Gravity in the infrared and effective nonlocal models

Enis Belgacem,[a] Yves Dirian,[b] Andreas Finke,[a] Stefano Foffa[a] and Michele Maggiore[a]

[a]Département de Physique Théorique and Center for Astroparticle Physics,
  Université de Genève, 24 quai Ansermet, CH-1211 Genève 4, Switzerland

[b]Center for Theoretical Astrophysics and Cosmology,
  Institute for Computational Science, University of Zürich, CH-8057 Zürich, Switzerland

E-mail: enis.belgacem@unige.ch, yves.dirian@ics.uzh.ch, Andreas.Finke@unige.ch, stefano.foffa@unige.ch, Michele.Maggiore@unige.ch

Abstract. We provide a systematic and updated discussion of a research line carried out by our group over the last few years, in which gravity is modified at cosmological distances by the introduction of nonlocal terms, assumed to emerge at an effective level from the infrared behavior of the quantum theory. The requirement of producing a viable cosmology turns out to be very stringent and basically selects a unique model, in which the nonlocal term describes an effective mass for the conformal mode. We discuss how such a specific structure could emerge from a fundamental local theory of gravity, and we perform a detailed comparison of this model with the most recent cosmological datasets, confirming that it fits current data at the same level as ΛCDM.

Most notably, the model has striking predictions in the sector of tensor perturbations, leading to a very large effect in the propagation of gravitational wave (GWs) over cosmological distances. At the redshifts relevant for the next generation of GW detectors such as Einstein Telescope, Cosmic Explorer and LISA, this leads to deviations from GR that could be as large as 80%, and could be verified with the detection of just a single coalescing binary with electromagnetic counterpart. This would also have potentially important consequences for the search of the counterpart since, for a given luminosity distance to the source, as inferred through the GW signal, the actual source redshift could be significantly different from that predicted by ΛCDM. At the redshifts relevant for advanced LIGO/Virgo/Kagra the effect is smaller, but still potentially observable over a few years of runs at target sensitivity.

Contents

1 Introduction

The infrared (IR) dynamics of quantum field theories with massless particles can in general be highly non-trivial. In General Relativity (GR), where the coupling constant $G_N = 1/M_{\rm Pl}^2$ has the dimension of inverse mass squared, perturbation theory is organized in powers of $G_N E^2$, with E a typical energy scale, and one might think that the IR limit $E \to 0$ is fully perturbative, contrary to the large energy limit, where an UV completion is needed at the latest at $E \simeq M_{\rm Pl}$. In fact, several lines of investigation indicate that this conclusion might be too naive. Already for pure gravity in flat space a non-trivial vacuum structure at infinity emerges, relating the asymptotic symmetries of flat space-time (the BMS group) to soft theorems and memory effects, leading to the conclusion that the vacuum in GR is not unique [1]. In de Sitter space the Feynman propagator grows without bounds for large separations, leading to IR divergences in gauge-invariant scattering processes [2]; the strongest divergence comes from the propagator of the conformal mode, a point that will be relevant in the following. Infrared divergences also appear when performing computation of physical quantities during inflation [3]. When gravity is coupled to massless fields, further non-trivial dynamics for the conformal mode of the metric, σ, arises through the conformal anomaly. The latter gives a contribution to the quantum effective action that, in flat four-dimensional space-time, is proportional to $(\Box \sigma)^2$, i.e. is fourth-order in the derivatives. In turn, this leads to a propagator for the conformal mode $G_\sigma(x, x') \propto \log(x - x')^2$, which again grows without bounds in the IR, leading to the possibility of non-trivial large-distance dynamics, including the possibility of non-trivial IR fixed points [4, 5]. IR divergences are the sign that something is missing in our understanding of the physics, and that the long-distance behavior of the theory could be different from that suggested by a perturbative analysis. A nontrivial IR dynamics could of course be relevant for understanding the origin of dark energy, and indeed the idea that quantum effects in gravity could have cosmological relevance already appeared in older works [6].

In the presence of strong IR effects a system often reacts by generating a mass scale dynamically. At first, it might seem that this is precluded for the gravitational field, since a mass term would be forbidden by diffeomorphism invariance. However, once quantum effects enter into play, the relevant quantity is no longer the fundamental action of the theory, but rather the quantum effective action. Whenever the theory has massless particles, such as the graviton in GR, the quantum effective action unavoidably develops nonlocal terms. As we will review in section 2, with nonlocal terms it is possible to construct gauge-invariant mass terms for gauge fields, and diffeomorphism-invariant mass terms for different modes of the gravitational field. The question then arises whether, in the IR limit of GR, a nonlocal mass term of this form could emerge. This question is very difficult to answer, since it basically involve non-perturbative physics,[1] and non-perturbative techniques for gravity are still far from being fully established. After recalling in section 2.3 some generic features that are important for understanding the proper way of dealing with nonlocal terms, in section 2.4 we will discuss tentative evidence, both from lattice gravity and from functional renormalization group equations, that suggests that the generation of such a mass scale is in principle possible. We will also stress that one can generate very specific nonlocal structures, rather than the most general nonlocal theory. If one assumes the generation of nonlocal terms relevant in

---

[1]Although a possibly simpler alternative is provided by theories with extra dimensions, that, when projected onto a four-dimensional brane, can indeed induce nonlocal terms relevant in the IR, as we will illustrate below with the example of the DGP model.

the IR as a sensible working hypothesis, the next question is what terms can give a viable cosmology, consistent with the wealth of current data. Of course, ideally one would like to derive the form of the nonlocal terms from first principle, and then study their consequences. Given the difficulty of such a top-down approach, it make sense to start by a study of the cosmological consequences of possible nonlocal terms. This turns out to be already a rather interesting study; summarizing previous work our group, we will indeed see that the condition of obtaining a viable cosmology are very stringent, and basically select a single model (the so-called RT model), among a large class of alternatives that have been studied. In this model, the nonlocal term corresponds to a diff-invariant mass for the conformal mode. We will then perform in section 3 an updated and detailed study of the phenomenological consequences of the RT model, including comparison with the most recent cosmological datasets, as well as predictions for gravitational-wave (GW) detectors that could turn out to be the smoking gun of this model.

2 Nonlocal mass terms for gauge and gravitational fields

In this section we review how one can construct nonlocal mass terms for gauge fields and for different components of the gravitational field, while preserving gauge or diffeomorphism invariance, respectively. Even if the formal manipulations might seem to hold already at the level of the classical action, we will recall in section 2.3 that these nonlocal terms only make sense at the level of quantum effective action, i.e. must be considered as generated by quantum effects (we will recall in section 2.3.3 some basic properties of the quantum effective action). Indeed, writing such nonlocal terms directly at the level of the fundamental action, one would run into fatal problems with ghost-like degrees of freedom and with causality. In contrast, nonlocal terms routinely appears at the level of quantum effective actions, and, in this context, create no problems with causality or with ghosts.

2.1 Nonlocal and gauge-invariant mass term for gauge fields

The simplest example is given by massive electrodynamics [7]. Consider the Proca action for a massive photon coupled to an external conserved current j^μ

$$S = \int d^4x \left(-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2} m_\gamma^2 A_\mu A^\mu - j_\mu A^\mu \right). \tag{2.1}$$

The equation of motion derived from this action is $\partial_\mu F^{\mu\nu} - m_\gamma^2 A^\nu = j^\nu$. Acting with ∂_ν on both sides and using $\partial_\nu j^\nu = 0$ one finds $m_\gamma^2 \partial_\nu A^\nu = 0$, so, for $m_\gamma \neq 0$, we get $\partial_\nu A^\nu = 0$. Note that this condition, that in massless electrodynamics can be obtained as a choice of gauge, is here obtained dynamically from the equations of motion; indeed, the action (2.1) is not gauge invariant, and there is no gauge freedom to fix. Then, using this condition, the equation of motion becomes $(\Box - m_\gamma^2) A^\mu = j^\mu$. In summary, the equations of motion derived from eq. (2.1) can be written as

$$(\Box - m_\gamma^2) A^\mu = j^\mu, \qquad \partial_\nu A^\nu = 0, \tag{2.2}$$

and describe the three degrees of freedom of a massive spin-1 particle. Now, let us compare this with the nonlocal action

$$S = \int d^4x \left[-\frac{1}{4} F_{\mu\nu} \left(1 - \frac{m_\gamma^2}{\Box} \right) F^{\mu\nu} - j_\mu A^\mu \right]. \tag{2.3}$$

The corresponding equation of motion is

$$\left(1 - \frac{m_\gamma^2}{\Box}\right)\partial_\mu F^{\mu\nu} = j^\nu\,. \tag{2.4}$$

The action (2.3) is nonlocal but gauge invariant. We can therefore impose $\partial_\mu A^\mu = 0$ as a choice of gauge, and then eq. (2.4) reduces to $(\Box - m_\gamma^2)A^\mu = j^\mu$. We therefore get back the two equations in (2.2), showing that eqs. (2.1) and (2.3) provide two equivalent formulations of the same classical theory. The equivalence of the two theories can also be shown by using the "Stückelberg trick", as originally done in [7]. One introduces the Stückelberg field φ and replaces

$$A_\mu \to A_\mu + (1/m_\gamma)\partial_\mu\varphi\,, \tag{2.5}$$

in the Proca action (2.1), which becomes

$$S[A_\mu, \varphi] = \int d^4x \left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}m_\gamma^2 A_\mu A^\mu - \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - m_\gamma A^\mu\partial_\mu\varphi - j_\mu A^\mu\right]\,. \tag{2.6}$$

We have added a new degrees of freedom φ, but we have gained a gauge symmetry, defined by the transformation

$$A_\mu \to A_\mu - \partial_\mu\theta\,, \qquad \varphi \to \varphi + m_\gamma\theta\,, \tag{2.7}$$

since the combined transformation leaves invariant the right-hand side of eq. (2.5). The equations of motion obtained from the action (2.6) are

$$\partial_\mu F^{\mu\nu} = m_\gamma^2 A^\nu + m_\gamma\partial^\nu\varphi + j^\nu\,, \tag{2.8}$$
$$\Box\varphi = -m_\gamma\partial_\mu A^\mu\,. \tag{2.9}$$

Equation (2.9) can be formally solved by $\varphi(x) = -m_\gamma\Box^{-1}(\partial_\mu A^\mu)$. Inserting this into eq. (2.8) we get eq. (2.4) [or, equivalently, inserting it into $S[A_\mu, \varphi]$, eq. (2.6), we get eq. (2.3)]. Thus, eq. (2.3) provides an alternative description of a massive photon which is explicitly gauge invariant, but nonlocal. An equivalent way of expressing the same result is to observe that the Ward identities of QED do not forbid a photon mass term [8]. Indeed, they only imply that the photon self-energy $\Sigma_{\mu\nu}$ is transverse, so that, in momentum space, it can be written as

$$\Sigma_{\mu\nu}(p) = \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right)F(p^2)\,. \tag{2.10}$$

If, in the limit $p^2 \to 0$, $F(p^2) \neq 0$, then the photon acquires a nonzero mass, which is precisely the one described by the nonlocal term in eq. (2.3).

It is interesting to rewrite the nonlocal mass term in a way that will be useful to make contact with the gravitational case [9]. We separate the gauge field into its transverse and longitudinal parts,

$$A_\mu = A_\mu^{\mathrm{T}} + \partial_\mu\alpha\,, \tag{2.11}$$

where $\partial^\mu A_\mu^{\mathrm{T}} = 0$. Under a gauge transformations $A_\mu \to A_\mu - \partial_\mu\theta$, we have $\alpha \to \alpha - \theta$ and $A_\mu^{\mathrm{T}} \to A_\mu^{\mathrm{T}}$, so A_μ^{T} is gauge invariant. To invert eq. (2.11) we take the divergence, which gives $\partial^\mu A_\mu = \Box\alpha$. This can be formally inverted as $\alpha = \Box^{-1}\partial^\mu A_\mu$. Substituting this into $A_\mu^{\mathrm{T}} = A_\mu - \partial_\mu\alpha$ we get

$$A_\mu^{\mathrm{T}} = A_\mu - \frac{1}{\Box}\partial_\mu\partial^\nu A_\nu \equiv P_\mu^\nu A_\nu\,, \tag{2.12}$$

where

$$P^\nu_\mu \equiv \delta^\nu_\mu - \frac{\partial_\mu \partial^\nu}{\Box} \tag{2.13}$$

is a nonlocal operator. The transverse part A^{T}_μ is therefore a gauge-invariant and nonlocal functional of the gauge field A_μ. In terms of A^{T}_μ, it is straightforward to check that the action (2.3) can be rewritten as

$$S = \int d^4x \left(-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2} m_\gamma^2 A^{\mathrm{T}}_\mu A^{T\mu} \right) - j^\mu A_\mu \,. \tag{2.14}$$

We can further replace A_μ with A^{T}_μ in the kinetic term, since $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = \partial_\mu A^T_\nu - \partial_\nu A^T_\mu$, and in $j^\mu A_\mu$, since $j^\mu \partial_\mu \alpha$ vanishes upon integrating by parts and using current conservation. This is a consequence of the fact that α is a pure gauge degree of freedom and can be set to zero with a gauge transformation.

In the case of massive electrodynamics the nonlocality is only apparent, since we have seen that the nonlocal term in the equation of motion (2.4) can be made local with the gauge choice $\partial_\mu A^\mu = 0$. In this sense, the above manipulations can be performed even at the level of fundamental action, since anyhow the nonlocality can be gauged away. Less trivial examples can be constructed where the nonlocality is genuine (and therefore, as we will review in section 2.3, only makes sense at the level of quantum effective action). In particular, the non-abelian generalization of the nonlocal mass term in eq. (2.3) is

$$\frac{m_g^2}{2} \mathrm{Tr} \int d^4x \, F_{\mu\nu} \frac{1}{\Box} F^{\mu\nu} \,, \tag{2.15}$$

where $F_{\mu\nu} = F^a_{\mu\nu} T^a$, $\Box^{ab} = D^{ac}_\mu D^{\mu,cb}$ and $D^{ab}_\mu = \delta^{ab} \partial_\mu - g f^{abc} A^c_\mu$ is the covariant derivative. This nonlocal term corresponds to giving a mass m_g to the non-abelian gauge bosons, plus extra nonlocal interaction terms that, altogether, reconstruct a gauge-invariant quantity. This nonlocal mass term cannot be reduced to a local term with a gauge choice, and has been postulated to appear in the quantum effective action of QCD, in order to reproduce non-perturbative results on the running of the strong coupling constant and on the gluon propagator in the IR, obtained from operator product expansions and from lattice QCD [10–12]. It is therefore an example of a nonlocal term that can appear in a quantum effective action because of strong IR effects.

2.2 Nonlocal and diff-invariant mass term for the conformal mode

2.2.1 Linearized GR in nonlocal variables

We now discuss possible generalizations of the above construction to the gravitational field. A possible route is to begin with gravity linearized over Minkowski space. First of all, it is useful to see how linearized gravity can be rewritten in terms of nonlocal variables, analogous to A^T_μ of the previous section (we follow the discussion in [9, 13]). We begin by writing $g_{\mu\nu} = \eta_{\mu\nu} + \kappa h_{\mu\nu}$, where $\kappa = (32\pi G)^{1/2}$. To quadratic level, the Einstein-Hilbert action becomes

$$S^{(2)}_{\mathrm{EH}} = \frac{1}{2} \int d^4x \, h_{\mu\nu} \mathcal{E}^{\mu\nu,\rho\sigma} h_{\rho\sigma} \,, \tag{2.16}$$

where $\mathcal{E}^{\mu\nu,\rho\sigma}$ is the Lichnerowicz operator,[2] while the interaction with matter with energy-momentum tensor $T^{\mu\nu}$, to linear order in $h_{\mu\nu}$, is given by

$$S_{\text{int}}^{(1)} = \frac{\kappa}{2} \int d^4x \, h_{\mu\nu} T^{\mu\nu} \,. \tag{2.17}$$

The linearized equations of motion derived from $S_{\text{EH}}^{(2)} + S_{\text{int}}^{(1)}$ are therefore

$$\mathcal{E}^{\mu\nu,\rho\sigma} h_{\rho\sigma} = -\frac{\kappa}{2} T^{\mu\nu} \,. \tag{2.18}$$

We next decompose the metric as

$$h_{\mu\nu} = h_{\mu\nu}^{\text{TT}} + \frac{1}{2}(\partial_\mu \epsilon_\nu + \partial_\nu \epsilon_\mu) + \frac{1}{3}\eta_{\mu\nu} s \,, \tag{2.19}$$

where $h_{\mu\nu}^{\text{TT}}$ is transverse ($\partial^\mu h_{\mu\nu}^{\text{TT}} = 0$) and traceless ($\eta^{\mu\nu} h_{\mu\nu}^{\text{TT}} = 0$), and therefore has five independent components.[3] We have therefore decomposed the 10 independent components of the symmetric tensor $h_{\mu\nu}$ into the five components of $h_{\mu\nu}^{\text{TT}}$, the four components of ϵ_μ, and the scalar s. Under a linearized diffeomorphism $h_{\mu\nu} \to h_{\mu\nu} - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu)$ we have $\epsilon_\mu \to \epsilon_\mu - \xi_\mu$ while $h_{\mu\nu}^{\text{TT}}$ and s are gauge invariant. Thus ϵ_μ describes the four pure gauge degrees of freedom, while s plus the five components of the TT tensor $h_{\mu\nu}^{\text{TT}}$ describe the six gauge-invariant degrees of freedom of the gravitational field. Notice that, at this linearized level, s is equivalent to the conformal mode of the metric. Indeed, restricting to the scalar sector (i.e. setting $\epsilon_\mu = 0$ and $h_{\mu\nu}^{\text{TT}} = 0$) and writing $g_{\mu\nu} = e^{2\sigma}\eta_{\mu\nu}$, comparison with eq. (2.19) shows that, at the linear level, $2\sigma = s/3$.

Similarly to the electromagnetic case of section 2.1, the quantities that appear in the right-hand side of eq. (2.19) are nonlocal functionals of the original metric perturbation $h_{\mu\nu}$. The inversion of eq. (2.19) is straightforward [9]. It is convenient to further separate ϵ_μ into its transverse and longitudinal parts, $\epsilon_\mu = \epsilon_\mu^{\text{T}} + \partial_\mu \alpha$, where $\partial^\mu \epsilon_\mu^{\text{T}} = 0$. Then, taking the trace of eq. (2.19) we get $h = (4/3)s + \Box\alpha$, while contracting eq. (2.19) with $\partial^\mu\partial^\nu$, gives $\partial^\mu\partial^\nu h_{\mu\nu} = \Box[s/3 + \Box\alpha]$. Combining these equations we get

$$s = \left(\eta^{\mu\nu} - \frac{1}{\Box}\partial^\mu\partial^\nu\right) h_{\mu\nu} \,, \qquad \alpha = -\frac{1}{3}\frac{1}{\Box}\left(\eta^{\mu\nu} - \frac{4}{\Box}\partial^\mu\partial^\nu\right) h_{\mu\nu} \,. \tag{2.20}$$

We can now extract ϵ_μ^{T} by applying ∂^μ to eq. (2.19) and using the above expressions for α and s. This gives $\epsilon_\mu^{\text{T}} = 2\Box^{-1} P_\mu^\rho \partial^\sigma h_{\rho\sigma}$. Finally, substituting these expressions into eq. (2.19) we get

$$h_{\mu\nu}^{\text{TT}} = h_{\mu\nu} - \frac{1}{3}\left(\eta_{\mu\nu} - \frac{\partial_\mu\partial_\nu}{\Box}\right) h - \frac{1}{\Box}(\partial_\mu\partial^\rho h_{\nu\rho} + \partial_\nu\partial^\rho h_{\mu\rho}) + \frac{1}{3}\eta_{\mu\nu}\frac{1}{\Box}\partial^\rho\partial^\sigma h_{\rho\sigma}$$

$$+ \frac{2}{3}\frac{1}{\Box^2}\partial_\mu\partial_\nu\partial^\rho\partial^\sigma h_{\rho\sigma} \,. \tag{2.21}$$

---

[2]The Lichnerowicz operator is defined by $\mathcal{E}^{\mu\nu,\rho\sigma} \equiv \frac{1}{2}(\eta^{\mu\rho}\eta^{\nu\sigma} + \eta^{\mu\sigma}\eta^{\nu\rho} - 2\eta^{\mu\nu}\eta^{\rho\sigma})\Box + (\eta^{\rho\sigma}\partial^\mu\partial^\nu + \eta^{\mu\nu}\partial^\rho\partial^\sigma) - \frac{1}{2}(\eta^{\mu\rho}\partial^\sigma\partial^\nu + \eta^{\nu\rho}\partial^\sigma\partial^\mu + \eta^{\mu\sigma}\partial^\rho\partial^\nu + \eta^{\nu\sigma}\partial^\rho\partial^\mu)$, where $\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu$ is the flat-space d'Alembertian. We use the signature $\eta_{\mu\nu} = (-,+,+,+)$ and MTW [14] sign conventions.

[3]We assume here $3+1$ spacetime dimensions. See [13] for the corresponding equations in $d+1$ spacetime dimensions, with d generic.

These results can be written more compactly using the projector $P^{\mu\nu} = \eta^{\mu\nu} - (\partial^\mu \partial^\nu / \Box)$. In particular,

$$s = P^{\mu\nu} h_{\mu\nu} \,, \tag{2.22}$$

$$h^{\mathrm{TT}}_{\mu\nu} = \left(P^\rho_\mu P^\sigma_\nu - \frac{1}{3} P_{\mu\nu} P^{\rho\sigma} \right) h_{\rho\sigma} \,. \tag{2.23}$$

The fact that this expression for $h^{\mathrm{TT}}_{\mu\nu}$ is indeed transverse and traceless is easily checked by using the properties of $P_{\mu\nu}$, $\partial^\mu P_{\mu\nu} = 0$, $\eta^{\mu\nu} P^\rho_\mu P^\sigma_\nu = P^{\rho\sigma}$ and $\eta^{\mu\nu} P_{\mu\nu} = 3$.

