021). However, a systematic error in this case is whether the approximation of the theory as Einstein–Maxwell in vacuum holds true. Given that the theory admits a well-posed initial value problem, it is straightforward to move beyond this approximation.

Moving forward one would like to obtain NR beyond-GR mergers that are free of these systematic effects by being valid for all values of coupling constants and avoiding secular dephasing. In order to mitigate dephasing growth, one can ‘stitch’
together results for simulations with different starting points of beyond-GR effects (hence different points of zero dephasing), as discussed for the case of EMRIs in Pound and Poisson (2008); Warburton et al. (2012). Similarly, one can apply the multiscale and dynamical renormalization group methods of Galley and Rothstein (2017); Hinderer and Flanagan (2008); Kunihiro (1995); Pound (2015b) to get rid of secular dephasing effects. This is an active area of work.

In order to produce beyond-GR BBH waveforms valid for all values of coupling parameter, one must move beyond perturbative methods and consider simulating BBHs in beyond-GR theories with well-posed initial value problems. Such approaches have been discussed using methods from fluid dynamics in Allwright and Lehner (2019); Cayoso et al. (2017). In particular, Witek et al. (2020) recently derived a 3+1 split for the (non-perturbative) equations of motion of EdGB in a step towards NR simulation.

7.4.2 PN waveforms

Incomplete inspiral waveforms in theories beyond GR may give rise to systematic errors in tests of GR with GWs. The inspiral waveform is computed via the PN method. In most non-GR theories, only the leading PN correction has been computed, which can be mapped to the PPE framework (see Sect. 6.2.1). However, there are certain theories in which corrections to the waveform have been computed also to higher PN orders (Sennett et al. 2016; Shiralilou et al. 2021, 2022; Yunes et al. 2012; Zhang et al. 2020). One such example is the Brans–Dicke theory, which is one of the simplest forms of scalar-tensor theories. The effect of higher PN corrections in this theory to constrain the theoretical coupling constant (Brans–Dicke parameter) has been studied in Yunes et al. (2016). The authors showed that the bound on the Brans–Dicke parameter with GW150914 only changes by \( \approx 10\% \) when one includes higher PN corrections or not. This shows that the higher-order corrections are not important when constraining Brans–Dicke theory, though there may be other non-GR theories where sub-leading PN corrections become important, and thus deriving such corrections is an important direction for future research (see Sect. 6.4).

7.4.3 Effects of bosonic dark matter

The effects of bosonic dark matter in the context of GR are discussed in Sect. 4. Here we discuss effects related to modified gravity.

Ultralight bosonic dark matter around spinning BHs can trigger superradiant instabilities, forming long-lived bosonic condensates outside the horizon which can alter the inspiral GW signal. A number of modified gravity theories with additional (scalar and/or vector) degrees of freedom exhibit phenomena such as dynamical scalarization (see e.g., Benkel et al. 2016; Berti et al. 2021; East and Ripley 2021; Ripley and Pretorius 2019a; Witek et al. 2019 for Einstein-Scalar-Gauss–Bonnet Theories and Fernandes et al. 2019; Herdeiro et al. 2018b; Hirschmann et al. 2018 for Einstein-Scalar-Maxwell theories), where BHs in those theories spontaneously or dynamically acquire scalar hair. How BHs in modified gravity behave around ultra-light bosonic dark matter environments, and what new effects could arise as well as
our ability to place constraints on deviations from GR and/or on dark matter has not yet been explored.

### 7.4.4 EMRI waveforms

With EMRI signals, we expect to be able to detect higher order-independent multipole moments of the central body (Ryan 1995, 1997a) and measure (if any) deviations from GR (Barausse et al. 2020). Most of the work in this direction, has so far considered BH spacetimes with a multipolar structure different from Kerr, such as bumpy BHs (Collins and Hughes 2004), and constructed approximate “kludge” waveforms (Glampedakis et al. 2002; Chua et al. 2017) generated considering geodesics in a perturbed Kerr spacetime with orbital parameters evolved using PN equations that assume radiation-reaction effects as in GR (Apostolatos et al. 2009; Barack and Cutler 2007; Gair et al. 2008; Glampedakis and Babak 2006; Moore et al. 2017).

