Black holes in General Relativity and beyond

Enrico Barausse\textsuperscript{1,2,3}

\textsuperscript{1}Institut d’Astrophysique de Paris, CNRS \& Sorbonne Universités, UMR 7095, 98 bis bd Arago, 75014 Paris, France
\textsuperscript{2}SISSA, Via Bonomea 265, 34136 Trieste, Italy and INFN Sezione di Trieste
\textsuperscript{3}IFPU - Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy

The recent detections of gravitational waves from binary systems of black holes are in remarkable agreement with the predictions of General Relativity. In this pedagogical mini-review, I will go through the physics of the different phases of the evolution of black hole binary systems, providing a qualitative physical interpretation of each one of them. I will also briefly describe how these phases would be modified if gravitation were described by a theory extending or deforming General Relativity, or if the binary components turned out to be more exotic compact objects than black holes.

I. INTRODUCTION

The binary black hole (BH) systems detected by ground based gravitational wave (GW) interferometers (first by the two LIGOs alone [1–5], then by the LIGO-Virgo network [6, 7]) are now ten, and their signals have been found to be in excellent agreement with the predictions of General Relativity (GR). While GR had previously passed experimental tests with flying colors in weak field and/or mildly relativistic regimes (in the solar system [8] and in binary pulsars [9, 10]), the LIGO-Virgo detections provide for the first time evidence that GR is viable in the strong gravity, highly relativistic and dynamical regime relevant for BH binaries [3, 11]. Moreover, they also support the hypothesis that the binary’s components, as well as the merger remnant, are really BHs, as opposed to more exotic compact objects [1].

Nevertheless, as always in physics, these statements on the correctness of GR and on the BH hypothesis come with several caveats, and they are only correct within the “experimental errors” affecting the measurements. In other words, while large non-perturbative effects in the GW signal seem to be ruled out by current observations [3, 11], deformations of the general relativistic signal that are either sufficiently small and/or in the frequency ranges where LIGO and Virgo are least (or not at all) sensitive may still be in agreement with observations.

The purpose of this note is to briefly scope out where deviations from GR or from the BH hypothesis might still hide in the GW signal from binary systems, while remaining in agreement with observations. I will do so by conventionally splitting the signal in four parts: an early low-frequency portion produced by the binary’s inspiral (Sec. II); one produced by the merger (Sec. III) with the subsequent ringdown phase (Sec. IV); and a post-ringdown phase produced by putative exotic near horizon physics (Sec. VI). (This fourth phase is of course absent in GR.) I will qualitatively interpret each part by relying on physical intuition from the test particle limit, Newtonian physics and BH perturbation theory.

II. THE INSPIRAL

At low orbital (and GW) frequencies, i.e. at large separations, the evolution of the binary is driven by the emission of GWs. The latter backreact on the system by carrying energy and angular momentum away from it, and thus cause the binary to slowly inspiral to smaller and smaller separations. In GR, this “(gravitational) radiation reaction” is well described, in the limit of small relative binary velocities $v \ll c$, by the quadrupole flux formula, according to which the energy flux emitted by a binary with masses $m_1$ and $m_2$ and separation $r$ is given by [12]

\[
\dot{E}_{\text{quadrupole}} = \frac{32G^2}{5c^5} \left( \frac{Gm_1m_2}{r^2} \right)^2 \left( \frac{v}{c} \right)^2.
\]  

For a quasi-circular binary, Kepler’s law yields $v = (GM/r)^{1/2}$ (with $M = m_1 + m_2$), and the GW frequency $f$ is given by twice the orbital frequency $f_{\text{orb}} = v/(2\pi r)$. One can then relate the GW frequency to the binary’s total energy $E$ (in the center of mass frame) by $E = -Gm_1m_2/(2r)$, and re-express Eq. (1) as [12]

\[
f \approx \frac{96G^{5/3}}{5c^5} \pi^{8/3} M^{5/3} f^{11/3},
\]  

where $M = M \nu^{3/5}$, with $\nu = m_1m_2/M^2 \leq 1/4$, is the chimp mass. A simple lesson to be drawn from this equation is that GW observations of the inspiral determine the chirp mass to high accuracy, simply by measuring the frequency and its rate of change. Indeed, the first GW detection (GW150914 [1]) determined the chirp mass of the binary source to be $M \approx 30M_\odot$, which translates into a total mass $M \gtrsim 70M_\odot$.