Plugging the decomposition (2.19) into the action (2.16) one finds that ϵ_μ cancels (an obvious consequence of the fact that it is a pure gauge mode), and [13]

$$S^{(2)}_{\mathrm{EH}} = \frac{1}{2} \int d^4x \left[h^{\mathrm{TT}}_{\mu\nu} \Box (h^{\mu\nu})^{\mathrm{TT}} - \frac{2}{3} s \Box s \right] \,. \tag{2.24}$$

Performing the same decomposition as in (2.19) for the energy-momentum tensor, the linearization of the interaction term becomes

$$S^{(1)}_{\mathrm{int}} = \frac{\kappa}{2} \int d^4x \left[h^{\mathrm{TT}}_{\mu\nu} (T^{\mu\nu})^{\mathrm{TT}} + \frac{1}{3} s T \right] \,, \tag{2.25}$$

where $T = \eta^{\mu\nu} T_{\mu\nu}$. The equations of motion (2.18) derived from $S^{(2)}_{\mathrm{EH}} + S^{(1)}_{\mathrm{int}}$ can then be rewritten as

$$\Box h^{\mathrm{TT}}_{\mu\nu} = -\frac{\kappa}{2} T^{\mathrm{TT}}_{\mu\nu} \,, \qquad \Box s = \frac{\kappa}{4} T \,. \tag{2.26}$$

At first, eq. (2.26) can be surprising, because it seems to imply that $h^{\mathrm{TT}}_{\mu\nu}$ and s describe six radiative gauge-invariant degrees of freedom. Of course, we know that in GR only the two degrees of freedom associated to the helicities ± 2 are radiative, while the remaining four gauge-invariant degrees of freedom are non-radiative and satisfy Poisson equations. Furthermore, the sign of the kinetic term of s in eq. (2.24) is such that the scalar s seems to be a ghost! As discussed in [13], the resolution of this apparent paradox is related to the nonlocal relation between the original metric perturbation $h_{\mu\nu}$ and the variables $\{h^{\mathrm{TT}}_{\mu\nu}, s\}$. The fact that this relation is nonlocal in time, and not only in space, implies that the initial data assigned on $h_{\mu\nu}$ on a given time slice are not sufficient to provide initial data on $\{h^{\mathrm{TT}}_{\mu\nu}, s\}$, so a naive counting of degrees of freedom in terms of $\{h^{\mathrm{TT}}_{\mu\nu}, s\}$ goes wrong.[4] Notice that this is different from what happens in the standard $3+1$ decomposition of the metric perturbations over flat space,

$$h_{00} = 2\psi \,, \qquad h_{0i} = \beta_i + \partial_i \gamma$$

$$h_{ij} = -2\phi \delta_{ij} + \left(\partial_i \partial_j - \frac{1}{3} \delta_{ij} \boldsymbol{\nabla}^2 \right) \lambda + \frac{1}{2} (\partial_i v_j + \partial_j v_i) + H^{\mathrm{TT}}_{ij} \,, \tag{2.27}$$

---

[4]A simple example to understand what exactly goes wrong, again discussed in [13], is provided by a scalar field ϕ that satisfies a Poisson equation $\boldsymbol{\nabla}^2 \phi = \rho$. If one introduces a field $\tilde{\phi}$ related to ϕ by a nonlocal relation such as $\tilde{\phi} = \Box^{-1} \phi$, the original Poisson equation can be rewritten as $\Box \tilde{\phi} = \tilde{\rho}$, where $\tilde{\rho} \equiv \boldsymbol{\nabla}^{-2} \rho$, so now $\tilde{\phi}$ looks like a propagating degree of freedom. However, for $\rho = 0$ the original equation $\boldsymbol{\nabla}^2 \phi = \rho$ (with vanishing boundary conditions at infinity) only has the solution $\phi = 0$. If we want to rewrite this equation in terms of $\tilde{\phi}$ without introducing spurious degrees of freedom we must therefore supplement the equation $\Box \tilde{\phi} = \tilde{\rho}$ with the condition that, when $\tilde{\rho} = 0$, the only acceptable solution is $\tilde{\phi} = 0$, which precisely kills the radiative solution.

where v^i and β^i are transverse spatial vectors, $\partial_i \beta^i = 0$ and $\partial_i v^i = 0$, and $H_{ij}^{\rm TT}$ is transverse and traceless with respect to the spatial indices, $\partial^j H_{ij}^{\rm TT} = 0$ and $\delta^{ij} H_{ij}^{\rm TT} = 0$. Indeed, this decomposition only involves spatial derivatives and therefore its inversion is nonlocal in space but local in time. From these variables, one can form six variables that are invariant under linearized gauge transformations: the two Bardeen variables, $\Phi = -\phi - (1/6)\boldsymbol{\nabla}^2 \lambda$ and $\Psi = \psi - \dot{\gamma} + (1/2)\ddot{\lambda}$, that are scalars under spatial rotations; the spatial vector $\Xi_i = \beta_i - (1/2)\dot{v}_i$, which, being transverse, has only two independent components; and the spatial tensor $H_{ij}^{\rm TT}$, which is already gauge-invariant (again, at the linearized level) and, being transverse and traceless (and carrying only spatial indices, contrary to $h_{\mu\nu}^{\rm TT}$), also has only two independent components. Standard analysis (see e.g. [15, 16] or chapter 18 of [17]) then shows that, after performing the same decomposition for the energy-momentum tensor, namely

$$T_{00} = \rho\,, \qquad\qquad T_{0i} = S_i + \partial_i S\,,$$
$$T_{ij} = P\delta_{ij} + \left(\partial_i \partial_j - \frac{1}{3}\delta_{ij}\boldsymbol{\nabla}^2\right)\Sigma + \frac{1}{2}(\partial_i \Sigma_j + \partial_j \Sigma_i) + \Sigma_{ij}\,, \tag{2.28}$$

where $\partial_i \Sigma^i = 0$, $\partial_i S^i = 0$, $\partial^i \Sigma_{ij} = 0$ and $\delta^{ij}\Sigma_{ij} = 0$, the linearized equations of motion can be rewritten as

$$\boldsymbol{\nabla}^2 \Phi = -4\pi G\rho\,, \qquad\qquad \boldsymbol{\nabla}^2 \Psi = -4\pi G(\rho - 2\boldsymbol{\nabla}^2 \Sigma)\,, \tag{2.29}$$

$$\boldsymbol{\nabla}^2 \Xi_i = -16\pi G S_i\,, \qquad\qquad \Box H_{ij}^{\rm TT} = -16\pi G \Sigma_{ij}\,. \tag{2.30}$$

We then get the standard result that only the two degrees of freedom of the tensor perturbations obey a wave equation, while the remaining gauge-invariant degrees of freedom described by Φ, Ψ and Ξ_i obey Poisson equations, and therefore are non-radiative.

Comparing the decompositions (2.19) and (2.27) one finds that the field s can be written explicitly as a nonlocal function of the Bardeen variables as [13]

$$s = 6\Phi - 2\Box^{-1}\boldsymbol{\nabla}^2(\Phi + \Psi)\,. \tag{2.31}$$

Just as in the example discussed in footnote 4, the apparent radiative nature of s in eq. (2.26) is an artifact due to this nonlocal relation, that introduces a spurious degree of freedom associated to the homogeneous equation $\Box s = 0$. Indeed, from eq. (2.29), $\boldsymbol{\nabla}^2(\Phi + \Psi)$ is fully determined by the source terms, and vanishes if the latter vanish. Thus, in order to eliminate this spurious degree of freedom we must supplement eq. (2.26) with the condition that $s = 0$ when $T = 0$, i.e. we must discard again the homogeneous solution of eq. (2.26) (and similarly for the helicities $0, \pm 1$ of $h_{\mu\nu}^{\rm TT}$) At the quantum level, this implies that there are no creation and annihilation operators associated to s, and s cannot appear on the external legs of a Feynman diagram. Therefore, the apparent ghost-like nature of s in eq. (2.24) is fictitious and, of course, in General Relativity there is no actual ghost.

2.2.2 Nonlocal mass terms at the linearized level

As we have seen, the use of the variables $\{h_{\mu\nu}^{\rm TT}, s\}$ is not convenient if we want to count the independent degrees of freedom of the theory and determine their radiative/non-radiative nature; for those purposes it is better to work directly with the original metric perturbation $h_{\mu\nu}$, or with the variables of the 3+1 decomposition (2.27), or with the ADM decomposition.

However, the variables $\{h_{\mu\nu}^{\rm TT}, s\}$ have the advantage that one can very easily see how a diff-invariant nonlocal mass term can be naturally written for different modes of the gravitational field, at the linearized level. Quite trivially, we can just modify eq. (2.24) into

$$\Gamma^{(2)} = \frac{1}{2} \int d^4x \left[h_{\mu\nu}^{\rm TT}(\Box - m_1^2)(h^{\mu\nu})^{\rm TT} - \frac{2}{3}\, s(\Box + m_2^2)s \right] \,, \tag{2.32}$$

for some masses m_1 and m_2.[5] This is analogous to eq. (2.14) in the case of a massive gauge field. These mass terms are clearly diff-invariant, since $h_{\mu\nu}^{\rm TT}$ and s are diff-invariant (again, at the linearized level). On the other hand, because of the relations (2.22), (2.23), once rewritten in terms of $h_{\mu\nu}$ they will be nonlocal. We have indeed used the notation Γ, rather than S, to stress that, because of the nonlocality, this modification makes sense at the level of the quantum effective action Γ, rather than for the fundamental action S.

To go beyond the linearized approximation, we can search for covariantizations of these expressions, as we will do in section 2.2.3. The second term gives a mass to s or, equivalently, to the conformal mode. The models that we will study in the following will be covariantizations of the above expression, with $m_1^2 = 0$ and $m_2^2 \equiv m^2 > 0$. We are therefore assuming that there exists a mechanism that, in the quantum effective action, generates a mass for the conformal mode, while leaving $h_{\mu\nu}^{\rm TT}$ massless. At the phenomenological level this is required by the fact that, as we will review in appendix A, among a large class of models explored, only those of this form appear to have a viable cosmological evolution. At the theoretical level, this is also suggested by various arguments, that will be discussed in section 2.4, that identify the conformal mode as the main candidate for producing strong IR quantum effects. Thus, we will look for a covariantization of a quantum effective action that, at quadratic level, has the form

$$\Gamma^{(2)} = \frac{1}{2} \int d^4x \left[h_{\mu\nu}^{\rm TT}\Box(h^{\mu\nu})^{\rm TT} - \frac{2}{3}\, s(\Box + m^2)s \right] \,, \tag{2.33}$$

so that the linearized equations of motion (2.26) are modified into[6]

$$\Box h_{\mu\nu}^{\rm TT} = -\frac{\kappa}{2} T_{\mu\nu}^{\rm TT}\,, \qquad (\Box + m^2)s = \frac{\kappa}{4} T\,. \tag{2.34}$$

To perform the covariantization, it is now convenient to go back to the original metric perturbation $h_{\mu\nu}$. Using eq. (2.22), we immediately see that eq. (2.33) can be rewritten as

$$\Gamma^{(2)} = \frac{1}{2} \int d^4x \left[h_{\mu\nu}\mathcal{E}^{\mu\nu,\rho\sigma}h_{\rho\sigma} - \frac{2}{3}\, m^2(P^{\mu\nu}h_{\mu\nu})^2 \right] \,, \tag{2.35}$$

while eq. (2.34) is equivalent to

$$\mathcal{E}^{\mu\nu,\rho\sigma}h_{\rho\sigma} - \frac{2}{3}\, m^2 P^{\mu\nu}P^{\rho\sigma}h_{\rho\sigma} = -\frac{\kappa}{2} T^{\mu\nu}\,. \tag{2.36}$$

In view of the covariantization, it is also convenient to rescale $h_{\mu\nu} \to h_{\mu\nu}/\kappa$, so that now $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$,

$$\Gamma^{(2)} = \frac{1}{64\pi G} \int d^4x \left[h_{\mu\nu}\mathcal{E}^{\mu\nu,\rho\sigma}h_{\rho\sigma} - \frac{2}{3}\, m^2(P^{\mu\nu}h_{\mu\nu})^2 \right] \,, \tag{2.37}$$

---

[5]We have chosen the signs in front of m_1^2 and m_2^2 so that $m_1^2 > 0$ and $m_2^2 > 0$ corresponds to 'non-tachyonic' masses, independently of the signs in front of the $\Box$ operators.

[6]As we will recall in section 2.3.3, the quantum effective action actually generates the equations of motion for the vacuum expectation values of the corresponding operators.

and

$$\mathcal{E}^{\mu\nu,\rho\sigma}h_{\rho\sigma} - \frac{2}{3}m^2 P^{\mu\nu}P^{\rho\sigma}h_{\rho\sigma} = -16\pi G\,T^{\mu\nu}\,. \tag{2.38}$$

2.2.3 Covariantizations: RT and RR models

We now look for possible covariantizations of the above expressions. Covariantizations, when they exists, are in general not unique. However, some choices can be more natural than others. We will see that, starting from the equation of motion (2.38) or from the quantum effective action (2.37), one ends up quite naturally with two different covariantizations, that define two possible models.

Let us start from the covariantization of eq. (2.38). The linearization of the Einstein tensor $G_{\mu\nu}$ is $G_{\mu\nu}^{(1)} = -(1/2)\mathcal{E}_{\mu\nu,\rho\sigma}h^{\rho\sigma}$, so the term $\mathcal{E}_{\mu\nu,\rho\sigma}h^{\rho\sigma} = -2G_{\mu\nu}^{(1)}$ in eq. (2.38) is uniquely promoted to $-2G_{\mu\nu}$ in the full covariant theory, by the requirement that we recover GR for $m = 0$. The nontrivial part is the covariantization of the mass term. At linear level the Ricci scalar becomes $R^{(1)} = -(\eta^{\rho\sigma}\Box - \partial^\rho\partial^\sigma)h_{\rho\sigma}$, that can be rewritten as $R^{(1)} = -\Box(P^{\rho\sigma}h_{\rho\sigma})$, so

$$P^{\rho\sigma}h_{\rho\sigma} = -\Box^{-1}R^{(1)}\,. \tag{2.39}$$

Therefore eq. (2.36) is equivalent to

$$-2G_{\mu\nu}^{(1)} + \frac{2}{3}m^2 P_{\mu\nu}\Box_\eta^{-1}R^{(1)} = -16\pi G\,T_{\mu\nu}\,, \tag{2.40}$$

where the notation $\Box_\eta$ stresses that, until now, the $\Box$ operator was the one with respect to the flat metric $\eta_{\mu\nu}$. After promoting $G_{\mu\nu}^{(1)}$ to $G_{\mu\nu}$, if we want to preserve energy-momentum conservation $\boldsymbol{\nabla}^\mu T_{\mu\nu} = 0$, we must promote $P_{\mu\nu}\Box_\eta^{-1}R^{(1)}$ to a transverse tensor, whose covariant derivative vanishes. To this purpose it is useful to observe that, in a generic Riemannian manifold, any symmetric tensor $S_{\mu\nu}$ can be decomposed as

$$S_{\mu\nu} = S_{\mu\nu}^{\mathrm{T}} + \frac{1}{2}(\boldsymbol{\nabla}_\mu S_\nu + \boldsymbol{\nabla}_\nu S_\mu)\,, \tag{2.41}$$

where $\boldsymbol{\nabla}^\mu S_{\mu\nu}^{\mathrm{T}} = 0$ [18, 19]. The extraction of the transverse part of a tensor is itself a nonlocal operation. In flat space, where $\boldsymbol{\nabla}_\mu \to \partial_\mu$, proceeding as we have done in the derivation of eqs. (2.20)–(2.21), one finds that

$$S_{\mu\nu}^{\mathrm{T}} = S_{\mu\nu} - \frac{1}{\Box_\eta}(\partial_\mu\partial^\rho S_{\rho\nu} + \partial_\nu\partial^\rho S_{\rho\mu}) + \frac{1}{\Box_\eta^2}\partial_\mu\partial_\nu\partial^\rho\partial^\sigma S_{\rho\sigma}\,. \tag{2.42}$$

Using this expression we can easily check that, in flat space, for a tensor $S_{\mu\nu}$ of the form $S_{\mu\nu}(x) = \eta_{\mu\nu}A(x)$, we have $S_{\mu\nu}^{\mathrm{T}} = P_{\mu\nu}A(x)$.[7] Thus, to linear order in an expansion over flat space, the term $P_{\mu\nu}\Box_\eta^{-1}R^{(1)}$ in eq. (2.40) is the same as the transverse part of the tensor $(\eta_{\mu\nu}\Box_\eta^{-1}R^{(1)})$, that we denote as $(\eta_{\mu\nu}\Box_\eta^{-1}R^{(1)})^T$, and eq. (2.40) is the same as

$$G_{\mu\nu}^{(1)} - \frac{1}{3}m^2(\eta_{\mu\nu}\Box_\eta^{-1}R^{(1)})^{\mathrm{T}} = 8\pi G\,T_{\mu\nu}\,. \tag{2.44}$$

---

[7]This could be derived even more simply by observing that, in flat space, where ∂_μ commutes with $\Box_\eta$ and therefore with $\Box_\eta^{-1}$, we can write

$$\eta_{\mu\nu}A = (\eta_{\mu\nu} - \Box_\eta^{-1}\partial_\mu\partial_\nu)A + (1/2)\partial_\mu(\Box_\eta^{-1}\partial_\nu A) + (1/2)\partial_\nu(\Box_\eta^{-1}\partial_\mu A) = P_{\mu\nu}A + (1/2)(\partial_\mu S_\nu + \partial_\nu S_\mu)\,, \tag{2.43}$$

where $S_\mu = \Box_\eta^{-1}\partial_\mu A$. Since $\partial^\mu(\eta_{\mu\nu} - \Box_\eta^{-1}\partial_\mu\partial_\nu)A = (\partial_\nu - \partial_\nu)A = 0$, $P_{\mu\nu}A$ is transverse, so $P_{\mu\nu}A = S_{\mu\nu}^{\mathrm{T}}$.

In this form, there is a natural covariantization given by

$$G_{\mu\nu} - \frac{1}{3}\, m^2 \left(g_{\mu\nu}\Box^{-1}R\right)^{\mathrm{T}} = 8\pi G\, T_{\mu\nu}\,, \tag{2.45}$$

where now $\Box^{-1}$ is the inverse of the covariant $\Box$ operator with respect to the generic metric $g_{\mu\nu}$, and the operation of taking the transverse part is the fully covariant operation defined by eq. (2.41). Equation (2.45) defines the so-called RT model, where R stands for the occurrence of the Ricci scalar and T for the extraction of the transverse part. This is the model that was first proposed in [20] (through a rather different route that we will review in appendix A). It was the first model of this class of nonlocal theories that looked cosmologically viable and even today, after the study of many alternative possibilities, it turns out to be the only viable one; the reasons that gradually eliminated all other alternatives will be discussed in appendix A. This model will therefore be the main focus of this paper. Notice that it is defined at the level of a nonlocal equation of motion rather than by a (quantum effective) action. Indeed, there is no known nonlocal action from which eq. (2.45) can be derived.

A different covariantization emerges naturally if we rather start from the quantum effective action (2.16). As usual, $d^4x\,(1/4)h_{\mu\nu}\mathcal{E}^{\mu\nu,\rho\sigma}h_{\rho\sigma}$ becomes $d^4x\sqrt{-g}R$ while, using eq. (2.39), $(P^{\mu\nu}h_{\mu\nu})^2$ is the same as $(\Box_\eta^{-1}R^{(1)})^2$, which is naturally covariantized into $(\Box^{-1}R)^2$. Thus, a natural covariantization of eq. (2.37) is

$$\begin{aligned}
\Gamma_{\mathrm{RR}} &= \frac{1}{16\pi G}\int d^4x\sqrt{-g}\left[R - \frac{m^2}{6}(\Box^{-1}R)^2\right]\\
&= \frac{1}{16\pi G}\int d^4x\sqrt{-g}\left[R - \frac{m^2}{6}R\frac{1}{\Box^2}R\right],
\end{aligned} \tag{2.46}$$

where in the last line we have integrated $1/\Box$ by parts.[8] This gives the model that was first proposed in [21]; we will refer to it as the RR model, after the two occurrences of the Ricci scalar in the nonlocal term.

The RT and RR model by construction coincide to linear order in an expansion over flat space. However, they are otherwise different, and have different cosmological predictions. As we will see in appendix A, the RR model shared most of the phenomenologically attractive properties of the RT model, such as viable cosmological background evolution, stable cosmological perturbations, good fit to Cosmic Microwave Background (CMB), Baryon Acoustic Oscillations (BAO), type Ia Supernovae (SNe) and structure formation data. However, we will also see that it does not pass the constraints from Lunar Laser Ranging, contrarily to the RT model, which is completely immune to it and, to date, passes all the observational tests. Thus, in this paper we will mostly focus on the RT model. Still, the RR model can be useful to illustrate some concepts in a somewhat simpler setting, also because of its relatively simple quantum effective action, and we will therefore also occasionally use it for pedagogical purposes.

2.3 Aspects of effective nonlocal theories

A correct treatment of nonlocal terms involves a few subtle points. It is particularly important to make clear that the nonlocality that we have introduced is not fundamental, i.e. it does not appear at the level of the fundamental action of the theory, that in our case could even

---

[8]Note that the inversions of the $\Box$ operator (and therefore also the integration by parts above) until now have been somewhat formal operations. We will justify them in more detail in section 2.3.3.

simply be Einstein-Hilbert gravity. Indeed, fundamental actions with nonlocal terms have problems with causality, and extra (typically ghost-like) degrees of freedom. However, even when the fundamental theory is local, if it contains massless particles (such as the graviton in GR) the corresponding quantum effective action will be unavoidably nonlocal, and this sort of nonlocality is not associated to any pathology. The same happens when nonlocal terms appear from a fundamental higher-dimensional theory, as in the DGP example, see eq. (2.72) below. These issues have been reviewed at length in [22, 23] (see also [24–27]). Here, for completeness, we summarize them briefly.