This approach reproduces the orbit’s main features but not to the precision required to determine that the inspiral is indeed an inspiral into a Kerr BH or not (Gair et al. 2008; Sasaki and Tagoshi 2003). Thus, further work on EMRI waveform modelling will require the consistent inclusion of radiation reaction not described as in GR, additional parameters related to the modification from GR, and environmental effects. Note that the presence of environmental effects may limit our ability to perform proper parameter estimation for some events (Barausse et al. 2014, 2015) if these are not properly taken into account.

These enhanced waveforms will increase model degeneracies that will impact our ability to learn about fundamental physics in the strong-field regime of gravity. That is why it is paramount to fully understand possible discernible features, such as characteristic variations of the amplitude and the energy emission rate (Suzuki and Maeda 2000) or the appearance of prolonged resonances (Apostolatos et al. 2009) in EMRI signals, as well as the development of novel data analysis techniques to extract information.

### 7.5 Burning questions and needed developments

In this chapter we have discussed how different systematics in the modelling of waveforms and induced by astrophysical environment can affect tests of GR. Below, we summarize the most important salient points.

- How does (residual) eccentricity induced by binary-disk or hierarchical triplet interactions limit our ability to perform fundamental physics tests with BH binaries?
- Can accretion, plasma effects or other stellar compact objects in the vicinity of an EMRI induce observable changes in the GW frequency evolution during the inspiral and/or ringdown that can spoil fundamental physics tests?
- Self-force calculations for generic BBHs in vacuum or embedded in a background (e.g. dark matter boson cloud) at second-order are necessary for proper modelling.
of EMRIs, so that reliable waveforms are available to test for fundamental physics.

- What PN accuracy is required for supermassive comparable-mass BH binaries so that we can perform precision tests of fundamental physics with LISA?
- Development of PN waveforms in beyond-GR theories may need to move beyond leading order corrections. The order to which one should go to constrain theories should be investigated.
- What is the accuracy level that numerical relativity simulations must reach in order to produce high-fidelity waveforms for LISA?
- Numerical relativity must push through existing non-convergence problems to achieve high-fidelity waveforms near merger. Moreover, code efficiency and scaling must increase considerably to achieve resolutions necessary to suppress systematic errors and to achieve template indistinguishability.
- Phenom and EOS models must be improved and calibrated against numerical relativity simulations through the entire parameter space of BBHs, and include eccentricity, arbitrary spins, and substantially different masses. Efforts must be made to include environmental effects such as dynamical friction or interactions with clouds of ultralight fields. Additionally, Phenom and EOB models must be extended to modified gravity.
- Numerical relativity efforts must expand to include more modified theories of gravity (especially those that admit a well-posed initial value problem). Efforts must be made to understand systematic errors of existing approaches that tackle ill-posed modified gravity theories.
- A combination of the effects of bosonic dark matter and modified gravity must be considered in order to be able to understand how more complex deviations from standard general relativity can take place.
- Development of waveforms for EMRIs in modified gravity must take place. In particular, further work on EMRI waveform modelling requires the consistent inclusion of radiation reaction in beyond-GR theories, additional parameters related to the modification from GR, and any environmental effects.
- An important potential source of systematics that has not been included in this section (because no work has been done on this yet), but that is surely important to ensure the robustness of tests of GR are *instrumental systematics*. These are uncertainties due to calibration errors, data gaps, or other issues with the instrument itself.