The quadrupole formula can be generalized by performing a “post-Newtonian” (PN) expansion, i.e. a perturbative expansion of the field equations in the small quantity $v/c \ll 1$ (see e.g. Ref. [13] for a review of the PN formalism). By solving the field equations under that approximation, one can find expressions for the fluxes that reduce to Eq. (1) in the limit of low frequencies (when $v \ll c$), but which also include subdominant terms in
$v/c$ (e.g. the mass-octupole and current-quadrupole contributions). These terms become important as the binary’s separation shrinks. Similar PN-expanded expressions can be found not only for the GW fluxes, but also for the gravitational waveforms and for the Hamiltonian describing the conservative interaction between the two binary components.

In particular, the PN expansion allows one to account for the effect of BH spins, which appear both in the conservative and dissipative sectors. One remarkable effect of BH spins during the inspiral is given by the modulations of the GW amplitude that they cause when they are misaligned with the binary’s orbital angular momentum \cite{14}. Indeed, save for special configurations giving rise to transitional precession \cite{14}, the spins and the orbital angular momentum undergo simple precession around the total angular momentum of the system, whose direction remains approximately constant during the binary’s evolution (including even the merger ringdown phase) \cite{14, 15}. Detecting these spin precession modulations, as well other spin related effects in the gravitational waveforms (e.g. the dependence of the plunge-merger frequency on the spins, which we will discuss below in Sec. III, and the enhancement of the GW amplitude caused by large spins aligned with the orbital angular momentum), one can, at least in principle, measure BH spins. Indeed, in two GW detections (GW151226 and GW170729) there seems to be evidence in favor of non-zero, probably precessing spins \cite{2, 3, 7}. Much more accurate measurements of BH spins (to within 1–10% or better) will also become possible with space based detectors such as LISA \cite{16, 17}.

As one moves away from GR to consider more general gravitational theories, the structure of the PN expansion also gets modified. The LIGO-Virgo collaboration has performed phenomenological tests of GR by “deforming” the PN expressions for the GW amplitude and phase as functions of frequency. In particular, they allowed for PN coefficients (i.e. coefficients akin e.g. to the 32/5 appearing in Eq. (1)) different from the GR predicted values, and tried to constrain the differences away from the latter with observations. This resulted in “large” deviations from GR being excluded by the data \cite{3, 11}. Note that these deformations of the PN coefficients may effectively account also for the possibility that the binary components may not be BHs but more exotic objects, which would present different tidal effects than BHs (see e.g. Ref. \cite{18}).

Another effect that would arise quite naturally from an extension/modification of GR is a low-frequency change of the fluxes and waveforms, resulting e.g. from the existence of BH dipole radiation \cite{19}. Within GR, gravitational monopole and dipole radiation are forbidden by the conservation of the stress energy tensor, i.e. respectively by the conservation of energy and linear momentum. This is similar to what happens in electromagnetism, where monopole emission is forbidden by the conservation of the electric charge. Suppose now, however, that the theory of gravity describing Nature is not GR (which only possesses a spin-2 graviton, i.e. which only describes gravity by means of a metric), but some more general theory in which gravity is described by a metric and additional gravitational degrees of freedom (e.g. a scalar field). To avoid the appearance of unwanted fifth forces in existing experiments, these additional gravitons must be coupled very weakly (if they are coupled at all) to matter, but they may be coupled non-minimally to the metric. As a result, the coupling to matter, which is negligible in weak gravity experiments such as those performed on Earth, may re-appear in strong gravity systems, where non-linearities of the metric become large enough to mediate an effective coupling between matter and the extra gravitational fields.

One may parametrize this effective coupling (also known as “Nordtvedt effect” \cite{20–22}) by “charges” carried by compact objects such as BHs and neutrons stars.\footnote{These charges are instead negligible for stars or objects with weak internal gravity} These charges, also referred to as “sensitivities” \cite{20, 23–28}, can be object-dependent, e.g. they can be different for BHs and neutron stars, or for stars of different compactness (as they need to vanish in the small compactness limit). In this sense, they may be viewed also as parametrizing violations of the universality of free fall (the equivalence principle). The latter holds in GR, but is generally violated beyond GR, at least in strong field regimes, because of energy and momentum exchanges between matter and the extra gravitational degrees of freedom, mediated by the large metric perturbations.

Now, just as in Maxwell’s theory a binary system of objects carrying unequal electric charges emits dipole radiation, a quasi-circular binary of BHs or neutron stars carrying unequal sensitivities $s_1 \neq s_2$ will emit (gravitational) dipole radiation. Such an emission would be enhanced relative to the quadrupole flux of Eq. (1) by a term $\sim (s_1 - s_2)^2 (v/c)^{-2}$, i.e. it would dominate the binary evolution at low frequencies \cite{20, 23–28}.