2.3.1 Localization and degrees of freedom

A nonlocal quantum effective action can be rewritten in local form by introducing auxiliary fields (see also [26–32]). This is quite convenient for working out the predictions of the theory (e.g. for studying the equations of motions of the theory, the cosmological perturbations, etc.), but requires some care at the level of interpretation, in order not to confuse the auxiliary fields with actual degrees of freedom of the theory. As a simple example, consider the theory of a massive photon discussed in section 2.1. We have seen that it can be formulated as a local but non gauge-invariant theory, as in eq. (2.1), or as a gauge-invariant theory at the price of nonlocality, as in eq. (2.3). One might also get a theory that is at the same time local and gauge-invariant, by introducing an auxiliary anti-symmetric tensor field $U^{\mu\nu}$ defined by $U^{\mu\nu} = \Box^{-1} F^{\mu\nu}$. In this way, one gets a local and gauge-invariant action written in terms of the two fields A^μ and $U^{\mu\nu}$. The equations of motion of the theory can then be rewritten as[9]

$$\partial_\mu F^{\mu\nu} = j^\nu + m^2 \partial_\mu U^{\mu\nu}, \qquad \Box U^{\mu\nu} = F^{\mu\nu}. \qquad (2.47)$$

While the steps leading to eq. (2.47) are formally correct, this local and gauge-invariant formulation seems to suggest that the theory has many more degrees of freedom than the Proca theory of a massive photon that was our starting point: we apparently have a massless gauge-invariant vector field A^μ, which carries two degrees of freedom, interacting with an antisymmetric tensor field $U^{\mu\nu}$, which apparently carries six degrees of freedom. This seems very different from the three degrees of freedom of a massive vector field from which we started. Of course, new degrees of freedom cannot pop out from nowhere, and the delicate point here is the passage from an equation such as $U^{\mu\nu} = \Box^{-1} F^{\mu\nu}$ to the equation $\Box U^{\mu\nu} = F^{\mu\nu}$, i.e. the inversion of the $\Box$ operator. By itself, the most general solution of an equation such as $\Box U^{\mu\nu} = F^{\mu\nu}$ is given by a solution of the inhomogeneous equation plus the most general solution of the associated homogeneous equation $\Box U^{\mu\nu} = 0$. The latter carries with itself the six degrees of freedom associated to $U^{\mu\nu}$. Clearly, if we want this local and gauge-invariant formulation to be equivalent to the original Proca theory, we cannot accept the most general solution of $\Box U^{\mu\nu} = F^{\mu\nu}$. In other words, the initial condition of the auxiliary field $U^{\mu\nu}$ cannot be taken as independent, but must be fixed in terms of the initial condition of the two transverse and the longitudinal components of A^μ, so that the theory indeed still has three independent degrees of freedom. In this sense, $U^{\mu\nu}$ is just an auxiliary field, and does not carry independent degrees of freedom. In particular, at the quantum level there are no creation/annihilation operators associated to it.

---

[9]This can be easily seen by implementing the definition $U^{\mu\nu} = \Box^{-1} F^{\mu\nu}$ by adding to the action a term $\lambda_{\mu\nu}(\Box U^{\mu\nu} - F^{\mu\nu})$, where $\lambda_{\mu\nu}$ is a Lagrange multiplier, and taking the variations with respect to $A_\mu, U^{\mu\nu}$ and $\lambda_{\mu\nu}$. A combination of the two latter equations gives $\lambda^{\mu\nu} = -(m^2/4) U^{\mu\nu}$, and the remaining two equations give eq. (2.47). Of course, eq. (2.47) can also be verified more simply by comparison with eq. (2.4).

A similar example, worked out in detail in [22], is given by the Polyakov quantum effective action in two dimensions. For two-dimensional gravity coupled to conformal matter it is possible to compute exactly the quantum effective action by integrating the conformal anomaly. This leads to the famous Polyakov quantum effective action, which can be written, in terms of the conformal mode, in a local form which is not explicitly invariant under diffeomorphism; equivalently, one can write it in a form which is nonlocal but diff-invariant. In the latter form the Polyakov quantum effective action is proportional to $R\Box^{-1}R$. One could further rewrite the theory in a form which is both local and diff-invariant by introducing an auxiliary field $U = -\Box^{-1}R$. However, in this case where the computation of the quantum effective action can be performed explicitly, it is easy to check that U is not an independent degree of freedom that popped out from nowhere; rather, its initial conditions are fixed in terms of the initial conditions of the conformal factor σ, the precise relation being simply $U_{\rm in} = 2\sigma_{\rm in}$, $\dot{U}_{\rm in} = 2\dot{\sigma}_{\rm in}$ [22].

In the following we will use a similar localization procedure for the RR and RT models. As in the examples above, the auxiliary fields that will be introduced are not new independent degrees of freedom; rather, their initial conditions should be understood as fixed in terms of the initial conditions on the metric, and there are no creation/annihilation operators associated to them (and, therefore, no issues of ghosts at the quantum level). If one had an explicit derivation of the nonlocal term from a fundamental theory, one would in principle be able to determine explicitly their initial conditions in terms of those on the metric. In practice, lacking such a derivation, these initial conditions must be taken as free phenomenological parameters. One might fear that this significantly reduces the predictive power of the theory. However, we will see in section 3 that, in the cosmological context in which we are interested, this introduces only very limited freedom, both at the level of background evolution and of cosmological perturbations, since these initial conditions turn out to be associated mostly to irrelevant directions in parameter space.

2.3.2 Localization of the RR and RT models

We next show how to write nonlocal gravity in a local form. We write the equations both for the RR model, and for the RT model that will eventually be our main focus, since the comparison between the two models can be instructive, and also the manipulations of the equations of the RR model are somewhat simpler. To write the RR model in a local form we introduce two auxiliary fields U and S, defined by $U = -\Box^{-1}R$ and $S = -\Box^{-1}U$ [21]. This can be implemented at the Lagrangian level by introducing two Lagrange multipliers ξ_1, ξ_2 into eq. (2.46),

$$\Gamma_{\rm RR} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \left[R\left(1 - \frac{m^2}{6}S\right) - \xi_1(\Box U + R) - \xi_2(\Box S + U) \right]. \qquad (2.48)$$

The variation with respect to the Lagrange multipliers ξ_1, ξ_2 gives $\Box U = -R$ and $\Box S = -U$, while the variation with respect to S and U, combined with $\Box U = -R$ and $\Box S = -U$, gives $\xi_1 = (m^2/6)S$ and $\xi_2 = (m^2/6)U$. Then, the variation with respect to $h_{\mu\nu}$, combined with the above relations, gives $G_{\mu\nu} = (m^2/6)K_{\mu\nu} + 8\pi G T_{\mu\nu}$, where

$$K^\mu_\nu \equiv 2SG^\mu_\nu - 2\nabla^\mu\partial_\nu S + 2\delta^\mu_\nu \Box S + \delta^\mu_\nu \partial_\rho S\partial^\rho U - \frac{1}{2}\delta^\mu_\nu U^2 - \left(\partial^\mu S\partial_\nu U + \partial_\nu S\partial^\mu U\right). \qquad (2.49)$$

Thus, the RR model is formally written as a scalar-tensor theory, with two scalar fields U and S, although, as we have discussed in section 2.3.1, U and S are not independent degrees of

freedoms, and their initial conditions are in principle fixed in terms of the initial conditions of the metric. In particular, there are no independent solutions associated to the homogeneous equations $\Box U = 0$ and $\Box S = 0$, and no corresponding quanta at the quantum level.

It is interesting to observe that exactly the same local formulation is obtained if we introduce a single auxiliary field $S = \Box^{-2}R$ rather than the pair $S = -\Box^{-1}U$ and $U = -\Box^{-1}R$. For the constraint equations this is evident, as the equation $\Box^2 S = R$ is equivalent to the system of two equations $\Box S = -U$ and $\Box U = -R$ (also from the point of view of the initial conditions on the auxiliary fields). As for the Einstein equations, they also remain the same. Indeed, let us consider the quantum effective action

$$\Gamma_{\rm RR} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \left[R \left(1 - \frac{m^2}{6} S \right) - \xi \left(\Box^2 S - R \right) \right] . \tag{2.50}$$

From the variation with respect to the Lagrange multiplier ξ and the auxiliary field S, we obtain the constraints $\Box^2 S = R$ and $\xi = -(m^2/6)S$. The variation with respect to the metric is more easily evaluated on the action obtained integrating by parts the $\Box^2$ operator in eq. (2.50), i.e.

$$\Gamma_{\rm RR} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \left[R \left(1 - \frac{m^2}{6} S \right) - (\Box\xi)(\Box S) + \xi R \right] . \tag{2.51}$$

The Einstein equations following from the variation of eq. (2.51) with respect to the metric will then contain both ξ and S. However ξ is traded for S using $\xi = -(m^2/6)S$ and, after this step, we find $G_{\mu\nu} = (m^2/6)K_{\mu\nu} + 8\pi G T_{\mu\nu}$, with

$$K_\nu^\mu \equiv 2SG_\nu^\mu - 2\nabla^\mu\partial_\nu S + 2\delta_\nu^\mu\Box S - \delta_\nu^\mu\partial_\rho S\partial^\rho\Box S - \frac{1}{2}\delta_\nu^\mu(\Box S)^2 + \partial^\mu S\partial_\nu\Box S + \partial_\nu S\partial^\mu\Box S . \tag{2.52}$$

Introducing the definition $U = -\Box S$, one can immediately see that the expression for K_ν^μ given in eq. (2.52) coincides with eq. (2.49), showing the equivalence of the two methods.

For the RT model the localization proceeds by defining again $U = -\Box^{-1}R$. We also introduce $S_{\mu\nu} = -Ug_{\mu\nu} = g_{\mu\nu}\Box^{-1}R$ and we extract its transverse part $S_{\mu\nu}^T$ by using eq. (2.41). Thus, eq. (2.45) is localized in terms of an auxiliary scalar field U and the auxiliary four-vector field S_μ that enters through eq. (2.41). The equations of motion then read [20, 33]

$$G_{\mu\nu} + \frac{m^2}{6} \left(2Ug_{\mu\nu} + \boldsymbol{\nabla}_\mu S_\nu + \boldsymbol{\nabla}_\nu S_\mu \right) = 8\pi G\, T_{\mu\nu} , \tag{2.53}$$

$$\Box U = -R , \tag{2.54}$$

$$(\delta_\nu^\mu\Box + \boldsymbol{\nabla}^\mu\boldsymbol{\nabla}_\nu)S_\mu = -2\partial_\nu U , \tag{2.55}$$

where eq. (2.55) is obtained by taking the divergence of eq. (2.41) with $S_{\mu\nu} = -Ug_{\mu\nu}$. The equations of motion of the RT model have a suggestive property in connection with the cosmological constant problem. Let us perform a shift $U(x) \to U(x) + u_0$, with u_0 a constant. Equations (2.54) and (2.55) are invariant while eq. (2.53) becomes

$$G_{\mu\nu} + \frac{m^2}{6} \left(2Ug_{\mu\nu} + \boldsymbol{\nabla}_\mu S_\nu + \boldsymbol{\nabla}_\nu S_\mu \right) = 8\pi G \left(T_{\mu\nu} - \lambda g_{\mu\nu} \right) , \tag{2.56}$$

where $\lambda = m^2 u_0/(24\pi G)$. Thus, u_0 (or, equivalently, the initial condition on U) generates a cosmological constant, and one could chose u_0 to cancel any vacuum energy term in $T_{\mu\nu}$.

It is also instructive to consider the equations of motion of the RR and RT models linearized over flat space, eq. (2.38), that were our starting point, and write them in terms of the auxiliary fields and of the metric variables of the $3 + 1$ decomposition (2.27). Since, by construction, the RR and RT model coincide when linearized over flat space, we use the RR model, whose localization is slightly simpler, since it involves two scalar fields U and S, rather than U and S_μ for the RT model. One then finds that eqs. (2.29)–(2.30) are modified into [21]

$$\boldsymbol{\nabla}^2\left[\Phi - (m^2/6)S\right] = -4\pi G\rho\,, \qquad \Phi - \Psi - (m^2/3)S = -8\pi G\Sigma\,, \qquad (2.57)$$

$$\boldsymbol{\nabla}^2\Xi_i = -16\pi GS_i\,, \qquad \Box H_{ij}^{\rm TT} = -16\pi G\Sigma_{ij}\,, \qquad (2.58)$$

$$(\Box + m^2)U = -8\pi G(\rho - 3P)\,, \qquad \Box S = -U\,, \qquad (2.59)$$

Equation (2.59) is needed to close the system, since S appears in eq. (2.57). Equations (2.57) and (2.58) shows that the original metric perturbations Φ, Ψ and Ξ_i remain non-radiative variables that satisfy Poisson equations, just as in GR.[10] The auxiliary fields U and S satisfy Klein-Gordon equations, but, as we have seen, their initial conditions are fixed in terms of the initial conditions on the metric, and therefore are not free radiative degrees of freedom either. From these equations it is also clear that the conformal mode s remains a non-propagating degree of freedom also in the RT or RR models. Indeed, combining the two equations in (2.57) we get

$$\begin{aligned}
\boldsymbol{\nabla}^2(\Phi + \Psi) &= 2\boldsymbol{\nabla}^2\left(\Phi - \frac{m^2}{6}S\right) + 8\pi G\boldsymbol{\nabla}^2\Sigma \\
&= -8\pi G\boldsymbol{\nabla}^2(\rho - \Sigma)\,.
\end{aligned} \qquad (2.60)$$

Then, from eq. (2.31) we get (again at the linearized level over flat space)

$$\begin{aligned}
s &= 6\Phi - 2\Box^{-1}\boldsymbol{\nabla}^2(\Phi + \Psi) \\
&= 6\Phi + 16\pi G\Box^{-1}\boldsymbol{\nabla}^2(\rho - \Sigma)\,.
\end{aligned} \qquad (2.61)$$

We see that the nonlocal term in s is fully determined by the energy-momentum tensor, in particular by the density ρ and by the anisotropic stress Σ that enters in T_{ij} through eq. (2.28). Thus, s remains a non-radiative degree of freedom, exactly as in GR, and vanishes if $\rho = 0$ and $\Sigma = 0$.

2.3.3 Causality and the quantum effective action

We next discuss why nonlocal terms would induce problems with causality if added at the level of a fundamental action, while they do not in a quantum effective action.

To illustrate the problem with causality of a nonlocal fundamental action, consider for instance an action with a nonlocal term proportional to $(1/2)\int d^4x\,\varphi\Box^{-1}\varphi$ where φ is a scalar field [13]. To complete the definition of this term we must specify the Green's function $G(x,x')$ used to define $\Box^{-1}$, and then

$$\frac{1}{2}\int d^4x\,\varphi(x)(\Box^{-1}\varphi)(x) \equiv \frac{1}{2}\int d^4x d^4x'\,\varphi(x)G(x,x')\varphi(x')\,. \qquad (2.62)$$

---

[10]This should be contrasted with what happens when one linearizes massive gravity with a Fierz-Pauli mass term, in which case Φ becomes a radiative field that satisfies $(\Box - m^2)\Phi =$ source term [16, 34, 35]. The fact that, for $m = 0$, $(\Box - m^2)\Phi$ does not reduces to its GR counterpart $\boldsymbol{\nabla}^2\Phi$ is a reflection of the van Dam-Veltman-Zakharov (vDVZ) discontinuity of linearized massive gravity.

Consider now the contribution of this term to the equation of motion. Taking the variation with respect to φ, we get

$$\frac{1}{2}\frac{\delta}{\delta\varphi(x)}\int dx'dx''\varphi(x')G(x';x'')\varphi(x'') = \frac{1}{2}\int dx'[G(x;x')+G(x';x)]\varphi(x') \equiv \Box_{\text{sym}}^{-1}\varphi\,, \quad (2.63)$$

where $\Box_{\text{sym}}^{-1}$ is the inverse d'Alembertian with respect to the symmetrized Green's function $[G(x;x')+G(x';x)]/2$. Thus, independently of the choice of $G(x,x')$, in the equations of motion we end up with a symmetric Green's function. Since the retarded Green's function is not symmetric, it cannot be obtained from such a variation. The equations of motion obtained from a nonlocal classical action are therefore in general acausal. This is one of the reasons why a fundamental action must be local.

The situation is however completely different for the quantum effective action. Let us recall, following standard textbook material, that, for a scalar field $\varphi(x)$ with fundamental action $S[\varphi]$, the quantum effective action is obtained by introducing an auxiliary source $J(x)$ and defining the generating functional of the connected Green's function $W[J]$ from

$$e^{iW[J]} \equiv \int D\varphi\; e^{iS[\varphi]+i\int J\varphi}\,, \quad (2.64)$$

where $\int J\varphi$ is a notation for $\int d^4x\, J(x)\varphi(x)$. Then $\delta W[J]/\delta J(x) = \langle 0|\varphi(x)|0\rangle_J$. We will use the notation $\langle 0|\varphi(x)|0\rangle_J \equiv \phi[J]$ for the vacuum expectation value of the field $\varphi(x)$ in the presence of the source $J(x)$. The quantum effective action $\Gamma[\phi]$ is defined as a functional of ϕ (rather than of the original field φ), obtained by performing the Legendre transform, $\Gamma[\phi] \equiv W[J] - \int \phi J$, where $J = J[\phi]$ is obtained by inverting $\phi = \phi[J]$. As a consequence, one immediately finds that

$$\delta\Gamma[\phi]/\delta\phi(x) = -J(x)\,. \quad (2.65)$$

From the path integral representation (2.64) it is also easy to show that

$$e^{i\Gamma[\phi]} = \int D\varphi\; e^{iS[\phi+\varphi]-i\int \frac{\delta\Gamma[\phi]}{\delta\phi}\varphi}\,. \quad (2.66)$$

Thus, the physical meaning of the quantum effective action $\Gamma[\phi]$ is that it is a functional of $\phi(x) = \langle 0|\varphi(x)|0\rangle$, obtained by integrating out the quantum fluctuations around it. From eq. (2.65) we also see that $\Gamma[\phi]$ is the quantity whose variation gives the exact equations of motion for the expectation values of the field, which by construction include (in principle, exactly, if one were able to compute Γ exactly) the contribution of the quantum fluctuations.

It is clear a priori that the quantum effective action obtained from a local and causal fundamental action cannot have problems with causality. To see explicitly how this comes out, one must take into account that, as we have seen, Γ does not give the equations of motion of the original field, but rather of its vacuum expectation value. We must however distinguish between the in-out and the in-in expectation values. The effective action computed using the standard Feynman path integral gives the equations of motion of the in-out vacuum expectation value, which are indeed acausal, because they involve the $\Box^{-1}$ operator constructed with the Feynman propagator. However, there is nothing wrong with this, since in-out matrix elements are not directly observable. Rather, they just appear in intermediate steps of the computation of observables, such as scattering cross section, and indeed the Feynman propagator appears everywhere in quantum field theory computations. In contrast,

the in-in matrix elements of the field are observables; for instance, $\langle 0_{\mathrm{in}}|\varphi(t,\mathbf{x})|0_{\mathrm{in}}\rangle$ is the vacuum expectation value of the quantum field φ at a given time t. To obtain the equations of motion of the in-in matrix elements one must evaluate the path integral in Γ by using the Schwinger-Keldish prescription. As a result, the in-in matrix elements automatically obey causal equations of motions in which the retarded propagator appears [36–38]. In practice, the result of the computation with the Schwinger-Keldish path integral turns out to be equivalent to that obtained by just performing a formal variation of the quantum effective action, without specifying the Green's function used to define $\Box^{-1}$, and then replacing the resulting occurrences of $\Box^{-1}$ in the equations of motion with the $\Box^{-1}$ operator defined with respect to the retarded Green's function (see section 12.1.6 of [38] for a pedagogical discussion in the quantum mechanical case, and [39] for a proof valid for the one-loop quantum effective action in curved space).[11]

2.4   Possible mechanisms for the generation of an IR mass scale

2.4.1   Perturbative loop corrections

We next discuss possible mechanisms for the generation of these nonlocal terms. We begin by observing that perturbative loop corrections due to massive matter fields cannot be responsible for them [40]. In gravity the one-loop corrections induced by matter fields can produce nonlocal form factors in the quantum effective action, associated to terms quadratic in the curvature [39, 41–45] (see [46–48] for reviews). The resulting quantum effective action has the form

$$\Gamma_{\mathrm{one-loop}} = \int d^4x \sqrt{-g} \left[\frac{m_{\mathrm{Pl}}^2}{2} R - R\, k_R(\Box) R - C_{\mu\nu\rho\sigma} k_W(\Box) C^{\mu\nu\rho\sigma} + GB \right], \qquad (2.67)$$

where $m_{\mathrm{Pl}}^2 = 1/(8\pi G)$, $C_{\mu\nu\rho\sigma}$ is the Weyl tensor and 'GB' denotes a similar nonlocal term that reduces to the topological Gauss-Bonnet term when its form factor is set to one. Consider the contribution to the form factor from a particle of mass M. When the particle is very massive compared to the energies or curvatures involved (so M much heavier than the center of mass energy E in a scattering experiment, or M much larger than the Hubble parameter $H(t)$ in a cosmological setting), according to the usual decoupling theorem, the particle decouples and its contribution to the form factor is local and suppressed by a factor $\mathcal{O}(\Box/M^2) \ll 1$. A nonlocal contribution instead emerges when the particle is light compared to the energy scale involved. In that case, the result has the form [44, 45, 49, 50]

$$k_R\left(\frac{-\Box}{M^2}\right) = \alpha \log\left(\frac{-\Box}{M^2}\right) + \beta\left(\frac{M^2}{-\Box}\right) + \gamma\left(\frac{M^2}{-\Box}\right)\log\left(\frac{-\Box}{M^2}\right) + \delta\left(\frac{M^2}{-\Box}\right)^2 + \dots, \quad (2.68)$$

and similarly for k_W. In [51] it was observed that the logarithmic terms and the term $(M^2/\Box)$ have little effect on the cosmological evolution in the present epoch, so one might hope that the leading term is actually given by the term $M^4/\Box^2$, which is the operator that appears in the RR model. Comparison with eq. (2.46) then shows that we must have $M^4 = \mathcal{O}(m^2 m_{\mathrm{Pl}}^2)$ and therefore $m = \mathcal{O}(M^2/m_{\mathrm{Pl}})$. Since the expansion (2.68) is valid, today,

---

[11]This also justifies the integration by parts of $\Box^{-1}$ that we have performed when constructing the RR model in eq. (2.46). At the level of the quantum effective action we can simply define $\Box^{-1}$ with a symmetric Green's function, $G(x,x') = G(x',x)$, which ensure a formal hermiticity of the action and for which the validity of the integration by parts of $\Box^{-1}$ is easily established (see appendix A of [9]). In any case, the equations of motion for the in-in expectation values will come automatically with a retarded Green's function.

only if $M \ll H_0$, such loop corrections could only generate a nonlocal term $m^2 R \Box^{-2} R$ with $m = \mathcal{O}(M^2/m_{\rm Pl}) \ll H_0(H_0/m_{\rm Pl})$. In contrast, we will see that the requirement of obtaining a dynamical dark energy density today of the order of the observed value fixes m to be of order H_0. Thus, loop corrections from light particles, i.e. (hypothetical) massive particles with masses $M \ll H_0$, fall short from providing the required value of m by a factor $\mathcal{O}(H_0/m_{\rm Pl}) \sim 10^{-52}$. On the other hand, in the present cosmological epoch particles with a mass $M \gg H_0$ are heavy compared to the relevant curvature scale fixed by H_0 and only give local contributions to the form factor, furthermore suppressed by $\mathcal{O}(H_0^2/M^2) \ll 1$. Thus, perturbative loop corrections are totally irrelevant to the IR dynamics of gravity.[12] Furthermore, they produce a generic nonlocal structure such as that given in eq. (2.67), while we have already anticipated that, for phenomenological reasons, we need a very specific nonlocal structure such as that in (2.45).