8 Future directions

This white paper has been focused on the fundamental physics we can extract from LISA data. The white paper first discussed tests of the theoretical pillars of GR, such as the speed of propagation of GWs and the GW memory. The white paper then moved on to tests of the nature of BHs, discussing constraints on deviations from the
Kerr hypothesis and other tests. Next, the white paper summarized tests of GR related to fundamental questions in cosmology, such as tests probing models that attempt to explain dark matter and others that attempt to explain dark energy (as an alternative to a cosmological constant). The white paper then concluded with a discussion of generic tests of GR, and the effects of systematics in all of these tests due to mismodelling and the influence of astrophysical environments.

As we hope this white paper demonstrates, there is a tremendous amount of work yet to be done before LISA flies in order to fully realize and exploit LISA’s scientific potential in this area. The most pressing questions have been discussed at the end of each of the sections summarized above, but let us here touch on a few of the key points. Regarding tests of GR, the most burning questions all involve the development of more detailed models for the GWs emitted in specified theories, including not just for the inspiral phases, but also the merger and the ringdown for comparable-mass binaries and EMRIs.

Much more work is also still needed in the context of LISA tests on the nature of BHs. Of particular note is the development of a systematic method to carry out Kerr hypothesis tests through the extraction of multipole moments of SMBHs in binaries. Moreover, the development of accurate waveform models to study the effect of ultralight bosonic fields around SMBHs, as well as the inspiral, merger and ringdown of exotica is needed. Regarding the latter, further studies to determine whether all exotica are a priori possible or whether some theoretical restrictions should be imposed is also important to limit the number of possibilities to study.

The extraction of fundamental physics information related to cosmology with LISA data will also require that we work hard to address several burning questions. Perhaps one of the most important ones is the construction of accurate waveform models that incorporate the effect of ultralight and heavier dark matter fields in the inspiral, merger and ringdown of compact objects of various mass ratios. Another important aspect that needs further studying is the effect of screening in the generation of propagating GWs, which would be particularly important when attempting to constrain modified gravity dark energy models.

Let us now discuss the burning questions related to model-independent tests of GR. As discussed in Sect. 6, perhaps the most important action item here is the development of model-independent modifications to waveform models in all three phases of coalescence. The merger phase, in particular, is not well-understood in enough modified theories to allow for a model-independent parameterization. Another important action item is related to the extension of LIGO/Virgo tests (Abbott et al. 2020a) to the LISA realm, such as the residual test and polarization tests. These tests will be different with LISA because LISA will be sensitive to a large number of sources that will be on during the entire observation, and because of the motion of LISA around the Sun.

Perhaps one of the most important points to consider is to what extent tests of GR will be affected by waveform systematics or other systematics that could be induced by astrophysical environments. Waveform systematics will be particularly problematic for SMBH signals, because of their expected high SNRs. But this could be a problem even for intermediate mass-ratio systems, for which we may not have sufficiently accurate generic waveforms yet. Waveform systematics may also be a
problem for EMRIs, although one expects second-order waveform calculations to be complete before LISA flies. However, astrophysical systematics may still be a particular problem for EMRIs, creating a floor to our ability to test GR.

Given all of this, it is hopefully clear that a great amount of work is still needed to extract the most fundamental physics from LISA data and to ensure such inferences are robust. A close collaboration between the LISA fundamental physics WG, and other WGs, such as the waveform modelling and the cosmology, is called for and will be paramount to successfully overcome all these difficulties. We remain nonetheless optimistic that through the infrastructure of the LISA Consortium these collaborations can be organized and structured, so that we can get the most science out of the data, when LISA flies.