Bounds on the existence of this dipole radiation are very strong in binary pulsar systems (see e.g. Ref. \cite{29}), but little can be inferred from those bounds about BH dipole emission, since sensitivities can be different for different classes of objects. Indeed, theories exist where the sensitivities are exactly zero for stars (including neutron stars), and non-zero only for BHs \cite{30, 31}. Bounds on BH dipole emission from LIGO-Virgo are rather loose, since these detectors cannot observe the low-frequency ($\lesssim 10$ Hz) signal from the early inspiral because of instrumental limitations (mainly due to seismic and Newtonian noise). However, much more stringent bounds will be made possible by the launch of the ESA-led mission LISA \cite{16}, a space-borne GW detector targeting the mHz frequency band. This will allow for observing the low frequency inspiral of LIGO-Virgo sources before they are detected from the ground \cite{32}, and also to observe massive BH...
III. THE MERGER

Under the effect of radiation reaction, a BH binary slowly evolves along quasi-circular orbits with decreasing separation. Unlike in Newtonian mechanics, where circular orbits can exist with arbitrarily small radii around a point mass, in GR there exists an innermost stable circular orbit (ISCO), inside which circular orbits can still exist but are unstable [34]. ISCOs do indeed exist for test particles (i.e. in the limit $m_2/m_1 \to 0$) around Schwarzschild and Kerr BHs, and they are present also away from the test particle limit, e.g. when one includes next-to-leading order corrections in $\mathcal{O}(m_2/m_1)$, or when one expands the dynamics of generic mass ratio binaries in a PN series (in the latter case, the ISCO corresponds to the circular orbit with minimum energy, as defined by the PN Hamiltonian mentioned in the previous section) [35–37].

The position of the ISCO, its energy and its angular momentum depend on the parameters of the BH binary, and most notably on the spins. This is easily seen in the test particle limit, i.e. for the ISCO of the Kerr geometry. As the spin of the Kerr BH (projected on the direction of the particle’s orbital angular momentum) increases, the ISCO moves inwards, while its energy and angular momentum decrease [38]. In particular, the ISCO for a test particle moving on retrograde orbits (with respect to the BH rotation) in a Kerr spacetime lies at larger radii than for a particle on prograde orbits [38]. This behavior is of course a manifestation of the more general “frame dragging” effect of GR, whereby matter near BH horizons tends to co-rotate with the BH as seen from infinity. It is also the reason why the radiative efficiency of thin accretion disks, whose inner edge lies at the ISCO of the central BH, is an increasing function of the BH spin, and why their inner edge gets closer and closer to the BH as the spin of the latter increases. Both facts are crucial to estimate the spin of BH candidates by electromagnetic observations such as continuum fitting and iron-K$\alpha$ lines [39, 40]. Similarly, any modification of the BH geometry causing an ISCO shift (as a result of a deviation of the gravity theory from GR, or simply because the central object is more exotic than a run-of-the-mill BH) can be constrained, at least in principle, by the same electromagnetic techniques [41, 42].

As in the test-particle limit, the position of the effective ISCO of a comparable mass binary moves to smaller separation for larger BH spins [35, 37, 43]. Once the binary reaches the ISCO, it cannot transition to a stable circular orbit and therefore plunges and merges. The plunge-merger phase takes place in a dynamical time $\sim GM/c^3$, i.e. it is much shorter than the inspiral phase. For this reason, the GW fluxes emitted in the plunge-merger phase can be neglected, to first approximation, with respect to those emitted during the inspiral, i.e. most of the energy and angular momentum emitted by a BH binary is lost from large separations down to the effective ISCO of the system. Therefore, as in the case of thin accretion disks (whose radiative efficiency increases with the spin of the central BH), the GW emission efficiency of BH binaries is a growing function of the spins [44] (because larger spins imply smaller ISCO separations, and thus longer inspiral phases and larger integrated GW fluxes).

One can also attempt to estimate the final spin of the BH remnant produced by the merger that follows the plunge by conserving angular momentum. Consider for instance a test particle falling into a BH from a quasi-circular equatorial inspiral. Neglecting the fluxes during the plunge-merger, and also neglecting the fraction of the binary’s mass emitted in GWs (see Ref. [45] for a treatment that avoids this second hypothesis), the final spin of the BH is

$$a_{\text{fin}} = \frac{1}{1 + q^2} \left( a_1 + a_2 q^2 + L_{\text{ISCO}}(a_1) q \right),$$