2.4.2 Nonlocal terms from extra dimensions

The above discussion shows that we must look for a different mechanism for the generation of nonlocal terms relevant in a cosmological setting. The Dvali-Gabadadze-Porrati (DGP) model [52], even if by now ruled out phenomenologically, still provides an instructive example of how a theory with a four-dimensional brane in a space with infinite extra dimensions can be rewritten as a four-dimensional covariant theory with nonlocal terms. The DGP action is

$$ S_{\rm DGP} = \frac{M_5^3}{2} \int d^5 X \sqrt{-G}\, R(G) + \frac{M_4^2}{2} \int d^4 x \sqrt{-g}\, R(g) + S_M \,, \qquad (2.69) $$

where $X^A = \{x^\mu, y\}$ are the five-dimensional bulk coordinates, $G_{AB}(X)$ is the 5d metric, and M_5 is the 5d Planck mass; the 4d coordinates, metric and Planck mass are denoted as x^μ, $g_{\mu\nu}(x)$ and M_4, respectively. The 4d metric $g_{\mu\nu}(x)$ is defined as the pullback of the 5d metric, $g_{\mu\nu}(x) = G_{AB}[X(x)]\partial_\mu X^A \partial_\nu X^B$. The matter action, S_M, is localized on the 4d brane.

One can expand the action to quadratic order over flat space, writing $G_{AB}(x,y) = \eta_{AB} + H_{AB}(x,y)$. Away from the brane, the corresponding equations of motions are just the 5d linearized Einstein equations in vacuum. One then finds that it is possible to write explicitly the solutions of the 5d equations of motion for $H_{AB}(x,y)$ in terms of the 4d metric perturbation on the brane, $h_{\mu\nu}(x)$, which plays the role of a boundary value in the equation of motion for $H_{AB}(x,y)$ (the computation is described in detail in section IX.A of ref. [53]). For instance, for the $(A = \mu, B = \nu)$ components of H_{AB}, the result (for a flat brane located at $y = 0$) is of the form

$$ H_{\mu\nu}(x,y) = e^{-|y|\sqrt{-\Box}}\, h_{\mu\nu}(x) \,, \qquad (2.70) $$

where $\sqrt{-\Box}$ is the formal square root of the d'Alembertian operator. Expanding the action (2.69) to quadratic order in H_{AB}, substituting the solution for $H_{AB}(x,y)$ in terms of the boundary value $h_{\mu\nu}(x)$, and integrating the explicit y dependence, one obtains an equivalent

---

[12]Note that this cannot be cured by including a large number N of particles with $M \ll H_0$ in the loops, as has been suggested. Since, for large N, the form factor $k_R(\Box)$ is proportional to N, in that case we rather get $m^2 \sim N M^4/m_{\rm Pl}^2$ and therefore, given that $M \ll H_0$, we now get $m \ll \sqrt{N} H_0(H_0/m_{\rm Pl})$. To obtain $m \sim H_0$ we would then need $N \sim (m_{\rm Pl}/H_0)^2 \sim 10^{104}$. Apart from the fact that existence of such a huge number of hypothetical particles with $m < H_0$ is very implausible, this would result in enhancing all loop corrections of gravity by this factor $\sqrt{N}$, and therefore the scale of quantum gravity would become $m_{\rm Pl}/\sqrt{N} \sim H_0$ in all observables.

nonlocal four-dimensional linearized action,

$$S_{\text{DGP}}^{(2)} = \frac{1}{64\pi G} \int d^4x \left[h_{\mu\nu} \mathcal{E}^{\mu\nu,\rho\sigma} h_{\rho\sigma} - m(h_{\mu\nu}\sqrt{-\Box}\, h^{\mu\nu} - h\sqrt{-\Box}\, h) \right] + \frac{1}{2} \int d^4x\, h_{\mu\nu} T^{\mu\nu} ,$$

$$(2.71)$$

where $m = 2M_5^3/M_4^2$ and, as usual, $M_4^2 \equiv m_{\text{Pl}}^2 = 1/(8\pi G)$. This has the form of Fierz-Pauli massive gravity, with the mass term m^2 replaced by $m\sqrt{-\Box}$.[13] We can now rewrite this expression in a form that involves only the linearized Einstein tensor.[14] The computation can be nicely performed following the steps in section 3 of [9]. That computation was done for Fierz-Pauli massive gravity, but goes through without any changes in our case, with the replacement $m^2 \to m\sqrt{-\Box}$. The strategy, which is analogous to that used in section 2.1 to rewrite the Proca action in nonlocal form, is to introduce a Stückelberg field A_μ by the replacement $h_{\mu\nu} \to h_{\mu\nu} + (1/m)(\partial_\mu A_\nu + \partial_\nu A_\mu)$, to obtain a theory that is explicitly invariant under linearized diffeomorphisms $h_{\mu\nu} \to h_{\mu\nu} - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu)$, $A_\mu \to A_\mu + m\xi_\mu$, and then integrate out A_μ using its own equations of motion, as in [7, 53, 56, 59]. The result can then be read from eq. (3.19) of [9], by replacing $m^2 \to m\sqrt{-\Box}$. Using furthermore $\mathcal{E}_{\mu\nu,\rho\sigma} h^{\rho\sigma} = -2G_{\mu\nu}^{(1)}$ (and rescaling $h_{\mu\nu} \to h_{\mu\nu}/\kappa$) we get

$$\left(1 + \frac{m}{\sqrt{-\Box}}\right) G_{\mu\nu}^{(1)} = 8\pi G \left(T_{\mu\nu} - \frac{1}{3} P_{\mu\nu} T\right) . \tag{2.72}$$

At the linearized level, this nonlocal four-dimensional equation of motion is completely equivalent to the local five-dimensional DGP model. Notice that the extra term on the right-hand side, that survives in the limit $m = 0$, is a reflection of the vDVZ discontinuity of the linearized theory.

One could then in principle look for the correct covariantization that would give back the DGP model at the full nonlinear level.[15] Independently of the correct covariantization (that should also reproduce the absence of the vDVZ discontinuity in the full nonlinear theory), for our purposes the above analysis is instructive because it shows how a nonlocal term, relevant in the IR, can in principle emerge from a theory with infinite extra dimensions. It also shows that, with a mechanism of this kind, one will generate a very specific and peculiar nonlocal structure, rather than the most general expression quadratic in the curvatures that is obtained from perturbative corrections, as in eq. (2.67).

Another interesting example of this type is given by the Karch-Randall model [60], which is a five-dimensional theory of gravity with a negative cosmological constant, giving rise to an AdS$_5$ spacetime, in which is embedded a 4d brane such that the induced metric on the brane is AdS$_4$. The peculiar feature of this compactification is that, despite the fact that four-dimensional general covariance is preserved, still it does not have a massless spin-2 state, but rather a tower of massive spin-2 states. As discussed in [61], from the point of view of an effective four-dimensional action the corresponding mass term cannot be obtained

---

[13]The meaning of the formal expression $\sqrt{-\Box}$ can be better understood by looking at the momentum dependence of the corresponding propagator in momentum space, $-i/(p^2 + m\sqrt{p^2})$. This propagator has a branch cut that corresponds to a continuum of resonances, the so-called 'resonance graviton' [54–56].

[14]We put in a more precise form a result discussed in [57, 58].

[15]A tempting guess for the correct covariantization is obtained by observing that the term $P_{\mu\nu} T$ can be eliminated in favor of $P_{\mu\nu} R^{(1)}$, where $R^{(1)}$ is the linearized Ricci tensor, by taking the trace of eq. (2.72), so that eq. (2.72) can be rewritten as $\left(1 + m/\sqrt{-\Box}\right) [G_{\mu\nu}^{(1)} - \frac{1}{6} P_{\mu\nu} R^{(1)}] = 8\pi G\, T_{\mu\nu}$. As we already observed below eq. (2.42), at linear level the transverse part of a tensor $\eta_{\mu\nu} A(x)$ is $P_{\mu\nu} A(x)$, so $P_{\mu\nu} A = (\eta_{\mu\nu} A)^T$. Then, one is naturally led to $\left[\left(1 + \frac{m}{\sqrt{-\Box}}\right)\left(G_{\mu\nu} - \frac{1}{6} g_{\mu\nu} R\right)\right]^T = 8\pi G\, T_{\mu\nu}$.

from terms quadratic in the curvature, and it was suggested that the proper description is in terms of a nonlocal effective action.

Observe also that, in the example of DGP, the mass scale m is given by $m = 2M_5^3/M_4^2$, i.e. is a combination of the five-dimensional and four-dimensional Planck masses, that were already explicitly present in the fundamental action (2.69). We will next explore a different possibility, namely that such a mass is generated dynamically by non-perturbative IR effects in gravity.

2.4.3 Dynamical mass generation

In this subsection we will discuss indications, from various non-perturbative techniques, in favor of the possibility of a dynamical mass generation in the IR limit of four-dimensional quantum gravity, in particular in relation to the conformal mode.

Lattice gravity. A possible non-perturbative tool is provided by lattice gravity, based either on a simplicial decomposition of the space-time manifold in Euclidean space (see [62] for review), or on causal dynamical triangulations (see [63, 64] for reviews).

In Euclidean quantum gravity one starts from a lattice discretization of the path integral over all Euclidean metrics, weighted with the Euclidean version of the Einstein-Hilbert action with bare cosmological constant Λ_0 and bare Newton constant G_0, and a suitable choice of the lattice measure (and possibly terms quadratic in the curvature). Euclidean lattice gravity is not at the same level of development as, say, lattice QCD, due to the difficulty of finding clear evidence for UV fixed points where one could take a nontrivial continuum limit, so the results should be taken with some qualifications. Still, numerical simulations indicate the existence of a critical coupling G_c such that, for $G_0 < G_c$, the lattice collapses into a degenerate collection of long, elongated simplices, and the four-dimensional geometry collapses into an effective two-dimensional manifold [62]. This phase, that takes place for $G_0 < G_c$ and therefore also in the perturbative regime $G_0 \to 0$, is interpreted as a result of the conformal mode instability in the Euclidean path integral for gravity. As the manifold collapses it reaches an effective dimension equal to two, where the Einstein-Hilbert action becomes a topological invariant, so the instability shuts off and the geometry does not collapse further. In contrast, for $G_0 > G_c$ the system is in a smooth phase. This is interpreted as an effect of the integration measure, that at sufficiently strong coupling suppresses singular spike-like curvature singularities that, in the phase $G_0 < G_c$, trigger the conformal mode instability. At the critical point the correlation length diverges, so a continuum limit can be taken, and, in the vicinity of the critical point the renormalized Newton's constant runs as [65–69] (see also [62, 70] for reviews)

$$G(k^2) = G_N \left[1 + \left(\Lambda_{\mathrm{grav}}^2/k^2 \right)^{\frac{1}{2\nu}} + \mathcal{O} \left(\Lambda_{\mathrm{grav}}^2/k^2 \right)^{\frac{1}{\nu}} \right] , \qquad (2.73)$$

where ν is a critical index which, within the numerical accuracy, turns out to be consistent with $\nu = 1/3$, and Λ_{grav} is a renormalization-group invariant mass scale which is dynamically generated, analogous to Λ_{QCD} in QCD. The expression (2.73) is only valid in the far UV regime $|k^2| \gg \Lambda_{\mathrm{grav}}^2$, and is not directly applicable to the IR regime relevant for cosmology.

The above results point toward the possibility of dynamical mass generation in the IR regime of quantum gravity, but do not yet give hints on what would be the precise role of this mass scale. Recent work using causal dynamical triangulation (CDT), however, indicates precisely the dynamical generation of a mass for the conformal mode [71]. In CDT one

defines the path integral from a sum over Lorentzian geometries, weighted with the factor e^{iS_L}, where S_L is the Einstein-Hilbert action in Lorentzian signature, discretized through triangulations of space-time in terms of simplices with time-like and space-like edges. The four-dimensional lengths of the space-like and time-like edges are defined as $\ell_s^2 = a^2$ and $\ell_t^2 = -\alpha a^2$, respectively, where a is the lattice spacing and $\alpha > 0$. The analytic continuation to $\alpha < 0$ transforms the factor e^{iS_L} in the Lorentzian path integral into e^{-S_E}, where S_E is the Euclidean Einstein-Hilbert action, allowing the use of tools from statistical physics and Monte Carlo techniques for the numerical evaluation of the path integral. Note however that the sum now is not over all (discretized) Euclidean geometries, but only over those that have a causal, Lorentzian, origin, i.e. those that, in the above sense of analytic continuation in α, can be obtained from a Wick rotation of discretized Lorentzian geometries. Thus, the approach to quantum gravity of causal dynamical triangulation is a priori different from that of Euclidean quantum gravity, meant as a sum over all Euclidean geometries.[16] Using numerical simulation of CDT it is possible to measure non-perturbatively the two-point correlation function of the fluctuations of the spatial three-volumes. The latter is related to the two-point function of the conformal mode, and ref. [71] showed that the numerical results provide evidence for a massive conformal mode, i.e. for a linearized nonlocal quantum effective action of the form (2.33) [or, equivalently (2.37)], whose covariantizations can be provided by the RR or RT models. A caveat of the result is that the simulation was performed at a single value of the coupling (κ_0, Δ) of the theory (which are related to the bare Newton constant and the parameter α), but the approach to the continuum was not studied. Still, this is a first indication that a mass for the conformal mode could indeed be generated dynamically in quantum gravity.

Functional renormalization group equations. In quantum field theory, exact renormalization group (RG) equations, such as the Polchinski equation [72] and the Wetterich equation [73], provide, in principle, an equivalent way of computing exactly a path integral, by transforming the functional integration into a functional differential equation. As such, both the path integral formulation and the functional renormalization group equations can be taken as equivalent non-perturbative definitions of a quantum field theory. In practice, just as the evaluation of the functional integral for an interacting theory requires approximations methods (perturbation theory, semiclassical methods such as instantons, etc.) or numerical evaluation through a lattice formulation, the functional RG equation, to be reduced to a manageable form, requires a truncation of the space of action functionals, projecting the intrinsically infinite-dimensional RG flow onto a manifold of finite (and manageable) dimension. In the end, the reliability of the non-perturbative results obtained depends on whether the truncation catches the most important features, and is the main uncertainty of the method. Still, functional RG method can provide important insight into the non-perturbative behavior of a theory. For gravity, functional renormalization group techniques have been developed particularly in connection with the asymptotic safety program, i.e. the search for a non-trivial UV fixed point (see [74] for review). More recently, these tools are being applied to the study

---

[16] The unboundedness of the Euclidean Einstein-Hilbert action, due to the fact that the conformal mode has a kinetic term with the 'wrong' sign, is now regularized by the lattice spacing a. In the limit $a \to 0$ it would again reappear; however, similarly to what we have seen in the case of Euclidean gravity, near a nontrivial fixed point it can happen that configurations with unbounded action are suppressed by the integration measure and play no role in the continuum limit, and this is indeed what happens near the fixed points obtained from causal dynamical triangulations [63]. This competition between configuration with unbounded action and entropy is precisely what gives rise to the Kosterlitz-Thouless transition in the two-dimensional XY-model.

of the IR behavior of gravity. It should be pointed out that the study of the IR behavior of Einstein-Hilbert gravity at the quantum level is completely independent from the issue of its UV completion. Independently of whether the latter is given by a non-trivial UV fixed point, string theory, or other options, we know that gravity, at, say, the laboratory or solar system scales, is very well described by the Einstein-Hilbert action, and we ask how this theory evolves with the RG flow as we run toward even lower energy scales. A number of recent functional RG studies, with different approximations, have found indication of strong quantum gravity effect in the IR [75–78]. The possibility of dynamical mass generation, in the functional RG language, is signaled by the fact that, running toward the IR, the RG flow encounters a singularity at some momentum scale k. An instance of this phenomenon was already found for some RG trajectories in [79], where, in a truncation of the theory including only the Einstein-Hilbert term $\int d^4x \sqrt{-g}\, R$ and the cosmological constant term $\int d^4x \sqrt{-g}$, it was found that, evolving the RG flow toward the IR, for some trajectories the running of Newton's constant hits a singularity and terminates at a finite scale $k_{\rm term}$. Of course, in general the singularity can be an artifact of the truncation. The same happens using functional RG equations in QCD; in that case, a simple truncation of the space of possible terms in the action is not sufficient, and a reliable description of the IR limit involves also nonlocal terms in the truncation ansatz [79], such as the one in eq. (2.15). Thus, by itself a RG flow that, within some truncation, becomes singular in the IR, can be a hint that a mass scale is generated and that at this scale nonlocal terms, that have not been included in the truncation, become important. For our purposes, an interesting observation is that a dynamical scale also appears using functional renormalization group equations because of the dynamics of the conformal mode. Indeed, in the functional RG approach, the would-be 'wrong' sign of the kinetic term of the conformal mode leads to functional differential equations that are perfectly well defined, contrary to the Euclidean path integral formulation, but 'backward-parabolic', i.e. the resulting flow toward the IR is not well defined, and reaches a singularity at a finite energy scale [77, 80]. It is quite natural to expect that, at this scale, nonlocal terms associated to the conformal mode, such as those defining the RR or RT models, become important to resolve the singularity and allow for a smooth flows that extends in the IR down to $k \to 0$.

A related interesting result is the one discussed in ref. [75], where it is found that, truncating the theory so to include only fluctuations of the transverse-traceless modes, there are strong non-perturbative infrared renormalization effects, that screen the cosmological constant. Together, these results can suggest a scenario where the cosmological constant is screened by strong IR effects due to the TT modes fluctuations; at the same time, the conformal mode fluctuations are responsible for generating a new IR mass scale and the nonlocal term that defines the RT model (2.45), that, as we have seen, at the linearized level is simply a mass term for the conformal mode. This nonlocal term, in turn, generates a viable dynamical dark energy model, as we will see in section 3.

It is also important to stress that, contrary to the perturbative corrections as in eq. (2.68), that unavoidably induce the most general structures consistent with the symmetries of the theory, it is perfectly conceivable, and indeed quite natural, that a non-perturbative phenomenon such as a dynamical mass generation could produce a mass for the conformal mode while still leaving massless the $h_{\mu\nu}^{\rm TT}$ mode, as in eq. (2.33). Indeed, even in the usual Higgs mechanism of the Standard Model, the photon remains massless while the $W^{\pm}$ and Z^0 get a mass. In our context, this is particularly natural because the conformal mode appears to be the most 'problematic' one, both because it is the mode of the grav-

itational field with the strongest IR divergences in de Sitter space [2] and because of the conformal mode instability in Euclidean quantum gravity.[17]

Finally, we comment on the naturalness and the numerical value of the mass scale that would be generated dynamically. For simplicity, we illustrate the argument using the RR model, so that we can explain the argument in the more immediate language of the quantum effective action. Equation (2.46) can be rewritten as

$$
\begin{aligned}
\Gamma_{\mathrm{RR}} &= \frac{m_{\mathrm{Pl}}^2}{2} \int d^4x \sqrt{-g} \left[R - \frac{1}{6} m^2 R \frac{1}{\Box^2} R \right] \\
&= \int d^4x \sqrt{-g} \left[\frac{m_{\mathrm{Pl}}^2}{2} R - R \frac{\Lambda_{\mathrm{RR}}^4}{\Box^2} R \right],
\end{aligned}
\tag{2.74}
$$

where $\Lambda_{\mathrm{RR}} = (1/12) m^2 m_{\mathrm{Pl}}^2$. In this form, it is clear that Λ_{RR} should be taken as the fundamental scale generated dynamically, corresponding to a dimensionless form factor $k_R(\Box) = \Lambda_{\mathrm{RR}}^4 / \Box^2$ in $R k_R(\Box) R$, while the parameter m is just a derived quantity introduced for convenience. The value of a scale generated generated dynamically in this way cannot be predicted, just as we cannot predict the value of Λ_{QCD}, and can only be obtained by comparison with the observation. In our case, as we will see below, we need $m = O(H_0)$ in order to have a dark energy that becomes important near the present epoch.[18] Therefore,

$$
\Lambda_{\mathrm{RR}} = O(H_0 m_{\mathrm{Pl}})^{1/2} = O(\mathrm{meV}).
\tag{2.75}
$$

The same holds for the RT model, as we see by rewriting eq. (2.45) as

$$
m_{\mathrm{Pl}}^2 G_{\mu\nu} - \Lambda_{\mathrm{RT}}^4 \left(g_{\mu\nu} \Box^{-1} R \right)^{\mathrm{T}} = T_{\mu\nu},
\tag{2.76}
$$

where we have now defined $\Lambda_{\mathrm{RT}}^4 = (1/3) m_{\mathrm{Pl}}^2 m^2$. Thus, in the RR or RT model dark energy can be explained by the dynamical generation of an energy scale whose value, of the order of the milli-eV (or some orders of magnitude smaller, see footnote 18), even if cannot be predicted, is not particularly surprising from the point of view of quantum field theory. This is different from attempts at explaining dark energy through the introduction of some particle of mass m, in which case m is the fundamental scale and should be fixed to the extremely small value $m \sim H_0 \sim 10^{-33}$ eV. Notice also that there is no problem of technical naturalness associated to a scale such as Λ_{RT} since, just as Λ_{QCD}, a mass scale which is generated dynamically in this way is a renormalization group invariant.

3 Phenomenology of the RT model

We now have all the elements for working out the predictions of nonlocal gravity. We focus on the RT model, that eventually turns out to be the most interesting phenomenologically. In order to make the paper self-contained, we begin by reviewing material on the background

---

[17]Indications for dynamical generation of a mass scale may also come from the running of the coupling constant associated to terms quadratic in the curvature [81], and in particular the Gauss-Bonnet term, whose coupling is asymptotically free and generates an IR scale through dimensional transmutation, exponentially suppressed, with respect to the Planck mass, by instanton effects [82, 83].

[18]More precisely, we will see in section 3.1.2 that the model has a significantly different evolution if initial conditions of order one are set during radiation dominance or during an earlier inflationary phase. In the former case $m \sim H_0$ with a numerical coefficient of order one, while in the latter case m can be numerically much smaller than H_0. Equation (2.75) therefore only holds in the former case.

evolution already discussed and reviewed in [20, 22, 23, 84]. We will then move to a detailed discussion of the perturbations and an updated comparison with the cosmological data. We will finally discuss GW propagation in the RT model and show that this leads to very interesting effects that could be detected in the near future with GW detectors.