Acknowledgements E. Berti is supported by NSF Grants No. PHY-1912550 and AST-2006538, NASA ATP Grants No. 17-ATP17-0225 and 19-ATP19-0051, NSF-XSEDE Grant No. PHY-090003, and NSF Grant PHY-20043. D. Blas is supported by a ‘Ayuda Beatrix Galindo Senior’ from the Spanish ‘Ministerio de Universidades’, grant BG20/00228. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. The research leading of to these results has received funding from the Spanish Ministry of Science and Innovation (PID2020-115845GB-100/AEI/10.13039/501100011033). K. Clough is supported by funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 693024). A. Cárdenas-Avendaño acknowledges funding from the Fundación Universitaria Konrad Lorenz (Project 5INV1) and from Will and Kacie Snellings. M. Crisostomi and E. Barausse are supported by the European Union’s H2020 ERC Consolidator Grant “GRavity from Astrophysical to Microscopic Scales” (Grant No. GRAMS-815673). P. Fleury received the support of a fellowship from “la Caixa” Foundation (ID 100010434). The fellowship code is LCF/BQ/PI19/11690018. C. Herdeiro thanks the support of the Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/ 04106/2020, the projects PTDC/FIS-OUT/28407/2017, CERN/FIS-PAR/0027/2019, PTDC/FIS-AST/ 3041/2020 and the European Union’s Horizon 2020 research and innovation (RISE) programme H2020- MSCA-RISE-2017 Grant No. FunFiCO-777740. P. Pani and E. Maggio acknowledge financial support provided under the European Union’s H2020 ERC, Starting Grant agreement no. DarkGRA–757480, and under the MIUR PRIN and FARE programmes (GW-NEXT, CUP: B8412000100001), and support from the Amaldi Research Center funded by the MIUR program “Dipartimento di Eccellenza” (CUP: B81I18001170001). N. Frusciante was supported by Fundación para a Ciência e a Tecnologia (FCT) through the research grants UIDB/04434/2020, UIDP/04434/2020, PTDC/FIS-OUT/29048/2017, CERN/FIS-PAR/0037/2019, the FCT project “CosmoTests—Cosmological tests of gravity theories beyond General Relativity” with ref. number CEECIND/00017/2018 and the FCT project “BEYLA –BEYond Lambda” with ref. number PTDC/FIS-AST/0054/2021. L. Lombriser was supported by a Swiss National Science Foundation Professorship grant (No. 170547). S.N. acknowledges support from the Alexander von Humboldt Foundation. D.N. acknowledges support from the NSF Grant No. PHY-2011784. R.B. acknowledges financial support from the FCT – Fundação para a Ciência e a Tecnologia, I.P., under the Scientific Employment Stimulus - Individual Call - 2020.00470.CEECIND. V. Paschalidis acknowledges support from NSF Grant PHY-1912619 and NASA Grant 80NSSC20K1542 to the University of Arizona. B.S.S. is supported by NSF grants No. AST-2006384 and PHY-2012083. C.F.S. is supported by contracts ESP2017-90084-P and PID2019-106515GB-I00/AEI/10.13039/501100011033 (Spanish Ministry of Science and Innovation) and 2017-SGR-1469 (AGAUR, Generalitat de Catalunya). T. P. S. acknowledges partial support from the STFC Consolidated Grant No. ST/P000703/1. M. Ruiz acknowledges support from NASA Grant 80NSSC17K0070 to the University of Illinois at Urbana-Champaign. I.D. Saltas is supported by the Greek Science Foundation GAÇR, Grant No. 21-16583M. N. Stergioulas is supported by the ESA Prodec grant PEA:4000132310 “LISA Stochastic Signals Analysis Pipeline”. F.V. acknowledges partial support from CNES. K.Y. acknowledges support from NSF Grant PHY-1806776, NASA Grant 80NSSC20K0523, a Sloan Foundation Research Fellowship and the Owens Family Foundation. K.Y. would like to also acknowledge support by the COST Action GWverse CA16104 and JSPS KAKENHI Grants No. JP17H06358. N. Yunes acknowledges support from NASA Grants No. NNX16AB98G, 80NSSC17M0041 and 80NSSC18K1352, NSF Award No. 1759615, and the Simons Foundation through

© Springer
MPS Award Number 896696. D.D. acknowledge financial support via an Emmy Noether Research Group funded by the German Research Foundation (DFG) under grant no. DO 1771/1-1.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References