(3)

where $q = m_2/m_1 \ll 1$ and $L_{\text{ISCO}}(a_1)$ is the ISCO angular momentum (normalized by $Gm_1m_2/c$ to make it dimensionless) as a function of the central (“primary”) BH spin $a_1$. Let us look for instance at the simple case of non-spinning BHs, $|a_1| = |a_2| = 0$, and extrapolate from the test-particle limit to equal masses, $q = 1$. The final predicted spin is then $a_{\text{fin}} \approx 0.866$, which is quite off compared to $a_{\text{fin}} \approx 0.686$ predicted by numerical relativity simulations of BH mergers [46]. However, if the primary BH spin $a_1$ used to compute $L_{\text{ISCO}}$ is replaced by $a_{\text{fin}}$, one is left with an algebraic equation in $a_{\text{fin}}$, which can be solved e.g. by iteration [47]. Remarkably, this procedure gives an excellent estimate for the final spin, $a_{\text{fin}} \approx 0.66$ [47]. This iterative method can be generalized to generic (even misaligned) spins and mass ratios, and turns out to be always in good agreement with numerical relativity simulations [45, 47]. (See also Refs. [15, 48–51] for other procedures to estimate the final spin based on similar physical arguments).

At least two lessons can be drawn from this method to estimate the final spin. The first is that there seems to be a mapping between the comparable mass BH binary problem and the dynamics of a test particle in a Kerr spacetime with spin equal to the final spin. The existence of this mapping is hardly surprising because similar mappings between the comparable mass and test-particle two body problems are known to exist, e.g. in Newtonian mechanics (where a binary can be mapped into a test particle of mass $m_1m_2/M$ around a point mass $M = m_1 + m_2$), in quantum electrodynamics (where one can map the energy levels of positronium into those of hydrogen [52]) and in PN theory itself (with the effective one body — EOB — model, applicable to the inspiral phase of non-spinning [53] and spinning [54, 55] BH binaries). Indeed, the success of the iterative method of Ref. [47] to compute the final spin, which we have out-
lined above, provides justification for extending the EOB model through the merger ringdown phase [56].

Secondly, it should be stressed that if deviations from GR or from the BH paradigm itself were to change the position and angular momentum of the effective ISCO of a BH binary, then Eq. (3) and the “amended” extrapolation procedure of Ref. [47] would produce a different final BH spin. While direct measurements of the final spin by LIGO and Virgo are currently impossible because of the low signal-to-noise ratio (SNR) of the ringdown [11], future detectors such as LISA will allow for measurements of the ringdown with SNR of several hundreds [57], thus making direct spin measurements possible. Moreover, even with the current LIGO-Virgo detections one can attempt to perform consistency tests between the inspiral part of the waveforms (which depends on the individual masses and spins) and the plunge-merger-ringdown part (which depends on the final mass and spin). These tests provide at present no signs of large deviations away from GR [3, 11] (see also the next section).

Another notable constraint on the nature of the components of the LIGO-Virgo binaries can be obtained from simple physical arguments involving the merger part of the waveform. As mentioned previously, the GW150914 BH merger event has chirp mass $M \sim 30M_\odot$ ($M \sim 70M_\odot$), as estimated from the inspiral. The peak of the GW amplitude, corresponding to the merger, is at about 150 Hz, corresponding to an orbital frequency of $\sim 75$ Hz and a separation (assuming a binary with roughly equal masses) of $\sim 350$ km [1]. Since $GM/c^2 \sim 200$ km, this means that the components of GW150914 merge when they are very close (in gravitational radii). If they were more exotic compact objects with sizes $R_1 + R_2 \gg GM/c^2$, they would be interacting well before reaching a separation of 350 km, which would trigger an earlier merger than for BHs. This seems to disfavor the hypothesis that the binary components could be e.g. boson stars [1], unless their compactness is very close to that of BHs (which in turns puts restrictions on the potential of the bosonic degrees of freedom). These bounds are much stronger than those for BHs in the massive/supermassive range, which come e.g. from tracking S-stars around the massive BH at the center of our Galaxy [58], and which still allow for a compactness $GM/(Rc^2) \gtrsim 10^{-3}$. Moreover, note that deviations from GR may also trigger early mergers, especially in the massive/supermassive range, which come e.g. from star systems.