3.1 Background evolution

3.1.1 Equations in FRW

We consider a spatially flat Friedman-Robertson-Walker (FRW) background, $ds^2 = -dt^2 + a^2(t)d\mathbf{x}^2$. For symmetry reasons the spatial component S_i of the auxiliary field S_μ must vanish, since there is no preferred spatial direction,[19] and the only variables are $U(t)$ and $S_0(t)$, together with the FRW scale factor $a(t)$. Eqs. (2.53)–(2.55) then become [20]

$$H^2 - \frac{m^2}{9}(U - \dot{S}_0) = \frac{8\pi G}{3}\rho \tag{3.1}$$

$$\ddot{U} + 3H\dot{U} = 6\dot{H} + 12H^2\,, \tag{3.2}$$

$$\ddot{S}_0 + 3H\dot{S}_0 - 3H^2 S_0 = \dot{U}\,, \tag{3.3}$$

where we have written $T_\nu^\mu = \mathrm{diag}(-\rho, p, p, p)$, and the dot denotes the derivative with respect to cosmic time t. It is convenient to define $Y = U - \dot{S}_0$, $h = H/H_0$, and $\Omega_i(t) = \rho_i(t)/\rho_c(t)$, where $i = \mathrm{M}, \mathrm{R}, \mathrm{DE}$ labels radiation, matter and dark energy, respectively, and $\rho_c(t) = 3H^2(t)/(8\pi G)$. We will also use the standard notation $\Omega_M \equiv \Omega_M(t_0)$, $\Omega_R \equiv \Omega_R(t_0)$ and $\Omega_{\mathrm{DE}} \equiv \Omega_{\mathrm{DE}}(t_0)$ (where t_0 is the present value of cosmic time) for the present values of $\Omega_i(t)$. We henceforth use the dimensionless variables

$$x \equiv \ln a(t) \tag{3.4}$$

instead of cosmic time t, and we denote $df/dx = f'$. Then the Friedmann equation (3.1) reads

$$h^2(x) = \Omega_M e^{-3x} + \Omega_R e^{-4x} + \gamma Y(x)\,, \tag{3.5}$$

where

$$\gamma \equiv m^2/(9H_0^2)\,. \tag{3.6}$$

---

[19]A recent paper [85] has studied the evolution of the RT model in FRW by setting $S_\mu(t) = (S_0(t), v(t), v(t), v(t))$ and claimed that this is the most general ansatz consistent with the rotational invariance of FRW. This is clearly wrong, since this ansatz selects a privileged spatial direction $\mathbf{S}(t) = v(t)(\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}})$, and therefore breaks the rotational invariance of FRW. The authors of [85] appear to have made confusion with the fact that, for a perfect fluid in FRW, T_ν^μ has the form $\mathrm{diag}(-\rho, p, p, p)$. Obviously, the trace T_i^i of a tensor is invariant under spatial rotations, while a spatial vector S_i is not! With a rotation we can bring the unit vector $(\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}})/\sqrt{3}$ onto the $\hat{\mathbf{z}}$ axis, and in this frame the choice of [85] becomes $\mathbf{S}(t) = \sqrt{3}v(t)\hat{\mathbf{z}}$. This ansatz therefore is not consistent with the isotropy of FRW at the background level. Perturbations over FRW do not have to respect the isotropy, so S_i will be non-vanishing at the perturbative level. As we will discuss in section 3.2.1, the vector S_i contributes to scalar perturbations through fluctuations of the form $S_i = \partial_i(\delta S)$. Since the i-th component of the vector $\mathbf{S} = (v(t), v(t), v(t))$ can be written as $S_i = \partial_i[v(t)r]$, where $r = |x|$, what ref. [85] is actually doing is to add to the background solution of the model an unphysical scalar perturbation $\delta S(t, r) = v(t)r$ that grows in space without bounds, radially from an arbitrarily chosen origin. Treating S_i correctly as a perturbation, the full equation for its evolution is not the one given in eq. (2.4) of ref. [85]. Rather, it involves all other first-order quantities [see eqs. (A.6)-(A.10) of [86] for the full set of equations] and, as we will review in section 3.2.1, the corresponding perturbations are stable.

This shows that there is an effective DE density

$$\rho_{\rm DE}(t) = \rho_0 \gamma Y(x) \,, \tag{3.7}$$

where $\rho_0 = 3H_0^2/(8\pi G)$. Using $U(x)$ and $Y(x)$, eqs. (3.2) and (3.3) take the form

$$Y'' + (3 - \zeta)Y' - 3(1 + \zeta)Y = 3U' - 3(1 + \zeta)U \,, \tag{3.8}$$

$$U'' + (3 + \zeta)U' = 6(2 + \zeta) \,, \tag{3.9}$$

where, using eq. (3.5),

$$\zeta(x) \equiv \frac{h'}{h} = -\frac{3\Omega_M e^{-3x} + 4\Omega_R e^{-4x} - \gamma Y'}{2(\Omega_M e^{-3x} + \Omega_R e^{-4x} + \gamma Y)} \,. \tag{3.10}$$

3.1.2   Initial conditions; the parameter ΔN

As a next step, we discuss the initial conditions on the auxiliary fields (we follow refs. [22, 87]). To get a first analytic understanding we observe that, in any given cosmological epoch, such as radiation dominance (RD), matter dominance (MD), or an earlier inflationary de Sitter (dS) phase, $\zeta(x)$ has an approximately constant value ζ_0, with $\zeta_0 = 0$ in dS, $\zeta_0 = -2$ in RD and $\zeta_0 = -3/2$ in MD. In the approximation of constant ζ eq. (3.9) can be integrated analytically, and has the solution [20]

$$U(x) = \frac{6(2 + \zeta_0)}{3 + \zeta_0}x + u_0 + u_1 e^{-(3+\zeta_0)x} \,. \tag{3.11}$$

The first term on the right-hand side is a particular solution of the inhomogeneous equation, while u_0 and u_1 parametrize the most general solution of the homogeneous equation $\Box U = U'' + (3 + \zeta_0)U = 0$. The initial conditions on U, i.e. $U(x_{\rm in})$ and $U'(x_{\rm in})$, are in one-to-one correspondence with the choice of the solutions of homogeneous equation, i.e. with u_0 and u_1. The constant u_0 corresponds to the reintroduction of a cosmological constant, as we have seen in eq. (2.56). Our aim is to see if we can obtain a self-accelerated evolution from the nonlocal term, without introducing by hand a cosmological constant, and we will therefore set $u_0 = 0$. A non-vanishing u_0 could always be reintroduced later, and, not surprisingly, produces an evolution that is intermediate between that of the RT model with $u_0 = 0$ and that ΛCDM, see section 7.4 of [22]. The other solution of the homogeneous equation, proportional to $e^{-(3+\zeta_0)x}$, is instead a decaying mode, in all cosmological phases. Thus, the solution with initial conditions $U(x_{\rm in}) = U'(x_{\rm in}) = 0$ has a marginally stable direction, corresponding to the possibility of reintroducing a cosmological constant, and a stable direction, i.e. is an attractor in the u_1 direction. Consider next eq. (3.8). Using eq. (3.11) and solving for $Y(x)$ we get [20]

$$Y(x) = -\frac{2(2 + \zeta_0)\zeta_0}{(3 + \zeta_0)(1 + \zeta_0)} + \frac{6(2 + \zeta_0)}{3 + \zeta_0}x + u_0 - \frac{6(2 + \zeta_0)u_1}{2\zeta_0^2 + 3\zeta_0 - 3}e^{-(3+\zeta_0)x}$$

$$+ a_1 e^{\alpha_+ x} + a_2 e^{\alpha_- x} \,, \tag{3.12}$$

where $\alpha_\pm = (1/2)[-3 + \zeta_0 \pm \sqrt{21 + 6\zeta_0 + \zeta_0^2}]$. In both RD and MD we have $\alpha_+ < 0$ and $\alpha_- < 0$, so both modes are decaying. This means that, if we start the evolution deep in the RD phase, with $u_0 = 0$ in order not to have a cosmological constant, and $u_1 \sim$

$a_1 \sim a_2 \sim \mathcal{O}(1)$, the solution will quickly approach the one obtained with initial conditions $U(x_{\rm in}) = U'(x_{\rm in}) = Y(x_{\rm in}) = Y'(x_{\rm in}) = 0$. We will refer to this solution as the 'minimal' RT model.

The situation becomes more interesting if we start the evolution during a primordial phase of de Sitter-like inflation, before RD. In dS there is a growing mode with $\alpha_+ = (-3 + \sqrt{21})/2 \simeq 0.79$. Then Y will grow during dS (exponentially in x, so as a power of the scale factor), and will then decrease again during RD and MD. In general, a growing mode during MD or the late RD phase would be fatal to the viability of the model, because any perturbation of the initial conditions would result in an activation of the unstable mode, and would bring the solution very far from a FRW solution driven by $T_{\mu\nu}$, as in standard cosmology (this is indeed a criterium that ruled out several other nonlocal models, as we will recall in appendix A). For the evolution during an early dS phase the situation is, however, different [22, 87]. Indeed, let us denote by $x_{\rm in}$ the value of $x = \ln a$ at a time, during inflation, when we set initial conditions $u_1 \sim a_1 \sim a_2 \sim \mathcal{O}(1)$, and by $x_{\rm end}$ the value when inflation ends and RD begins (we neglect for simplicity an intermediate reheating phase). We use the notation

$$x_{\rm end} - x_{\rm in} = \log\left(a_{\rm end}/a_{\rm in}\right) \equiv \Delta N \,, \tag{3.13}$$

so ΔN is the number of e-folds from the time where we set initial condition of order one, to the end of a de Sitter phase of inflation. Thus, if $Y(x_{\rm in})$ has a generic value of order one (i.e., is not fine-tuned to zero), by the end of inflation

$$Y(x_{\rm end}) \simeq \exp\{\alpha_+^{\rm dS} \Delta N\} \simeq \exp\{0.79 \Delta N\} \,. \tag{3.14}$$

The evolution of U can be computed similarly, using eq. (3.11). During a quasi-de Sitter phase of inflation, starting from a value of order one, we get

$$U(x_{\rm end}) \simeq 4\Delta N \,. \tag{3.15}$$

The important point is that, despite the exponential growth in eq. (3.14), even for very large values of ΔN the corresponding DE density $\rho_{\rm DE}(x) = \rho_0 \gamma Y(x)$ has no effect on the inflationary dynamics. This is due to the fact that $\rho_0 = 3H_0^2/(8\pi G) \sim (10^{-3} {\rm eV})^4$ is extremely small compared to the energy density during inflation. For instance, if $Y(x_{\rm in}) = \mathcal{O}(1)$ and we take $\Delta N = 60$, at the end of inflation we get $Y(x_{\rm end}) = \mathcal{O}(10^{20})$. Even with such a large value of Y, we have

$$[\rho_0 Y(x_{\rm end})]^{1/4} \sim 10^{-3} {\rm eV} \times Y^{1/4}(x_{\rm end}) \sim 10^2 \, {\rm eV} \,. \tag{3.16}$$

This is totally negligible compared to the inflationary scale M, that has typical values, say, of order 10^{13} GeV. Thus, during the inflationary phase the evolution of the scale factor is the same as in standard GR without the nonlocal term. So, the important conclusion is that, at the level of background evolution, there is no evident pathology associated with the exponential growth of $Y(x)$. Rather, one will have to study in detail the evolution through dS, RD and MD to see if it gives a viable and interesting background cosmology. As we will recall below, following [22, 87], indeed the corresponding background evolution is viable, and also quite interesting. As discussed in [87], even at the level of cosmological perturbations this growth during de Sitter is innocuous, again because of the smallness of the scale associated to the nonlocal term with respect to the inflationary scale.

Equations (3.14) and (3.15) give the values of $Y(x)$ and $U(x)$ when they enter the subsequent RD phase (apart from some minor modification due to reheating). As we will

see explicitly in section 3.1.3, even if in the RD and MD phases the solution obtained with vanishing initial conditions is an attractor, the fact that $Y(x)$ enters the RD phase with an exponentially large value gives an evolution that is sensibly different from that of the minimal model, simply because there is not enough time to relax to zero this exponentially large value by the end of the MD phase and the beginning of the current DE-dominated phase, when (having chosen m of order H_0) the energy scale associated to Y eventually becomes comparable to the total energy density. Thus, there is a residual dependence of the dark energy evolution near the present cosmological epoch, from the value ΔN that determines, through eqs. (3.14) and (3.15), the values of the auxiliary field when the enter the RD phase.

The conclusion is that, at the level of the background cosmological evolution, our ignorance on the initial conditions of the auxiliary fields can be reabsorbed into a single parameter ΔN, that gives the number of e-folds from the moment when these fields have initial conditions $\mathcal{O}(1)$ during inflation, until the end of inflation (plus the parameter u_0, that corresponds to reintroducing a cosmological constant, and that we will set to zero).

As discussed in [23], no further freedom emerges at the level of cosmological perturbations. Indeed, at the perturbation level we must consider all Fourier modes of the perturbations, so in principle we should assign the initial conditions on $\delta U_{\mathbf{k}}(x)$, $\delta Y_{\mathbf{k}}(x)$ and on their first time derivatives, at an initial time $x_{\rm in}$. The fact that the auxiliary fields do not represent arbitrary degrees of freedom but are fixed in terms of the metric means that the initial conditions for the perturbations of the auxiliary fields will be of order of the metric perturbations. One can therefore ask what happens if we start with initial conditions of this order of magnitude. The explicit numerical study in [23] shows that the effect of such a change in the initial conditions of the perturbations is completely negligible.

In the rest of this paper we will study the predictions of the RT model for a few values of ΔN. For this purpose, it is useful to recall that, for inflation taking place at a scale $M_{\rm infl} = (\rho_{\rm infl})^{1/4}$, assuming instantaneous reheating, the minimum number of e-folds required to solve the flatness and horizon problems is (see e.g. section 21.1 of [17])

$$(\Delta N)_{\rm min} \simeq 64 - \log \frac{10^{16}\,{\rm GeV}}{M_{\rm infl}} \,. \tag{3.17}$$

In the following, beside the 'minimal' model defined by initial conditions of order one during RD, which is equivalent to setting $\Delta N = 0$ (or, equivalently, ΔN of order one), we will study also the cases $\Delta N = 34, 50, 64$ that, according to eq. (3.17), approximately correspond to the minimum value of ΔN for $M_{\rm infl} = \{10^3, 10^{10}, 10^{16}\}\,{\rm GeV}$, respectively. These values are chosen because $M_{\rm infl} = 10^3\,{\rm GeV}$ corresponds to inflation at the electroweak scale, which is on the lower range of possible inflationary scales, while $M_{\rm infl} = 10^{16}\,{\rm GeV}$ is the highest value consistent with the non-detection of tensor perturbations in the CMB anisotropies, and $M_{\rm infl} = 10^{10}\,{\rm GeV}$ is an intermediate value which is quite often considered as a typical inflationary scale.

Of course, the number of e-folds during inflation at a given scale does not need to be the minimum required to solve the flatness and horizon problems and, for a given value of $M_{\rm infl}$, we could chose a higher value of ΔN. We have therefore studied also how the results change increasing ΔN for a fixed value of $M_{\rm infl}$. As already pointed out in [87], increasing ΔN the results eventually saturate to a limiting curve (as a function of redshift). In particular, setting $M_{\rm infl} = 10^{16}\,{\rm GeV}$, we find that this limiting curve is reached, within sub-percent level accuracy, already for $\Delta N \simeq 70$. In the following, beside the cases ($M_{\rm infl} = 10^3\,{\rm GeV}, \Delta N = 34$), ($M_{\rm infl} = 10^{10}\,{\rm GeV}, \Delta N = 50$) and ($M_{\rm infl} = 10^{16}\,{\rm GeV}, \Delta N = 64$), we will also show the

results for $(M_{\rm infl} = 10^{16}\,{\rm GeV}, \Delta N = 100)$, that represents the limiting curve for the various background quantities as a function of redshift. For brevity, we will refer to these cases as the RT model with $\Delta N = 34, 50, 64$ and 100, respectively. We have checked that the same limiting curve is obtained starting from a different value of $M_{\rm infl}$ and raising again sufficiently ΔN. This behavior is due to a scaling property of the equations when Y starts from a very large value at the beginning of RD [87].

A quite interesting aspect of the cosmological evolution of the RT model with initial conditions set during inflation, that will emerge clearly from the discussion below, is that the behavior of dark energy at the present epoch depends on the existence and duration (as quantified by ΔN) of a phase of primordial inflation, providing an unexpected connection between early- and late-time cosmology.

3.1.3 Results: $\rho_{\rm DE}(z)$, $w_{\rm DE}(z)$, $H(z)$

Given the initial conditions and a choice of values for the cosmological parameters Ω_M and h_0 (defined as usual from $H_0 = 100 h_0\,{\rm km\,s^{-1}Mpc^{-1}}$), the numerical integration of the equations for the background evolution, eqs. (3.5)–(3.9), is straightforward.[20] In the following figures we show the results for the minimal RT model and for the RT model with $\Delta N = 34, 50, 64, 100$.

The upper left panel of figure 1 shows the evolution of the dark energy density $\rho_{\rm DE}(x)$, normalized to the total energy density $\rho_{\rm tot}(x) = \rho_M(x) + \rho_R(x) + \rho_{\rm DE}(x)$, as a function of x [recall that here $x = \ln a$, and we normalize the scale factor so that $a(t_0) = 1$]. For orientation, matter-radiation equilibrium is at $x \simeq -8.1$, at the present epoch $x = 0$, and $x > 0$ corresponds to the cosmological future. We see from the plot that the DE density due to the nonlocal term is negligible until the relatively recent cosmological epoch, when eventually dominates.

---

[20]In practice, in the numerical implementation of our integration routine, we consider that the transition between inflation and RD takes place when, extrapolating backward in time the present energy density in radiation $\rho_{R,0}$, the energy density in radiation $\rho_{R,0}/a^4$ becomes equal to $M_{\rm infl}^4$, i.e. when the scale factor has the value $a_\star$ given by $a_\star = \rho_{R,0}^{1/4}/M_{\rm infl}$ (notice that the quantity $M_{\rm infl}$ defined in this way corresponds to the actual inflationary scale only in the approximation of instantaneous reheating). Using $\rho_{R,0}^{1/4} \simeq 2.41 \times 10^{-4}\,{\rm eV}$, the corresponding value of $x = \log a$ is $x_\star \simeq -65.9 + \log(10^{16}\,{\rm GeV}/M_{\rm infl})$. Assuming that initial conditions of order one have been set ΔN e-folds earlier, at the inflation-RD transition we take $Y = \exp\{0.79\Delta N\}$ and $U = 4\Delta N$. The numerical integration through the full RD phase would be numerically difficult, and not necessary, since we know that, until we are deep in RD, the solution for Y evolves according to the slowest-decaying mode, which decays as $\exp\{-0.70x\}$ and the solution for U stays constant. Thus, at a value x_0 still deep into RD (we take $x_0 = -15$; RD-MD equilibrium is at $x \simeq -8.1$) we have $U(x_0) = 4\Delta N$, $U'(x_0) = 0$, $Y(x_0) = \exp\{0.79\Delta N - 0.70(x_0 - x_\star)\}$ and $Y'(x_0) = -0.70Y(x_0)$. At this point we start the numerical evolution with these initial conditions. To produce figure 1, for the minimal model and for $\Delta N = 34, 50, 64$ we have used the respective mean values for Ω_M and h_0 from table 2, obtained from our MCMC chains. For the limiting curves $\Delta N = 100$ we have not rerun our MCMC and we have used the same values as for $\Delta N = 64$, which is an excellent approximation since we see from table 2 that, for large ΔN, the variation in the parameters are very small (and would give effects totally unappreciable on the scale of the figures). A final detail is that, in ΛCDM, assuming flatness and fixing Ω_M and Ω_R, directly fixes Ω_Λ from $\Omega_M + \Omega_R + \Omega_\Lambda = 1$, and one can immediately integrate the evolution equations. In contrast, in the nonlocal model, once fixed Ω_M and Ω_R (and assuming flatness), the remaining parameter in the equations is γ, which is fixed by trials and errors until the value of the dark energy energy fraction today, $\Omega_{\rm DE}$, obtained from the solution of the equations, satisfies the condition $\Omega_M + \Omega_R + \Omega_{\rm DE} = 1$, i.e. $\Omega_{\rm DE} \simeq 0.7$. The corresponding values of γ turn out to be $\gamma \simeq 5.13555 \times 10^{-2}$ for the minimal model, and $\gamma \simeq \{2.69512 \times 10^{-3}, 1.0321 \times 10^{-3}, 3.73915 \times 10^{-4}, 1.94944 \times 10^{-11}\}$ for $\Delta N = 34, 50, 64, 100$, respectively. For the mass m this means $m/H_0 \simeq 0.68$ for the minimal model, and $m/H_0 \simeq \{0.16, 0.10, 0.06, 4.2 \times 10^{-8}\}$ for $\Delta N = 34, 50, 64, 100$. We perform the numerical integration of the differential equations both with Mathematica and with CLASS, and we check the consistency of the results.
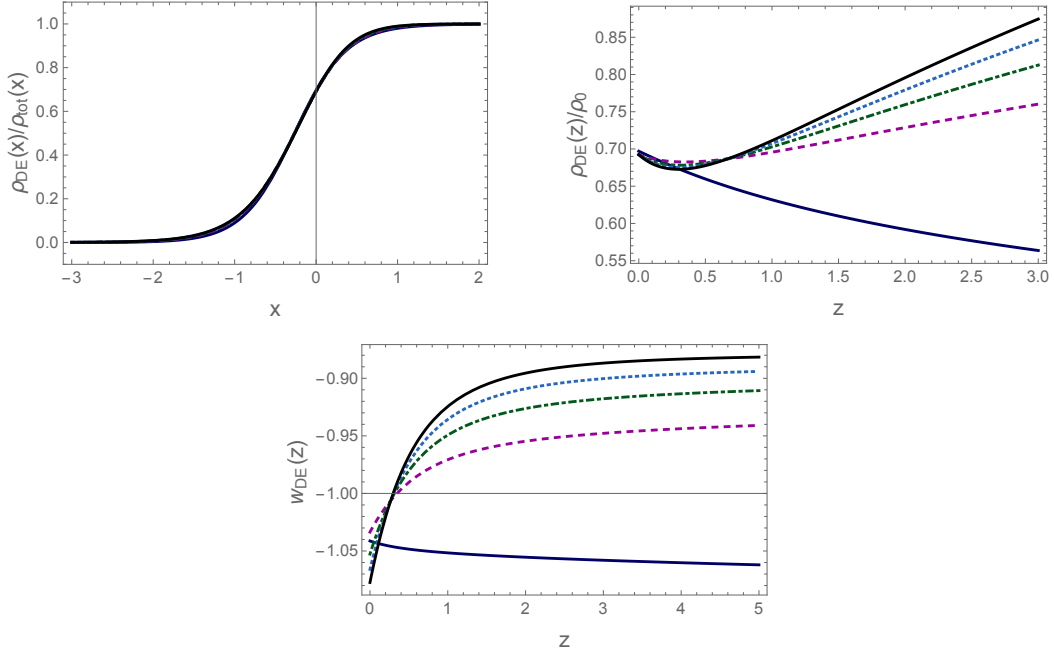
Figure 1. Upper left panel: $\rho_{\rm DE}(x)$ normalized to the total energy density $\rho_{\rm tot}(x)$ as a function of x. Upper right panel: $\rho_{\rm DE}(z)$ normalized to the critical energy density today, ρ_0, as a function of redshift z. Lower panel: the DE equation of state $w_{\rm DE}(z)$ as a function of redshift. The curves correspond to the minimal RT model (blue solid line) and the RT model with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed), $\Delta N = 64$ (cyan, dotted) and $\Delta N = 100$ (black solid line).