IV. THE RINGDOWN

When the two binary components coalesce, the remnant is initially highly perturbed as a result of the violent merger process. As the system evolves toward an asymptotically stationary state, the amplitude of these perturbations decreases and eventually becomes sufficiently small for them to be described by linear theory. Within GR, linear perturbations $h_{\mu\nu}$ over a generic vacuum spacetime with metric $g_{\mu\nu}$ satisfy the wave equation [61, 62]

$$\Box h^{\alpha\beta} + 2R^{\alpha}_{\mu} \beta h^{\mu\nu} = 0,$$  \hspace{1cm} (4)

with $h_{\mu\nu} \equiv h_{\mu\nu} - h^{\mu}_{\nu} g_{\mu\nu}/2$ obeying the Lorentz gauge condition $\nabla_{\mu} h^{\mu\nu} = 0$, and $\Box \equiv g^{\mu\nu} \nabla_{\mu} \nabla_{\nu}$. However, linear perturbations over a Kerr background of gravitational radius $r_g = GM/c^2$ ($M$ being the mass) and spin parameter $a \in [0, 1]$ are better described by the Newman-Penrose scalars [63]

$$\Psi_0 = -C_{\mu\nu\lambda} l^{\mu} m^\nu l^\lambda m^\sigma,$$  \hspace{1cm} (5)

$$\Psi_4 = -C_{\mu\nu\lambda\sigma} n^{\mu} m^\nu l^\lambda m^\sigma,$$  \hspace{1cm} (6)

where $C_{\mu\nu\lambda\sigma}$ is the Weyl curvature tensor, and $l, n, m, m^\sigma$ constitute a (complex) null tetrad, defined at each spacetime point. Note that $\Psi_0$ and $\Psi_4$ can be thought of as describing ingoing and outgoing GWs.

Because of the symmetries of the Kerr metric (and in particular the existence of a Killing tensor, besides the Killing vectors associated to stationarity and axisymmetry), the linear perturbation equations are dramatically simplified by a decomposition in Fourier modes and in spin-weighted spherical harmonics $S_{lm}(\theta, \phi)$ (with $s = 2$ and $s = -2$ for $\Psi_0$ and $\Psi_4$), i.e. by adopting the ansatz [64]

$$\psi(t, r, \theta, \phi) = \frac{1}{2\pi} \int e^{-i \omega t} \sum_{l=|s|}^{\infty} \sum_{m=-l}^{l} e^{im\phi} S_{lm}(\theta) R_{lm}(r) d\omega,$$  \hspace{1cm} (7)

where $\psi$ stands for either $\Psi_0$ or $\rho^{-4} \Psi_4$, with $\rho \equiv 1/(r - iar, \cos \theta)$. Indeed, with this decomposition the perturbed field equations reduce to a single ordinary differential equation for the radial function $R_{lm}$ [64]:

$$\Delta R_{lm}^2 R_{lm} + 2(s + 1)(r - r_s) \partial_r R_{lm} + V R_{lm} = 0,$$  \hspace{1cm} (8)

with $\Delta = r^2 - 2rr_s + a^2 r_s^2$ and $V$ a (complex) potential depending on $\omega, r, M, a, s, l$ and $m$. Moreover, by introducing the tortoise coordinate $r_\*$, which ranges from $-\infty$ at the horizon to $+\infty$ at spatial infinity, one can zoom in onto the near horizon region, and the equation becomes reminiscent of the one-dimensional Schrödinger equation with a potential barrier.

This equation should be solved with boundary conditions corresponding to no radiation incoming from infinity (i.e. the system is isolated) and no radiation exiting the event horizon (since nothing can escape the event horizon in GR). The boundary conditions (ingoing at the event horizon and outgoing at infinity) select a discrete spectrum of complex frequencies $\omega$ (c.f. the similar problem of the one-dimensional Schrödinger equation with a rectangular potential). The frequencies are known as “quasi-normal mode” (QNM) frequencies [65], and for non-extremal Kerr BHs they have strictly negative imaginary part, i.e. the modes that they describe
are damped. These damped oscillations constitute the “ringdown” part of the GW signal from a BH binary.

From the procedure whereby they are derived, which we have just outlined, it is clear that for given multipole numbers $\ell, m$, the QNM frequencies only depend (in GR) on the BH mass and spin. This is in a sense a consequence of the no-hair theorem of GR, which states that BHs can only be described by their mass and spin, and that they do not carry any additional “charges” or “hairs” [66]. (Note that BHs in GR can carry an electric charge, but for astrophysical BHs that is expected to be zero or negligible [67].)

The direct measurement of the QNM frequencies and decay times from the LIGO-Virgo detections is a difficult task because of the low SNR of the ringdown. Indeed, at present one can at most check the compatibility of the ringdown signal with the predictions of GR [3, 11], i.e. with the QNM frequencies and decay times that would be expected based on the final remnant spin predicted via Eq. (3) (or a similar method), once the masses and spin measurements from the inspiral part are factored in.

Stronger ringdown signals would allow not only for measuring, directly from the data, the frequency and decay time of the dominant, least damped QNM (the $\ell = m = 2$ mode for a quasicircular binary), but also the first subdominant mode (usually the $\ell = m = 3$ or $\ell = m = 4$, mode depending on the binary’s parameters). This would permit performing a null test of the no-hair theorem [68]. Indeed, as mentioned above, all QNM frequencies and decay times depend only on $M$ and $a$ in GR. Therefore, one could use the $\ell = m = 2$ measured frequency and decay time to infer $M$ and $a$, use those to predict the frequency/decay time of the second QNM, and finally check the agreement with the measured value(s).

This kind of test, however, would require not only that the ringdown signal be detectable (i.e. with SNR larger than $\sim 8$), but also that the SNR be sufficiently large to resolve both the dominant and first subdominant QNM. In practice, this requires an SNR of at least $\sim 90$ in the ringdown phase alone [57, 69], which will be achieved routinely with LISA, but which will require Voyager-type third generation detectors on Earth [57] (unless one manages to stack several ringdown signals together by rescaling them [70]).

Several observations can be made at this stage. First, in theories of gravity extending GR, not only can the spin-2 QNMs differ from GR (as a result e.g. of BHs carrying hairs, or simply because of the different field equations at linear order [71]), but there may also be QNMs with different helicities (e.g. scalars, vectors, or even additional tensor modes in bimetric gravitational theories) [72]. These QNMs would give rise to different polarizations in the response of a network of GW detectors, and would be in principle distinguishable from purely spin-2 modes [73]. However, as mentioned previously, non-GR gravitons must be coupled very weakly to a detector (if coupled at all) to ensure the absence of unwanted fifth forces in experiments testing the (weak)

equivalence principle on Earth. Therefore, the coupling of QNMs of helicity $s \neq \pm 2$ to the detector will be strongly suppressed (see e.g. Ref. [59] for an explicit example).

Secondly, let us note that in the geometric optics limit (i.e. $\ell \gg m \gg 1$) the equations of linear perturbation theory [Eq. (8)] must reduce to the null geodesics equation. This is most easily seen by inserting the ansatz $h_{\mu\nu} \approx A_{\mu\nu} \exp(iS)$ into Eq. (4), and expanding in the limit of large $\partial_j S$ (i.e. in the limit of large frequencies and wavenumbers). This gives the null condition $g^{\mu\nu} \partial_\mu \partial_\nu S = 0$, which coincides with the Hamilton-Jacobi equation for massless particles. From the derivative of this condition, one then obtains the null geodesics equation (c.f. e.g. section 7.8 of Ref [74] for details). This result is of course hardly surprising as it simply amounts to saying that the gravitational wavefronts follow null geodesics (i.e. GWs travel at the speed of light), but it has noteworthy implications for the physics of QNMs. Indeed, since the linear perturbation equations must reduce to the null geodesics equation in the geometric optics limit, the potential $V$ appearing in the QNM master equation (8) must reduce to the effective potential governing the motion of null geodesics in Kerr in the limit $\ell \gg m \gg 1$. In particular, in that limit the position of the peak of the potential must asymptote to the same location as the (unstable) circular photon orbit, and the frequencies of the QNMs are simply given by linear combinations of the orbital and frame-dragging precession frequencies of the circular photon orbit [75–77]. (Note that the frame-dragging precession frequency only appears in the QNM expressions for spinning BHs.) Moreover, one can show that the decay times of the QNMs for $\ell \gg m \gg 1$ are related to the Lyapunov exponents of null geodesics near the circular photon orbit, which in turn depend on the curvature of the effective potential for null geodesics near its peak [75, 76]. An intuitive interpretation of these results is that the QNMs are generated at the circular photon orbit and slowly leak outward, since that orbit is unstable to radial perturbations.

From this point of view, it is clear that the physics of the QNMs is very similar (if not the same) as that of the circular photon orbit, and electromagnetic experiments aiming to explore the latter present clear synergies with GW detectors. One example is given by the Event Horizon Telescope [78], an ongoing very long baseline radio interferometry experiment aiming to observe the near horizon region of Sgr A* (the massive BH at the center of our Galaxy), and which has recently provided the first “shadow image” of the horizon of the massive BH at the center of M87 [79]. Another implication of realizing that QNMs are generated near the circular photon orbit is that BH mimickers that deviate significantly from the Schwarzschild/Kerr geometry in that region may be severely constrained once GW interferometers detect the ringdown part of a binary’s signal with sufficiently large SNR.
V. THE POST-RINGDOWN

From the derivation of the QNMs outlined in the previous section, it is clear that if the boundary conditions at the event horizon are not perfectly ingoing, the QNM frequencies and decay times will in general change. Partially or fully reflective boundary conditions may in fact be physically relevant in case the remnant is not a BH but a more exotic horizonless object (e.g. a wormhole [18], or an ensemble of horizonless quantum states [80–86]), perhaps motivated by attempts to solve the information loss paradox [87]. Other possible deviations from the GR event horizon paradigm may arise in theories that violate Lorentz symmetry in the purely gravitational sector [28, 88].