When $\rho_{\rm DE}(x)$ is normalized to $\rho_{\rm tot}(x)$, which includes the contribution of $\rho_{\rm DE}(x)$ itself, the result for the minimal model and for the RT models with large ΔN look all very similar, and the various curves are basically indistinguishable. However, the individual behaviors of $\rho_{\rm DE}(x)$ are quite different. This is shown in the upper right panel of figure 1, where $\rho_{\rm DE}$ is shown as a function of the redshift z [related to x by $x = -\log(1 + z)$], and normalized to the constant critical energy density today ρ_0. We see that, as we approach the present epoch from large z, in the minimal model $\rho_{\rm DE}$ increases, until it reaches the present value $\rho_{\rm DE}/\rho_0 \simeq 0.7$, which is fixed by our choice of $\Omega_M \simeq 0.3$. In contrast, for large ΔN, $\rho_{\rm DE}$ starts from a very large value deep in RD (a consequence of the large value of the auxiliary field Y at the end of inflation), and then decreases for most of its evolution, until the present epoch. This behavior can be understood observing that, for $\Delta N = 0$, the evolution of Y is determined by the particular solution of the inhomogeneous equation (3.8), which stays close to zero during RD and then starts to increases with time during MD, until we enter in a regime dominated by DE; in contrast, for large ΔN the solution starts from a very large initial value at the beginning of RD and then decays according to the decaying modes of the associated homogeneous equation, until, close to the recent epoch, the decaying modes have become smaller than the solution of the inhomogeneous equation, that takes over, so the solution for Y starts to rise again.

As mentioned before, for sufficiently large ΔN, the results saturate toward a limiting curve, independent of the chosen value of $M_{\rm infl}$. As explained in [87], this is due to the fact that, for sufficiently large ΔN, an increase in the initial values of Y at the beginning of RD is exactly compensated by a decrease in γ, and we end up on the same solution. This
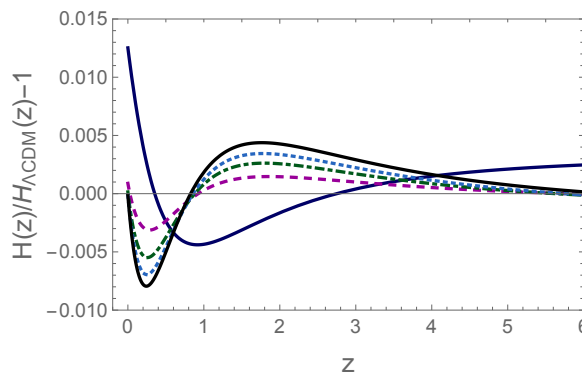
Figure 2. Relative difference of Hubble rate with respect to ΛCDM for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed), $\Delta N = 64$ (cyan, dotted) and $\Delta N = 100$ (black solid line).

	RT, minimal	$\Delta N = 34$	$\Delta N = 50$	$\Delta N = 64$	$\Delta N = 100$
w_0	-1.041	-1.034	-1.053	-1.066	-1.077
w_a	-0.023	$+0.127$	$+0.218$	$+0.283$	$+0.335$

Table 1. Values of w_0 and w_a for the RT model, minimal and with various values of ΔN.

limiting curve is shown as the black solid line in figure 1, obtained for definiteness setting $(M_{\rm infl} = 10^{16}\,{\rm GeV}, \Delta N = 100)$. For instance, in this and in all similar plots below, on the scale of the figure all the curves with $M_{\rm infl} = 10^{16}\,{\rm GeV}$ and $\Delta N \gtrsim 70$ are indistinguishable, and fall on this asymptotic curve.

The lower panel in figure 1 shows the DE equation of state, defined as usual from the conservation equation

$$\dot\rho_{\rm DE} + 3H(1 + w_{\rm DE})\rho_{\rm DE} = 0\,. \tag{3.18}$$

The different evolutions of $\rho_{\rm DE}$ for the minimal model and for large ΔN result in different, and quite distinctive behaviors of $w_{\rm DE}$ as a function of redshift. For the minimal model $w_{\rm DE}(z)$ is always on the 'phantom' side, $w_{\rm DE}(z) < -1$, while, for large ΔN, the evolution exhibits 'phantom crossing' at $z \simeq 0.30 - 0.35$. In all cases, we see that the DE density starts to dominate near the present cosmological epoch, and its equation of state corresponds to accelerated expansion. Thus, the nonlocal term generates a dynamical DE density that drives an accelerated expansion of the Universe at the current cosmological epoch. This is already a very non-trivial result: it means that giving a mass to the conformal mode, and covariantizing it as discussed in section 2.2.3, provides an explanation for the observed accelerated expansion of the Universe.

Figure 2 shows the relative difference $[H_{\rm RT}(z) - H_{\Lambda{\rm CDM}}(z)]/H_{\Lambda{\rm CDM}}(z)$ between each RT model (minimal and with $\Delta N = 34, 50, 64, 100$) and ΛCDM. Once again, the predictions of each model are computed using the respective mean values of the cosmological parameters in table 2. At $z = 0$ the difference between the various curves is due to the different mean values for H_0, and at large z (but still within MD) it is determined by the different mean values for Ω_M. We see that, at $z = 0$, the minimal RT model differs from ΛCDM by about 1%, while the RT models with large ΔN give a prediction for H_0 basically indistinguishable from that of ΛCDM. Away from $z = 0$, $|\Delta H(z)|/H(z)$ is of order 0.5% or less. The evolution with redshift is, however, quite distinctive, with $\Delta H(z)/H(z)$ oscillating and changing sign
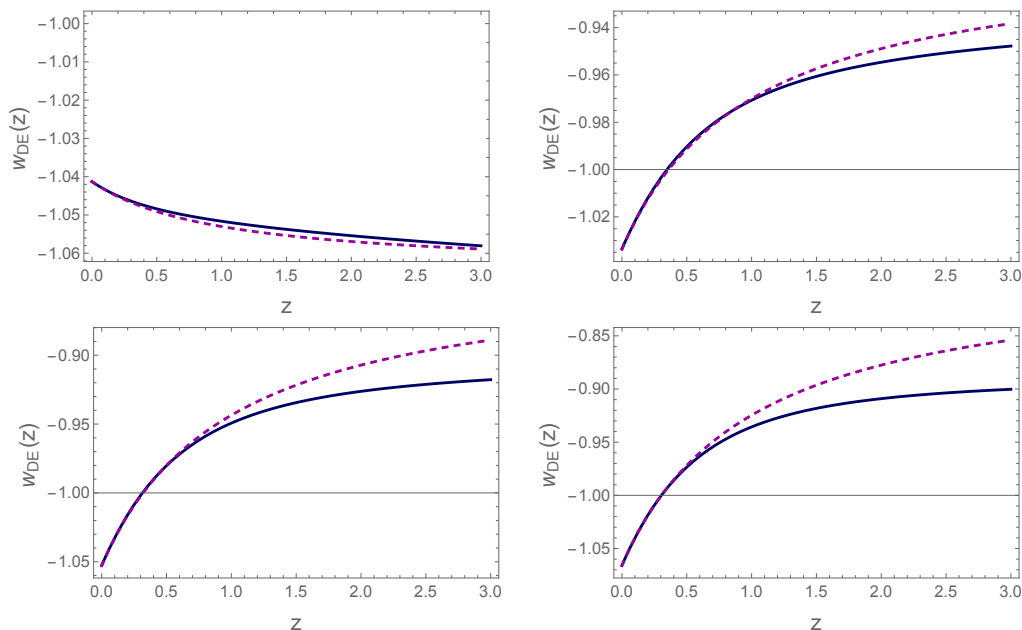
Figure 3. The DE equation of state $w_{\rm DE}(z)$ from the numerical integration of the equations (blue solid lines), compared with the parametrization (3.19) (magenta dashed lines) for RT minimal (upper left panel) and RT with $\Delta N = 34$ (upper right), $\Delta N = 50$ (lower left) and $\Delta N = 64$ (lower right).

as z increases. These differences with respect to ΛCDM can be compared to a compilation of measurements of $H(z)$ at different redshifts. We will perform this test in section 3.3, after having performed the Bayesian parameter estimation for the models.

It is interesting to compare the actual predictions of the model to the results obtained with the standard (w_0, w_a) parametrization $w_{\rm DE}(a) = w_0 + (1-a)w_a$ [88, 89], or, in terms of redshift,

$$w_{\rm DE}(z) = w_0 + \frac{z}{1+z} w_a \,. \tag{3.19}$$

Setting $w_0 \equiv w(a = 1)$ and $w_a \equiv -(dw/da)_{|a=1}$ we get the values of w_0 and w_a given in table 1. In figure 3 we compare the actual numerical result for $w(z)$ to the fit provided by this parametrization. We see that, for large ΔN, the parametrization (3.19) is not very accurate beyond some value of z, with the range in z shrinking as ΔN increases.

3.2 Scalar perturbations

3.2.1 Formalism

Cosmological scalar perturbations for the RR and RT model (in the minimal case) have been studied in detail in [86, 90] (see also [22] for review). Here, after recalling the basic formalism, we will extend the results to the RT model with large ΔN and we will present updated results on various indicators of cosmological perturbations, using the values of the cosmological parameters that will be determined in section 3.3 by the comparison with observations. We work in the Newtonian gauge, where, in the scalar perturbation sector, the perturbed FRW metric has the form

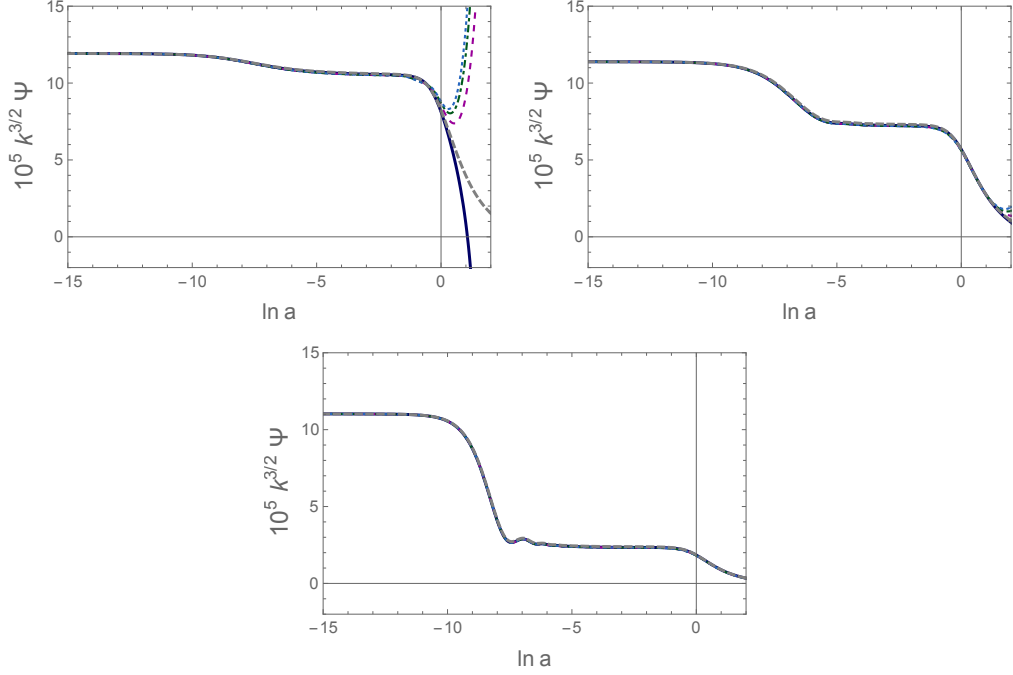$$ds^2 = -(1 + 2\Psi)dt^2 + a^2(t)(1 + 2\Phi)\delta_{ij}dx^i dx^j \,, \tag{3.20}$$

Figure 4. $k^{3/2}\Psi(a;k)$ in ΛCDM (gray dashed line), the minimal RT model (blue solid line) and RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted) for $\kappa = 0.1$ (upper left panel), $\kappa = 1$ (upper right) and $\kappa = 5$ (lower panel). On the scale of these figures, the results for $\kappa = 5$ are indistinguishable among the models, while for $\kappa = 1$ one can barely distinguish some small differences in the cosmological future, $x > 0$. Observe that the quantity that we plot is $k^{3/2}\Psi(a;k)$ multiplied by a factor 10^5. Matter-radiation equilibrium is at $x \simeq -8.1$, and in this region one sees the usual transition between two different plateaux in Ψ.

where Φ and Ψ are the Bardeen variables. We similarly perturb the auxiliary fields, writing

$$U(t, \mathbf{x}) = \bar{U}(t) + \delta U(t, \mathbf{x}), \qquad S_\mu(t, \mathbf{x}) = \bar{S}_\mu(t) + \delta S_\mu(t, \mathbf{x}), \tag{3.21}$$

where, in this section, background quantities are denoted with an overbar. In FRW, $\bar{S}_i$ vanishes because at the background level there is no preferred spatial direction, but its perturbation δS_i is non-vanishing. As with any vector, we can decompose it into a transverse and longitudinal part, $\delta S_i = \delta S_i^{\mathrm{T}} + \partial_i(\delta S)$, where $\partial_i(\delta S_i^{\mathrm{T}}) = 0$. Since we are considering scalar perturbations, we only retain δS. Thus, in the RT model the metric perturbations in the scalar sector are described by $\Psi, \Phi, \delta U, \delta S_0$ and δS. It is convenient to trade S_0 and S for

$$V = H_0 S_0, \qquad Z = H_0^2 S, \tag{3.22}$$

so we eventually work with the variables $\{\Psi, \Phi, \delta U, \delta V, \delta Z\}$.[21] We similarly perform the usual expansion of the energy-momentum tensor, writing

$$T_0^0 = -(\bar{\rho} + \delta\rho), \qquad T_i^0 = (\bar{\rho} + \bar{p})v_i, \qquad T_j^i = (\bar{p} + \delta p)\delta_j^i + \Sigma_j^i, \tag{3.23}$$

---

[21]Note that here we are using coordinates $(t, \mathbf{x})$, where t is cosmic time, and $S_0 \equiv S_t$ is the $\mu = 0$ component of S_μ with respect to these coordinates. If one rather uses conformal time η, defined as usual by $dt = a(\eta)d\eta$, then the corresponding $\mu = 0$ component S_η is related to S_t by $S_\eta = aS_t$ and then $V = H_0 a^{-1} S_\eta$. In appendix A of [86], where the perturbation equations for the RT model where first computed, the equations are written in conformal time and the notation S_0 is used for S_η.

where $\bar{\rho}$ and $\bar{p}$ are the unperturbed density and pressure. The matter perturbation variables are therefore $\delta\rho, \delta p, v_i$, and the anisotropic stress tensor Σ^i_j, which is symmetric and traceless, $\Sigma^i_i = 0$. The pressure perturbations can be written as $\delta p = c_s^2 \delta\rho$, where c_s^2 is the speed of sound of the fluid, and we define as usual $\delta \equiv \delta\rho/\bar{\rho}$ and $\theta \equiv \delta^{ij}\partial_i v_j$, with δ_R, θ_R referring to radiation and δ_M, θ_M to matter. We only consider the contribution to $T_{\mu\nu}$ from radiation and non-relativistic matter, so $\Sigma^i_j = 0$. We transform the perturbation equations to Fourier space and we denote comoving momenta by k. We further define

$$\hat{k} = k/(aH), \qquad \hat{\theta} = \theta/(aH). \tag{3.24}$$

We also use

$$\kappa \equiv k/k_{\rm eq}, \tag{3.25}$$

where $k_{\rm eq} = a_{\rm eq}H_{\rm eq}$ is the wavenumber of the mode that enters the horizon at matter-radiation equilibrium. Numerically, $k_{\rm eq} \simeq 0.014\,h_0\,{\rm Mpc}^{-1} \simeq 0.010\,{\rm Mpc}^{-1}$. To illustrate our numerical results, we use as reference values $\kappa = 0.1, 1$ and 5 (or just $\kappa = 0.1$ and 1, when the results for $\kappa = 5$ turn out to be graphically indistinguishable from $\kappa = 1$). The mode with $\kappa = 5$ entered inside the horizon already during RD, while the mode $\kappa = 1$ reentered at matter-radiation equality. In contrast, the mode with $\kappa = 0.1$ was outside the horizon during RD and most of MD, and re-entered at $z \simeq 1.5$.

The full set of equations for the perturbations are given by eqs. (A.6)-(A.10) of [86]. In figure 4 we show the time evolution of the Fourier modes of the Bardeen variable $\Psi_{\bf k}$ for the RT model (minimal and with $\Delta N = 34, 50, 64$), obtained from the numerical integration of these perturbation equations, and we compare with the result in ΛCDM, for $\kappa = 0.1$, $\kappa = 1$ and $\kappa = 5$. We actually plot $k^{3/2}\Psi_k$, whose square gives the variance of the field per unit logarithmic interval of momentum. We see that, up to the present time $x = 0$, the evolution of the scalar perturbations is well-behaved, and very close to that of ΛCDM, and become closer and closer as k increases. This can be understood from the fact that any instability induced by the nonlocal term on the cosmological evolution can only develop on a timescale t such that mt is (much) larger than one. However, we have seen that m is of order H_0, and in fact numerically smaller, with $m \simeq 0.68H_0$ for the minimal RT model and even smaller for large ΔN, see footnote 20. Thus, any instability induced by the nonlocal term can only develop on a timescale larger or equal than to a few times H_0, and therefore in the cosmological future, where these modes could eventually enter a non-linear regime.

3.2.2 Indicators of deviations from GR: $(G_{\rm eff}, \eta)$ and (μ, Σ)

The full set of perturbation equations is needed for implementing the model into a Boltzmann code and comparing its predictions to CMB, BAO and SNe observations, as we will do in section 3.3. For a first qualitative understanding, however, it is convenient to introduce some simpler indicators of deviations from ΛCDM. One such quantity is the effective Newton's constant, which is defined so that the modified Poisson equation for the Fourier modes $\Phi_{\bf k}$ can be rewritten as in GR, with G replaced by $G_{\rm eff}(x, k)$ [recall that here $x \equiv \ln a(t)$ is used to parametrize the time evolution, and should not be confused with a spatial variable],

$$k^2 \Phi_{\bf k}(x) = 4\pi G_{\rm eff}(x; k) a^2 \rho_0 \tag{3.26}$$

$$\times \left[\Omega_R e^{-4x} \left(\delta_{R,{\bf k}}(x) + \frac{4}{\hat{k}^2}\hat{\theta}_{R,{\bf k}}(x) \right) + \Omega_M e^{-3x} \left(\delta_{M,{\bf k}}(x) + \frac{3}{\hat{k}^2}\hat{\theta}_{M,{\bf k}}(x) \right) \right].$$
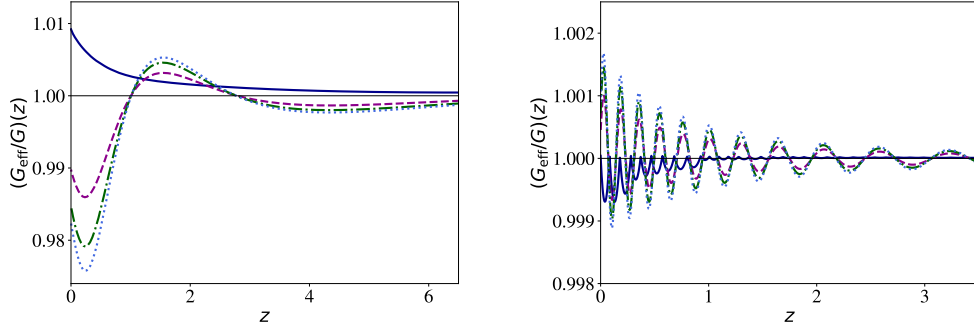
Figure 5. $G_{\rm eff}/G$ as a function of z for fixed κ, for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted), for $\kappa = 0.1$ (left panel) and $\kappa = 1$ (right panel).

Its explicit expression in terms of the perturbed fields can be read from eq. (A.6) of [86],

$$\frac{G_{\rm eff}(x;k)}{G} = 1 + \gamma \frac{\delta U_{\bf k} + h\left(2\Psi_{\bf k}\bar{V}' + \Psi'_{\bf k}\bar{V} - \delta V'_{\bf k}\right) + 3h^2\left(\delta Z_{\bf k} - \frac{1}{2}\delta Z'_{\bf k}\right) + 3h\left(\Psi_{\bf k}\bar{V} - \frac{1}{2}\delta V_{\bf k}\right)}{\Omega_R e^{-4x}\left(\delta_{R,{\bf k}}(x) + \frac{4}{\hat{k}^2}\hat{\theta}_{R,{\bf k}}(x)\right) + \Omega_M e^{-3x}\left(\delta_{M,{\bf k}}(x) + \frac{3}{\hat{k}^2}\hat{\theta}_{M,{\bf k}}(x)\right)}.$$
(3.27)

From this expression one finds that, for sub-horizon modes, i.e. in the limit $\hat{k} \gg 1$, we have [86, 90]

$$\frac{G_{\rm eff}(x;k)}{G} = 1 + \mathcal{O}\left(\frac{1}{\hat{k}^2}\right).$$
(3.28)

As we will see in section 3.4.2, this property, which is not shared by other modified gravity models and in particular by the RR nonlocal model, is crucial, since it allows the RT model to evade limits on the time variation of the (effective) Newton's constant obtained from Lunar Laser Ranging.