In the presence of such deviations from purely ingoing boundary conditions, part of the GW radiation generated at the peak of the potential (i.e. roughly at the circular photon orbit, where QNMs are produced) is reflected off the horizon. Part of this reflected radiation then tunnels through the potential barrier at the circular photon orbit (eventually reaching infinity), while the rest is reflected back again toward the horizon, where the process can restart. Therefore, after the ringdown described in the previous section, which corresponds to radiation generated at the potential’s peak and leaking outward, there appears a series of “echoes” delayed (relative to the initial ringdown signal) by multiples of twice the light travel time between the circular photon orbit and the effective “mirror” at the horizon [67, 89–92].

A similar phenomenon can take place in the presence of matter far away from the BH, which can act as a partially reflective mirror for outgoing radiation. Indeed, the matter’s gravitational potential can give rise to a “bump” in the potential $V$ appearing in Eq. (8) [67, 89]. In this case, after the initial ringdown “burst”, one would have echoes delayed by multiples of twice the light travel time between the circular photon orbit and the far away matter. An example from Ref. [89] is shown in Fig. 1. As can be seen, the initial ringdown transient is governed by the QNM frequencies and decay times of the pure Kerr/Schwarzschild system (without any matter), while only after a delay of twice the light travel time between the circular photon orbit and the matter does the signal start to be described by the QNMs of the composite system. The same applies to the case of a BH mimicker with near horizon reflectivity [18, 67, 89]: the initial ringdown transient has spectrum comprised of the Kerr/Schwarzschild QNMs, while the “true” QNMs (accounting for the modified boundary conditions at the horizon) only become apparent after a time roughly equal to twice the light travel time between the near horizon mirror and the potential’s peak.

The presence of echoes in the post-ringdown signal of some of the LIGO-Virgo systems has been put forward by Ref. [93–95], but the claim is highly controversial as later analyses have found it to have low significance [96, 97]. Such a low significance (if any) is not surprising since the systems detected by LIGO-Virgo have very small ringdown SNR. Nevertheless, third generation ground-based detectors and LISA will allow for measuring the ringdown phase with much higher SNR [57], which could open the way to detecting echoes, if there are any.

A more robust way to assess the possible presence of echoes is to look at their interaction with BH spins. Energy conservation dictates that the echoes should have successively decreasing amplitude around a non-spinning BH, but their amplitude can actually increase around BHs with sufficiently large spins. This is due to a phenomenon called super-radiance [98–101], which is strictly related to the existence of an ergoregion (i.e. a near-horizon region where modes with negative energy can exist as seen by an observer at infinity), and whereby echoes can extract rotational energy from the spinning BH. Indeed, BHs with sufficiently large spins and surrounded by a “mirror” at the horizon become linearly unstable because of super-radiance, meaning that the QNMs of the system are not damped but exponentially growing (at the expense of the BH’s rotational energy). This instability is also known as “BH bomb” instability [102, 103], and if present would have important consequences for the background of unresolved GWs that LIGO and Virgo are trying to detect [104]. Indeed, if all astrophysical BH candidates in the universe (in binaries or isolated) had a perfectly reflective mirror at the horizon, or if they had no horizon at all and they let gravitational radiation go through them undamped, under reasonable assumptions on the distributions of their spins, their exponentially growing QNMs would sum up to create a stochastic background of GWs orders of magnitude larger than the current upper bounds from LIGO-Virgo [104]. This allows us to exclude that all BH candidates in the universe have a perfectly reflective horizon surface, or that they let gravitational radiation go through them undamped. Indeed, LIGO-Virgo at design sensitivity will only allow for 1% or less of the BH candidate population to have such properties. The constraint can be extended to partially
reflective mirrors or to objects that damp passing radiation. However, a transmissivity — or energy loss through the object — of at least 60% (6%) can quench the instability for any spin (for spins \(a \lesssim 0.9\)), thus evading these constraints [104, 105].