Together with $G_{\rm eff}$, a second useful indicator is [91]

$$\eta(x;k) = \frac{\Phi_{\bf k}(x) + \Psi_{\bf k}(x)}{\Phi_{\bf k}(x)},$$
(3.29)

which, in GR, vanishes in the absence of anisotropic stress. Alternatively, two useful quantities are the functions $\mu(x;k)$ [92] and $\Sigma(x;k)$ [91] which are defined through[22]

$$\Psi = [1 + \mu(x;k)]\Psi_{\rm GR} \qquad \Psi - \Phi = [1 + \Sigma(x;k)](\Psi - \Phi)_{\rm GR},$$
(3.30)

where the subscript denotes the same quantities computed in GR, assuming a ΛCDM model with the same value of Ω_M as the modified gravity model. The advantage of this parametrization is that it separates the modifications to the motion of non-relativistic particles, which is described by μ, from the modification to light propagation, which is encoded in Σ. Therefore μ is sensitive to structure formation and Σ is sensitive to lensing.

In figure 5 we show the numerical results for the effective Newton constant as a function of redshift, for the minimal RT model and for the RT model with $\Delta N = 34, 50, 64$, for $\kappa = 0.1$

---

[22]In the literature the quantity that we call $1 + \mu$ is sometimes denoted by μ, and similarly our $1 + \Sigma$ is sometimes denoted by Σ. Our definitions are such that, in GR, $\mu = \Sigma = 0$.
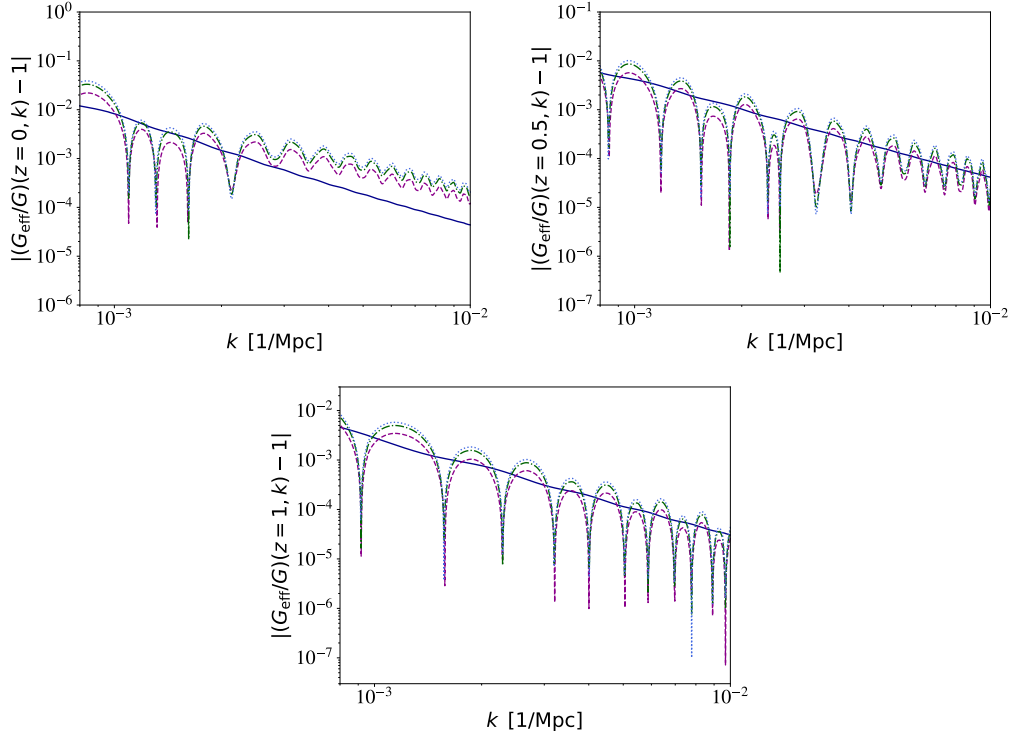
Figure 6. $|(G_{\rm eff}/G) - 1|$ as a function of k for fixed z, for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted), on a logarithmic scale. The three panels refers to $z = 0$ (upper left panel), $z = 0.5$ (upper right) and $z = 1$ (lower panel). The sign of $(G_{\rm eff}/G) - 1$ is such that, close to the vertical axis, $G_{\rm eff}/G > 1$ for the minimal model and $G_{\rm eff}/G < 1$ for the other cases, and the sign changes each time the logarithmic plot has a downward spike.
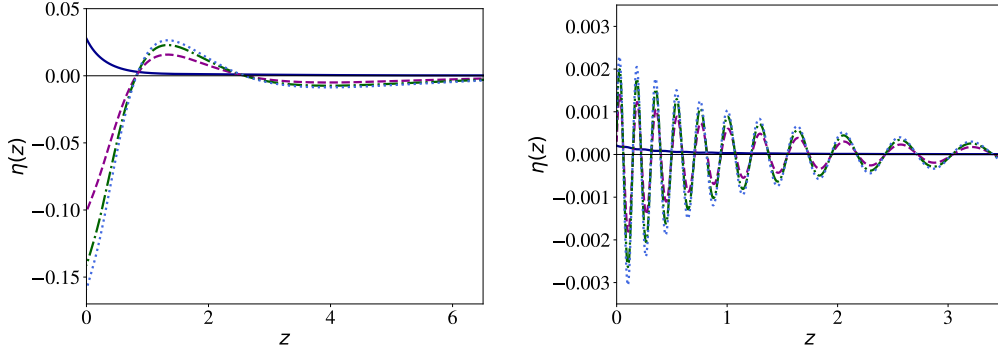

Figure 7. η as a function of z, for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted), for $\kappa = 0.1$ (left panel) and $\kappa = 1$ (right panel).

and 1. We see that, already for $\kappa = 0.1$ (i.e. $k = 0.1 k_{\rm eq} \simeq 0.001\,{\rm Mpc}^{-1}$), $G_{\rm eff}$ differs by G by less than 1%, and, for higher values of k, $G_{\rm eff}$ goes quickly to G, in agreements with eq. (3.28) (for instance, in the plot for $\kappa = 5$, $|G_{\rm eff}/G|$ would always be below 1.001). For these values of k, there are also some oscillations as a function of z and, for given z, the envelop of the oscillations reproduces the $1/k^2$ behavior found analytically in eq. (3.28). Notice that,
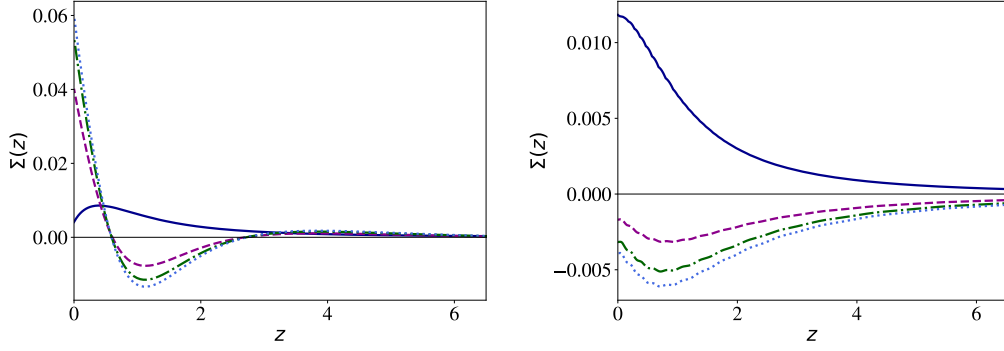
Figure 8. Σ as a function of z, for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted), for $\kappa = 0.1$ (left panel) and $\kappa = 1$ (right panel).

Figure 9. μ as a function of z, for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted), for $\kappa = 0.1$ (left panel) and $\kappa = 1$ (right panel).
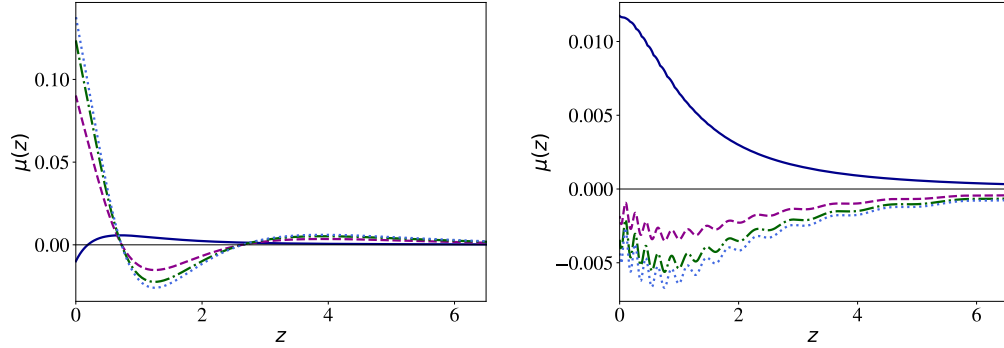
because of eq. (3.28), on small scales $G_{\rm eff}$ reduces to the standard Newton's constant G probed by solar system or by laboratory experiments. However, at typical cosmological scales such as $k \sim k_{\rm eq}$, its value is different, even at $z = 0$. In particular, in the RT models with large ΔN, on these scales $G_{\rm eff} < G$, i.e. gravity is weakened on cosmological scales, while for the minimal RT model it is strengthened. Figure 6 shows, on a logarithmic scale, the dependence of $|(G_{\rm eff}/G) - 1|$ on the wavenumber k, for three different values of the redshift, $z = 0, 0.5$ and 1.

Fig 7 shows η as a function of z, again for $\kappa = 0.1$ and 1, while in figures 8 and 9 we show the same results for the indicators Σ and μ. Notice in particular that both Σ and μ have a rather non-trivial dependence on k for cosmological scales $k \sim k_{\rm eq}$. We see from the plots that, at small redshifts, in the RT model with large ΔN, for $k = 0.1 k_{\rm eq}$ both Σ and μ are positive (with $1 + \Sigma$ higher by about 5% than the ΛCDM value of unity, and μ by about 10% in $z = 0$), while for $k = k_{\rm eq}$ or larger the situation is reversed and Σ and μ become negative at small z.

Another useful derived quantity is the growth rate $f(z, k) \equiv d \log \delta_M / d \ln a$. As is well known, in ΛCDM, for the typical wavenumbers relevant for structure formation, $f(z, k)$ is basically independent of wavenumber k and very well fitted by $f(z) = [\Omega_M(z)]^\gamma$ with γ a constant, numerically close to 0.55. More precisely, writing $f(z) = [\Omega_M(z)]^{\gamma(z)}$, the function
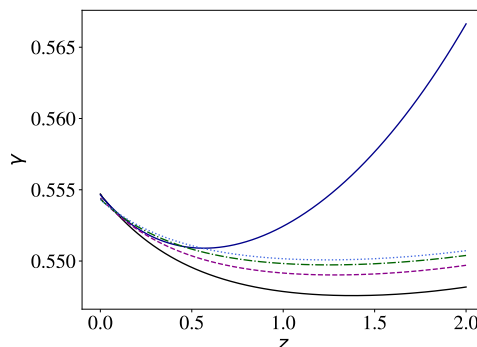
Figure 10. The function $\gamma(z)$ related to the growth rate $f(z)$ by $f(z) = [\Omega_M(z)]^{\gamma(z)}$, for ΛCDM (gray solid line), for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted).

$\gamma(z)$ for ΛCDM is shown as the gray solid line in figure 10, so it is indeed approximately constant and given numerically by $\gamma \simeq 0.55$, within percent level accuracy. We find that the fit $f(z) = [\Omega_M(z)]^{\gamma(z)}$ also holds for the RT model, again with a function $\gamma(z)$ independent of the wavenumber k. The corresponding functions $\gamma(z)$ are shown in figure 10 for the RT model, minimal and with large ΔN. We see that, for large values of ΔN, $\gamma(z)$ is indeed independent of z within percent level accuracy, just as in ΛCDM, and again is given numerically by $\gamma \simeq 0.55$.[23] For the minimal RT model the variation of $\gamma(z)$ with redshift is somewhat larger, but still it stays between $0.55 - 0.56$ up to $z = 2$. Notice that the growth index γ is a useful quantity only as long as we are in the epoch where DE is still important. When we are deep into MD, $\Omega_M(z) \to 1$, and $[\Omega_M(z)]^\gamma \to 1$ independently of γ.

Two main conclusions emerge from this study of the cosmological perturbations of the RT model in the scalar sector. First, they are well-behaved. This is already a rather non-trivial result. Several modified gravity models have indeed been ruled out by the presence of instabilities in their perturbations. This was for instance the case for the DPG model [52], which opened the way to the study of IR modifications of GR and has a self-accelerated solution [93, 94] but had a ghost-like instability on the self-accelerated branch [95–99]. Massive gravity [100–102] has difficulties already in obtaining a viable background FRW evolution [103], while in bigravity [104] a background FRW solutions exist, but, in a branch of solutions that has a dynamical dark energy, the cosmological perturbations have instabilities in both the scalar and tensor sectors [105–111] (see [112] for a recent comprehensive review of modifications of GR at the cosmological scale). Thus, already the fact of producing quite naturally a viable cosmological background evolution with self-acceleration, and stable scalar perturbations, is a non-trivial results.

The second conclusion that emerges from this study is that, both in the background evolution and in the scalar perturbations, the RT model is very close to ΛCDM, with deviations of at most a few percent, for all ΔN. This already indicates that the model is a good candidate for fitting well the current cosmological observations. In the next section we will confirm this conclusion by comparing the RT model with ΛCDM from the point of view of the quality of the fit to the cosmological observations, and we will perform Bayesian parameter estimation for the values of the cosmological parameters.

---

[23]Of course the growth index, traditionally denoted by γ, should not be confused with the parameter γ of the RT model, defined in eq. (3.6)!

3.3 Comparison with cosmological observations

We now perform a detailed comparison with cosmological observations, using the most recent cosmological datasets in order to update the results presented in [113, 114] for the minimal RT model, and in [115] for the RT model with large ΔN. As in these previous works, we implement the perturbations of the RT model computed in [86] into the CLASS cosmological Boltzmann code [116] (v2.7), that we have modified so to describe the background evolution and scalar perturbations of the RT model. Our code has been tested against other Einstein-Boltzmann solvers in [117], and the most recent version is publicly available on GitHub [118] (evolved from [119]).

3.3.1 Datasets and methodology

For ΛCDM, the *Planck* baseline analysis uses six independent cosmological parameters: the Hubble parameter today $H_0 = 100h_0 \, \mathrm{km \, s^{-1} Mpc^{-1}}$, the physical baryon and cold dark matter density fractions today $\omega_b = \Omega_b h_0^2$ and $\omega_c = \Omega_c h_0^2$, respectively, the amplitude A_s and tilt n_s of the primordial scalar perturbations, and the reionization optical depth τ_{re}. Note that, assuming flatness, the energy fraction Ω_Λ associated to a cosmological constant is a derived parameter, fixed by the flatness condition. In the RT model we have a mass scale m [or, equivalently, the dimensionless parameter γ, eq. (3.6)] which replaces the cosmological constant, and again can be taken as a derived parameter, fixed by the flatness condition. Thus, for the RT model, we can take the same six independent cosmological parameters, as in ΛCDM.

An important extension, however, is provided by the sum of neutrino masses, $\sum_\nu m_\nu$. As discussed in [120], their inclusion can a priori be important when comparing a modified gravity model to ΛCDM. Oscillation experiments give a lower limit $\sum_\nu m_\nu \gtrsim 0.06 \, \mathrm{eV}$ [121] (assuming a normal mass hierarchy dominated by the heaviest neutrino mass eigenstate). In the *Planck* baseline analysis the sum of neutrino masses is kept fixed to this minimum allowed value. As discussed in the *Planck* papers [122, 123], there is actually no compelling theoretical reason for this choice, and there are other possibilities, including a degenerate hierarchy with $\sum_\nu m_\nu \gtrsim 0.1 \, \mathrm{eV}$. The choice of fixing the sum of neutrino masses to the minimum allowed values is justified by the fact that, in ΛCDM, letting the sum of neutrino masses as a free parameter, one finds that its marginalized posterior is peaked in zero, and if we let it vary with the prior $\sum_\nu m_\nu \geq 0.06 \, \mathrm{eV}$ the data drive $\sum_\nu m_\nu$ back to the prior (see figure 34 of [123]). In contrast, in a modified gravity model, the posterior for $\sum_\nu m_\nu$ could be peaked at a value higher than the lower bound $0.06 \, \mathrm{eV}$.[24] A uniform comparison of a modified gravity model with ΛCDM therefore requires to let $\sum_\nu m_\nu$ as a free parameter in both models, as we have done in [23] and as we will do below. We will denote by $\nu\Lambda$CDM the ΛCDM model in which $\sum_\nu m_\nu$ is added to the list of free parameters.

In summary, we will perform Bayesian parameter estimation for both $\nu\Lambda$CDM and the RT model (minimal, and with $\Delta N = 34, 50, 64$), and we will compare the quality of their fits to the datasets discussed below, using, as free parameters,

$$\theta = \{H_0, \omega_b, \omega_c, A_s, n_s, \tau_{\mathrm{re}}, \textstyle\sum_\nu m_\nu\} \,. \tag{3.31}$$

For CMB, SNe and BAO we use the following likelihoods:

---

[24]This is indeed what happens in the RR nonlocal model [23, 120], and we will find that this also happens for the minimal RT model.

- For CMB we use the *Planck* 2018 data release, using the low-ℓ temperature-only likelihood, the low-ℓ EE likelihood, and the high-ℓ temperature and polarization PLIK likelihood described in ref. [124], as well as the lensing likelihood based on temperature+polarization map-based lensing reconstruction [125]. This provides a significant update of our previous studies based on the *Planck* 2015 data release.

- For type Ia supernovae we use the likelihood of the Pantheon type Ia supernova sample [126], which includes data from the Pan-STARRS1 (PS1) Medium Deep Survey. This updates our previous study based on the JLA [127] dataset.

- For BAO we still use the likelihoods of the BAO detection of the 6dF Galaxy Survey [128] and the BAO scale measurement of SDSS DR7 Main Galaxy Sample [129], and we update the SDSS data using the power spectrum of BAO from the Data Release 12 [130].

For the RT model the initial conditions of the perturbations of the auxiliary fields δU and δS_μ are set to zero. As we have already shown in [23], taking different initial conditions, of the order of the metric perturbations (which is their natural scale, since, as discussed in section 2.3.1, the initial conditions on the auxiliary fields are in principle fixed by the initial conditions on the metric perturbations) has a totally negligible effect.

After having determined in this way the mean values of the parameters of the models (ΛCDM and RT, minimal and with various ΔN), we will use these values to compare the models with further datasets, namely measurements of $H(z)$ ("cosmic chronometers") and $f\sigma_8$ data.[25]

3.3.2 Comparison with CMB, BAO, SNe, cosmic chronometers and $f\sigma_8$

Fit to CMB+BAO+SNe and Bayesian parameter estimation. Table 2 shows the results for the Bayesian parameter estimation and the resulting χ^2 for $\nu\Lambda$CDM and the RT model (minimal, and with $\Delta N = 34, 50, 64$), using the combined CMB+SNe+BAO data. Beside the values of the seven fundamental independent parameters given in (3.31), we also give some useful derived parameters, namely Ω_M, the reionization redshift $z_{\rm re}$, and the amplitude of matter density fluctuations in spheres of radius $8h_0^{-1}$ Mpc, σ_8. In the last line we

---

[25]Some technical details on our MCMC. We use the statistical framework Cobaya 2.0.2 (https://github.com/CobayaSampler/cobaya, developed by Jesus Torrado and Antony Lewis) to let Markov chains sample the posterior distribution for the cosmological parameters. Cobaya uses the sampler developed for CosmoMC [131, 132] tailored for parameter spaces with a speed hierarchy (it also implements the "fast dragging" procedure described in [133]). We determine the best-fit cosmological parameters as follows. For each model, we select from its Markov Chain samples the N samples that are closest to the highest-posterior sample and fit a generic quadratic function using least squares. In practice, we choose $N = 5d^2$ where $d = 28$ is the total number of parameters including the fiducial ones and the factor 5 is chosen as a compromise between locality and numerical stability. The norm used to determine the closest samples is Euclidean after normalization of the sample coordinates by their standard deviations as estimated from all samples of the chain. We then identify the convex subspace of the quadratic fit using an eigen-decomposition of the Hessian. Finally, we minimize the quadratic fit within the convex subspace under the constraint that some of the parameters must be positive (for example the neutrino mass). Given the Markov Chain samples, this procedure gives a best-fit candidate within seconds of runtime. The posterior is then evaluted at the candidate point predicted by this procedure. Typically, the prediction and the actual evaluation are close. The values for χ^2 given here are always corresponding to the true evaluation at the predicted minimum. We consistently get better results following this method than using Cobaya's BOBYQA [134–136] minimizer, which takes into account the previous samples only via their covariance matrix. A python notebook is available at https://github.com/AndreasFinke/quadfit.