Let us note that a similar BH bomb super-radiant instability affects BHs with sufficiently large spins and a mirror at large distances from the BH [102, 103]. Indeed, if this mirror has sufficiently large reflectivity, it can focus gravitational radiation back to the BH. This reflected radiation would in turn be scattered back by the effective potential of Eq. (8), and so on. Since the ergoregion of Kerr BHs extends (on the equatorial plane) to \(r = 2r_s\) in Boyer-Lindquist coordinates, while the circular photon orbit has radius that goes to \(r_s\) in the same coordinates for \(a \to 1\) [38], it is clear that while bouncing back and forth between the potential’s peak and the mirror the wave can get amplified by producing negative energy ergoregion modes, as in the case of a mirror at the horizon.

While matter around the BH will not typically provide sufficient reflectivity for this instability to pick up, a physical implementation of this idea is provided by ultralight bosonic degrees of freedom. Indeed, such boson fields (be them scalar, vectors or tensors) can become super-radiance unstable around BHs with sufficiently large spin, provided that their Compton wavelength is comparable to the BH horizon radius [106]. The instability is akin to the BH bomb with an external mirror, where the mirror is provided here by the mass, which constrains the boson field near the BH by allowing for bosonic bound states. As a result of the instability, the BH spin quickly decreases, and the angular momentum is transferred to the boson field, which evolves into a rotating dipolar condensate. This condensate is in turn expected to emit almost monochromatic GWs, as can be understood simply from the quadrupole formula [107–113]. Summing up the monochromatic waves that would be produced by all the (isolated) BHs in the universe, under reasonable assumptions for their spins, one obtains a stochastic background of GWs that exceeds the current LIGO-Virgo bounds by several orders of magnitude if the boson’s mass is in the range \(\sim eV\) [114, 115]. Further constraints may come from trying to detect the monochromatic GWs produced by the bosonic condensate directly, either in isolated BHs, or in the remnant resulting from a BH merger [108, 113–115, 117–120].

VI. CONCLUSIONS

In this pedagogical note, I have tried to show how the GW signal from the merger of two BHs can be understood qualitatively based on very simple physical ingredients, including the quadrupole formula, which allows for understanding the low frequency inspiral and its dependence on the chirp mass; the PN precession of the spins during the inspiral; the presence of an effective ISCO, whose position depends on the BH spins as a result of “frame dragging” and which affects the overall power emitted in GWs; and the circular photon orbit and its effective potential, whose physics determines the QNMs. I have also shown how modifications of GR and/or of the BH paradigm can affect each of these ingredients and produce GW signals differing from GR. In particular, I have briefly outlined low frequency modifications to the gravitational waveform that may arise from BH dipole radiation (caused in turn by BH hairs); possible changes to the conservative dynamics, which may result in changes of the ISCO properties and therefore in a different final spin; changes in the physics of the circular photon orbit, which would affect the properties of QNMs; and changes in the boundary conditions satisfied by the QNMs at the horizon and at infinity (as a result of exotic near horizon physics, matter far from the binary, or the presence of an ultralight boson field), which would result in echoes and super-radiant instabilities.

ACKNOWLEDGMENTS

This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. GRAMS-815673; project title “GRavity from Astrophysical to Microscopic Scales”). This work was also supported by the H2020-MSCA-RISE-2015 Grant No. StronGrHEP-690904. The author would like to acknowledge networking support by the COST Action CA16104.

[1] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 116, 061102 (2016).
[2] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. Lett. 116, 241103 (2016), arXiv:1606.04855 [gr-qc].
[3] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. X6, 041015 (2016), [erratum: Phys. Rev.X8,no.3,039903(2018)], arXiv:1606.04856 [gr-qc].
[4] B. P. Abbott et al. (LIGO Scientific, VIRGO), Phys. Rev. Lett. 118, 221101 (2017), [Erratum: Phys. Rev. Lett.121,no.12,129901(2018)], arXiv:1706.01812 [gr-qc].
[5] B. P. Abbott et al. (LIGO Scientific, Virgo), Astrophys. J. 851, L35 (2017), arXiv:1711.05578 [astro-ph.HE].
[6] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. Lett. 119, 141101 (2017), arXiv:1709.09660 [gr-qc].
[118] A. Arvanitaki, M. Baryakhtar, and X. Huang, Phys. Rev. D91, 084011 (2015), arXiv:1411.2263 [hep-ph].

[119] A. Arvanitaki, M. Baryakhtar, S. Dimopoulos, S. Dubovsky, and R. Lasenby, Phys. Rev. D95, 043001 (2017), arXiv:1604.03958 [hep-ph].

[120] P. S. B. Dev, M. Lindner, and S. Ohmer, (2016), arXiv:1609.03939 [hep-ph].