Parameter	$\nu\Lambda$CDM	RT, minimal	RT, $\Delta N = 34$	RT, $\Delta N = 50$	RT, $\Delta N = 64$
H_0	67.89 ± 0.47	$68.74^{+0.59}_{-0.51}$	67.95 ± 0.48	67.90 ± 0.47	67.88 ± 0.48
$\sum_\nu m_\nu$ [eV]	< 0.057 (at 1σ)	$0.071^{+0.024}_{-0.066}$	< 0.048 (at 1σ)	< 0.044 (at 1σ)	< 0.041 (at 1σ)
ω_c	0.1193 ± 0.0009	0.1120 ± 0.0009	0.1191 ± 0.0009	0.1190 ± 0.0009	0.1189 ± 0.0009
$100\omega_b$	2.242 ± 0.013	2.237 ± 0.013	2.243 ± 0.013	2.244 ± 0.013	2.244 ± 0.013
$\ln(10^{10}A_s)$	3.045 ± 0.014	3.043 ± 0.014	3.047 ± 0.014	$3.048^{+0.013}_{-0.015}$	3.049 ± 0.014
n_s	0.9665 ± 0.0036	0.9649 ± 0.0036	0.9670 ± 0.0036	0.9673 ± 0.0035	0.9672 ± 0.0035
$\tau_{\rm re}$	0.0555 ± 0.0072	0.0537 ± 0.0072	0.0565 ± 0.0073	$0.0572^{+0.0065}_{-0.0075}$	0.0575 ± 0.0071
Ω_M	0.3085 ± 0.0060	$0.3029^{+0.0061}_{-0.0070}$	0.3075 ± 0.0061	0.3076 ± 0.0060	0.3076 ± 0.0060
$z_{\rm re}$	7.76 ± 0.72	7.60 ± 0.73	7.86 ± 0.72	7.93 ± 0.70	7.96 ± 0.70
σ_8	$0.8164^{+0.0097}_{-0.0068}$	$0.823^{+0.0130}_{-0.0087}$	$0.8141^{+0.0089}_{-0.0067}$	$0.8134^{+0.0088}_{-0.0064}$	$0.8129^{+0.0084}_{-0.0066}$
$\Delta\chi^2$	0	1.30	-0.48	-0.20	-0.00

Table 2. Mean values (with 1σ errors) of the parameters for $\nu\Lambda$CDM and the RT model (minimal, and with $\Delta N = 34$, 50, 64), using CMB, BAO and SNe. H_0 is in units of $\rm km\,s^{-1}\,Mpc^{-1}$. The last line gives the difference in the χ^2 of each given model with respect to $\nu\Lambda$CDM. The RT model with $\Delta N = 34$ or with $\Delta N = 50$ fits the data slightly better than $\nu\Lambda$CDM, but the difference is not statistically significant.

show the differences in χ^2, with respect to the value for $\nu\Lambda$CDM. We recall that, for models with the same number of free parameters, as $\nu\Lambda$CDM and the RT models, the conventional interpretation is that a difference $|\Delta\chi^2| \leq 2$ implies statistical equivalence between the two models, while $2 \lesssim |\Delta\chi^2| \lesssim 6$ suggests "weak evidence" in favor of the model with lower χ^2, and $|\Delta\chi^2| \gtrsim 6$ indicates "strong evidence" in favor of the model with lower χ^2. Thus, all models considered fit the data at a statistically equivalent level.[26]

The result of Bayesian parameter estimation shows that all models with large ΔN give predictions extremely close to those of $\nu\Lambda$CDM, consistently with the analysis of the previous sections, that showed that these models are very close to ΛCDM both in the background evolution and in the cosmological perturbations. The minimal RT model differs a bit more, and in particular predicts a slightly higher value of H_0, which in any case is not enough to significantly relieve the tension with the local H_0 measurement [137, 138]. Indeed, as discussed in [139, 140], it might not be possible to solve the H_0 tension, together with other potential tensions within ΛCDM, with a modification of only the late-Universe dynamics (as in our nonlocal model). The other difference of the minimal RT model is that it predicts a non-zero value for the sum of the neutrino masses, while all other models considered only give an upper bound.

Figures 11 and 12 show the two-dimensional likelihoods for (Ω_M, σ_8), (Ω_M, H_0) and $(\sum_\nu m_\nu, H_0)$. The pattern that emerges, from this and similar plots, is that the RT model with large values of ΔN is extremely close to ΛCDM, as we already saw from table 2, while the minimal RT model has some more significant differences, such as a slightly higher value of H_0 (although, as mentioned above, not enough to significantly decrease the tension with local measurements), of σ_8, and of the sum of neutrino masses.

---

[26]Note that ΔN is not a free parameter varied so to minimize the χ^2. Rather, we have used a very limited sample of values of ΔN, chosen a priori on the basis of the fact that, according to the relation (3.17), they correspond to significant choices for the inflationary scale $M_{\rm infl}$, see the discussion below eq. (3.17).
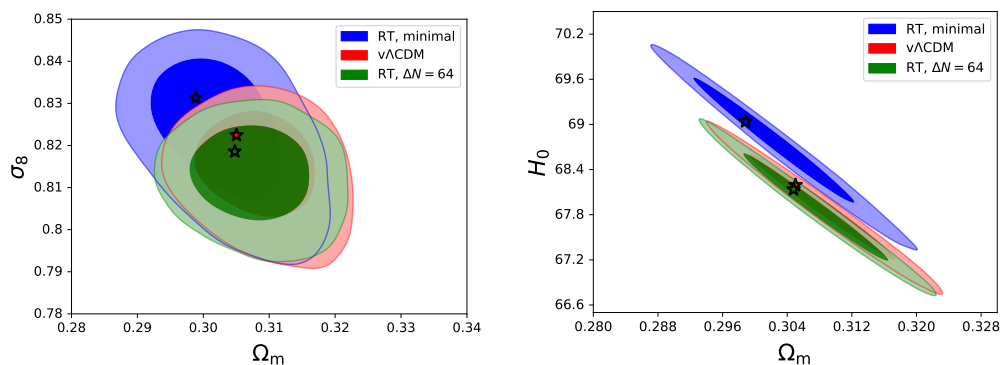
Figure 11. Left panel: the two-dimensional likelihood in the (Ω_M, σ_8) plane for $\nu\Lambda$CDM (red), the minimal RT model (blue) and the RT model with $\Delta N = 64$ (green). The stars are the best-fit values of the parameters (note that the values reported in table 2 are rather the mean values). Right panel: the same for (Ω_M, H_0).
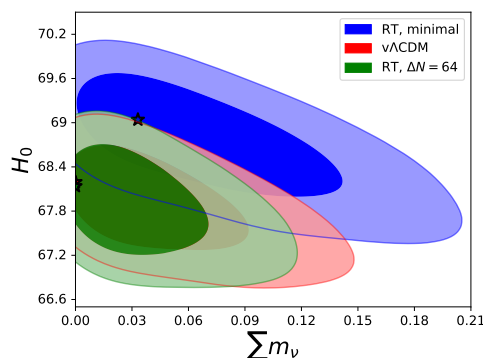

Figure 12. As in figure 11, for $(\sum_\nu m_\nu, H_0)$.

	$\nu\Lambda$CDM	RT, minimal	$\Delta N = 34$	$\Delta N = 50$	$\Delta N = 64$
$\Delta\chi^2$	0	-1.13	0.22	0.42	0.57

Table 3. Values of $\Delta\chi^2$, with respect to $\nu\Lambda$CDM, for the RT model, minimal and with various values of ΔN, from the fit to a compilation of measurements of $H(z)$.

	$\nu\Lambda$CDM	RT, minimal	$\Delta N = 34$	$\Delta N = 50$	$\Delta N = 64$
$\Delta\chi^2$	0	1.41	-0.05	-0.18	-0.28

Table 4. Values of $\Delta\chi^2$, with respect to $\nu\Lambda$CDM, for the RT model, minimal and with various values of ΔN, from the fit to a compilation of measurements of $f\sigma_8$.

Cosmic chronometers. Another useful observational test is provided by measurements of $H(z)$ at different redshifts ("cosmic chronometers" [141]). We use a compilation of 36 measurements of $H(z)$ between $z = 0.07$ and $z = 2.34$, given in table I of [142]. Using the respective prediction for $H(z)$ in ΛCDM and in the RT models (with the respective mean values of Ω_M and H_0 from table 2, obtained from the MCMC comparison to CMB+BAO+SNe) to fit these $H(z)$ measurements, we find the difference in χ^2, with respect to $\nu\Lambda$CDM, shown in table 3. The corresponding reduced χ^2, all of order $0.63 - 0.64$, show that, by themselves, all the models fits these data well.
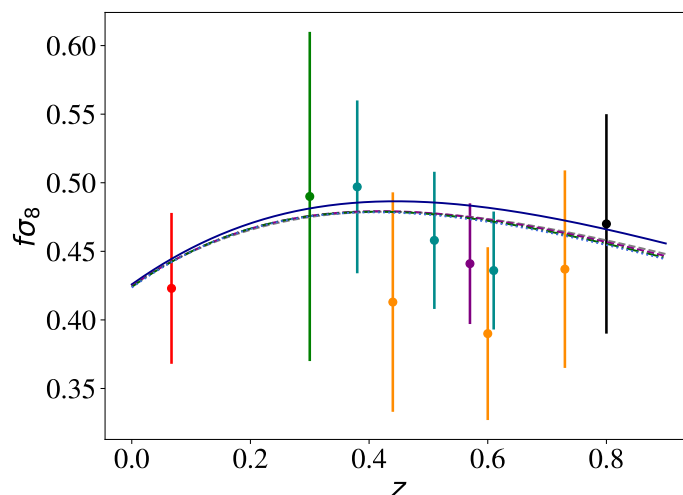
Figure 13. A collection of measurements of $f\sigma_8$ and the corresponding predictions of $\nu\Lambda$CDM and of the RT model, minimal and with $\Delta N = 34, 50$ and 64. The curve for the minimal RT model is the upper one, while all others are almost indistinguishable on this scale. The data points are from 6dF GRS [144] (red), SDSS LRG [145] (green), BOSS CMASS [146] (purple), WiggleZ [147] (orange), VIPERS [148] (black) and BOSS DR12 [130] (cyan).

Structure formation and $f\sigma_8$ data. The properties of the models with respect to structure formation are already partly tested by the inclusion of BAO in our MCMC analysis. We further compare the models to a set of measurements of $f\sigma_8$, using the datapoints that we already used in [23].[27]

Figure 13 shows the data and the predictions of ΛCDM and of the RT model, minimal and with $\Delta N = 34, \Delta N = 50, \Delta N = 64$, obtained using for each model the respective mean values of Ω_M and H_0 from table 2. The corresponding differences of χ^2, with respect to the value in $\nu\Lambda$CDM, are given in table 4. We see that, once again, the differences between ΛCDM and the RT model with various ΔN are not statistically significant. From the plots of $G_{\rm eff}$ in figure 5 we see that at low k (upper left panel) the minimal RT model predicts $G_{\rm eff}/G > 1$, while the RT models with large ΔN predict $G_{\rm eff}/G < 1$. The data favor a weakening of gravity at these scales, so the RT models with large ΔN are slightly preferred with respect to ΛCDM, and the minimal RT model is slightly disfavored, but in all cases at a statistically insignificant level.

Finally, figure 14 shows the relative difference in the linear power spectrum of the RT models with respect to ΛCDM (each one computed using their respective mean values of the cosmological parameters) as a function of k, for $z = 0$ (left panel), and as as a function of z, for the mode with $k = 0.1/{\rm Mpc}$ (right panel).

The conclusion of this analysis is that, on the one hand, the RT model, for all values of ΔN, is very close to ΛCDM at the level of background evolution and scalar perturbations, and

---

[27]Actually, many more measurement exists: ref. [143] provides a compilation of 63 measurement of $f\sigma_8$ from 2006 to 2018. However, as stressed in [143], many of these datapoints are correlated, due to overlap in the galaxy samples used, and no covariance matrix is available for the full dataset, nor for most of its subsets. Furthermore, one must also take care of the fact that different datapoints have been obtained with different fiducial cosmologies, and that survey systematic may vary with time of publication and lead to inhomogeneities in the data. Therefore the use of the full dataset, without the appropriate covariance matrix and corrections, would lead to results of dubious interpretation.
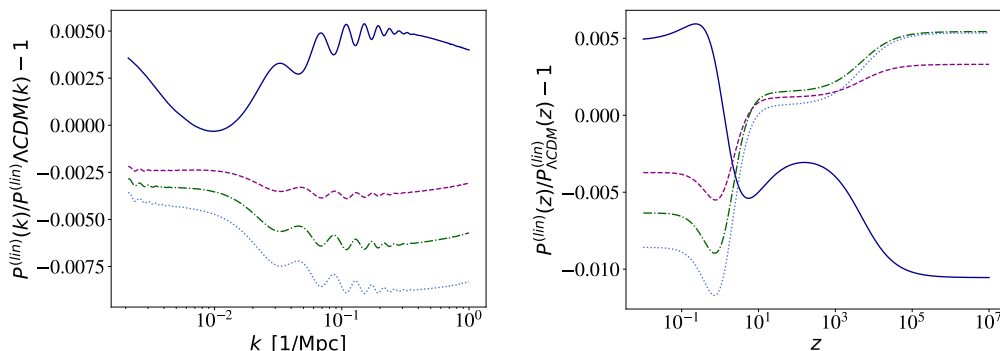
Figure 14. Relative difference of total linear matter power spectrum, with respect to best-fit ΛCDM, for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted). Left panel: as a function of k, at $z = 0$. Right panel: as a function of redshift, for the mode with $k = 0.1/$Mpc.

fits the observations at the same level as ΛCDM. On the other hand, the deviations, which for both the background and scalar perturbations are typically at the percent or sub-percent level, could in principle be within reach for future missions. For instance, assuming that the function $\mu(a, k)$ that characterizes deviations from the Poisson law is scale independent and parametrizing its dependence on the scale factor as $\mu(a) = \mu_s a^s$, a future survey such as EUCLID [149], for fixed cosmological parameters, is expected to measure μ_s with an error $\Delta\mu_s = 0.0046$ for $s = 1$ and $\Delta\mu_s = 0.014$ for $s = 3$ [150]. The RT model has indeed been selected by the Dark Energy Science Collaboration (DESC) of the Large Synoptic Survey Telescope (LSST), among a few modified gravity models, for further studies and development of dedicated pipelines [151].

3.4 Recovery of GR at short scales

3.4.1 Solar system constraints and absence of vDVZ discontinuity

Any cosmological model that modifies GR on cosmological scale must also be able to reproduce the successes of GR at much smaller scales, such as the solar system and laboratory scales. In theories that introduces extra fields, as in scalar-tensor theories, or extra polarization of the gravitons, as in massive gravity, this is highly non-trivial. The linearized theory does not reduce to GR, and screening mechanisms involving the non-linearities of the theory are needed. In the RT model, in contrast, the situation is much simpler. Already at the linear level the theory reduces smoothly to GR, and there is no discontinuity such as the vDVZ discontinuity of massive gravity. These issues have been discussed at length in [20, 33] and here we summarize these results, for completeness.

GR limit in the linearization over Minkowski space. Let us consider first the GR limit for the RT model linearized over flat space. In this case eq. (2.45) reduces to eq. (2.38). In order to compute the matter-matter interaction induced by this coupling of $T_{\mu\nu}$ with $h_{\mu\nu}$ we proceed as follows [20]. We use the gauge invariance of the linearized theory to fix the De Donder gauge $\partial^\mu[h_{\mu\nu} - (1/2)h\eta_{\mu\nu}] = 0$. Going in momentum space, the resulting equation can be solved for the Fourier transform $\tilde{h}_{\mu\nu}(k)$, obtaining

$$\tilde{h}_{\mu\nu}(k) = \frac{16\pi G}{k^2}\left[\tilde{T}_{\mu\nu}(k) - \frac{\eta_{\mu\nu}k^2}{2(k^2 - m^2)}\tilde{T}(k) + \frac{m^2}{3(k^2 - m^2)}\left(\eta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2}\right)\tilde{T}(k)\right], \quad (3.32)$$

where $T = \eta^{\mu\nu} T_{\mu\nu}$. Plugging this result into the linearized interaction term[28]

$$S_{\text{int}} = \frac{1}{2} \int d^4x \, h_{\mu\nu} T^{\mu\nu} \,, \tag{3.33}$$

and using $k^\mu \tilde{T}_{\mu\nu}(k) = 0$ we get

$$S_{\text{int}} = 8\pi G \int \frac{d^4k}{(2\pi)^4} \, \tilde{T}_{\mu\nu}(-k) \Delta^{\mu\nu\rho\sigma}(k) \tilde{T}_{\rho\sigma}(k) \,, \tag{3.34}$$

where

$$\Delta^{\mu\nu\rho\sigma}(k) = \frac{1}{2k^2} \left(\eta^{\mu\rho}\eta^{\nu\sigma} + \eta^{\mu\sigma}\eta^{\nu\rho} - \eta^{\mu\nu}\eta^{\rho\sigma} \right) + \frac{1}{6} \left[\frac{1}{k^2} - \frac{1}{k^2 - m^2} \right] \eta^{\mu\nu}\eta^{\rho\sigma} \,. \tag{3.35}$$

The term in the first line is the usual GR result due to the exchange of the helicities ± 2 of a massless graviton. The term in brackets vanishes for $m \to 0$. Therefore the RT model has no vDVZ discontinuity, and reduces smoothly to GR as $m \to 0$. In the regime where a linearization over flat space is adequate, for modes with $|k^2| \gg m^2$ the predictions of the RT model differ from the predictions of GR by a factor $1 + \mathcal{O}(m^2/k^2)$. We have seen that the comparison with cosmological observations fixes $m \sim H_0$ (or even smaller for large ΔN, see footnote 20). For $|k| = (1\,\text{a.u.})^{-1}$ (as appropriate to solar system experiments), $m^2/k^2 \sim (1\,\text{a.u.}/H_0^{-1})^2 \sim 10^{-30}$, and the predictions of the RT model are indistinguishable from that of GR.

The absence of vDVZ discontinuity can also be understood observing that the term in bracket in eq. (3.35) induces a matter-matter interaction

$$8\pi G \int \frac{d^4k}{(2\pi)^4} \frac{1}{6} \tilde{T}(-k) \left[\frac{1}{k^2} - \frac{1}{k^2 - m^2} \right] \tilde{T}(k) \,. \tag{3.36}$$

Comparing with eq. (2.33) one realizes that the two terms in brackets corresponds to the exchange of the helicity zero component of $h_{\mu\nu}^{\text{TT}}$ and of the trace mode s. In GR, where s is massless and both $h_{\mu\nu}^{\text{TT}}$ and s appear with a $\Box$ factor in the quadratic lagrangian [see eq. (2.24)] these two terms cancel exactly, while here the cancellation is only partial but is recovered for $m \to 0$. Notice that both the helicity zero component of $h_{\mu\nu}^{\text{TT}}$ and s are non-propagating degrees of freedom in GR and remain non-propagating in the RT model. We have indeed seen in eqs. (2.57)–(2.59) that, in the RR or RT models linearized over flat space, the only radiative degree of freedom of the metric are still given by the helicity ± 2 modes described by H_{ij}^{TT} and s remains non-radiative, see eq. (2.61). Exactly as in GR, the negative sign in front of the $1/(k^2 - m^2)$ term in eq. (3.36), which would correspond to a ghost if it were due to a propagating particle, is therefore innocuous from the point of view of quantum vacuum stability. The helicity zero component of $h_{\mu\nu}^{\text{TT}}$ and s are not associated to creation/annihilation operators and cannot appear on the external lines of a Feynman diagram.[29]

---

[28]In [20] the overall factor $1/2$ was missed in S_{int}.

[29]Notice that, in GR, vacuum stability it is not related to the fact that the contribution of s to the interaction (3.34) is canceled by the contribution of the helicity zero component of $h_{\mu\nu}^{\text{TT}}$. This is a cancelation that, in the language of Feynman graphs, takes place only in the internal lines. However, if s were a propagating degree of freedom it would also appear in external lines, where it would induce vacuum decay into negative-energy ghost states plus positive-energy particles (and this, of course, cannot be canceled by graphs with the helicity zero mode of $h_{\mu\nu}^{\text{TT}}$ on external lines, since these contribute to different S-matrix elements). The crucial point for vacuum stability in GR is rather that s is non-propagating, so it is not associated to creation/annihilation operators and cannot appear on the external lines (just as A_0 in electrodynamics).

GR limit for the Schwarzschild solution. After having checked the recovery of the GR limit in Minkowski space, let us consider the GR limit for the Schwarzschild solution, by studying the static spherically symmetric solution of the RT model. A typical issue of massive gravity theories is that they become non-linear when r is smaller than a distance, the Vainshtein radius, which is parametrically larger than the Schwarzschild radius r_S of the source; e.g. $r_V = (GM/m^4)^{1/5}$ in the theory defined by adding a Fierz-Pauli mass term to the Einstein-Hilbert action [152, 153], and $r_V = (GM/m^2)^{1/3}$ [154] in the dRGT theory [100, 101]. For $m = \mathcal{O}(H_0)$ and $M = M_\odot$, we have $(GM/m^2)^{1/3} \sim 100$ pc. Since linearized theory only holds for $r > r_V$, in massive gravity in the whole range of distances probed by solar system and laboratory experiments the linearized expansion is not valid, and one must show that a Vainshtein mechanism is at work, i.e. that the inclusion of classical non-linearities restore the continuity with GR at $r \ll r_V$. Explicit examples of this type have indeed been found for the dRGT theory [155, 156].

For the RT model, however, the situation is much simpler, and the limit $m \to 0$ of the Schwarzschild solution is smooth. The Schwarzschild solution in the RT model has been worked out in [33]. In the limit $r \gg r_S$, the result for the metric is

$$ds^2 = -A(r)dt^2 + B(r)dr^2 + r^2(d\theta^2 + \sin^2\theta \, d\phi^2)\,, \tag{3.37}$$

where

$$A(r) = 1 - \frac{r_S}{r}\left[1 + \frac{1}{3}(1 - \cos mr)\right]\,, \tag{3.38}$$

$$B(r) = 1 + \frac{r_S}{r}\left[1 - \frac{1}{3}(1 - \cos mr) + \frac{1}{3}mr\sin mr\right]\,, \tag{3.39}$$

In the limit $mr \ll 1$ (but still $r \gg r_S$), eqs. (3.38) and (3.39) give

$$A(r) \simeq 1 - \frac{r_S}{r}\left(1 + \frac{m^2 r^2}{6}\right)\,, \tag{3.40}$$

and (to first order in r_S/r) $B(r) = 1/A(r)$.[30] For comparison, in massive gravity the analogous computation gives [53, 152]

$$A(r) = 1 - \frac{4}{3}\frac{r_S}{r}\left(1 - \frac{r_S}{12m^4 r^5}\right)\,. \tag{3.41}$$

The factor $4/3$ in front of r_S/r is due to the extra contribution coming from the exchange of the helicity-0 graviton, and gives rise to the vDVZ discontiuity. In contrast, no vDVZ discontinuity is present in eq. (3.40). Furthermore, in eq. (3.41) the correction blows up as r decreases, and for $r \sim r_V = (GM/m^4)^{1/5}$ it becomes of the order of the leading term, signaling the breakdown of the linearized approximation. In eq. (3.40), in contrast, the correction becomes smaller and smaller as r decreases, and perturbation theory is valid at all scales $r \ll m^{-1}$, until we arrive at $r \simeq r_S$, where eventually also GR becomes non-linear.

In conclusion, in the RT model (as well as in the RR model, where the analysis is very similar), in static situations GR is smoothly recovered, with correction $O(m^2 r^2)$. Given that

---

[30]The solution for the auxiliary field $U = -\Box^{-1}R$ is given simply by $U(r) = (r_S/r)\cos mr$. For the auxiliary field $S_\mu(x)$, in spherical coordinates only the component $S_r(r)$ is non-vanishing, and furthermore depends only on r. It is convenient to define $V(r)$ from $S_r(r) = B^{1/2}(r)rV(r)$. The (not very illuminating) solution for $V(r)$ is given in eq. (3.8) of [33], and reduces to $V(r) \simeq -r_S/(2r)$ for $r \gg r_S$.

m is of order H_0, these corrections are utterly negligible for all r of order of solar system scale or smaller; e.g. $m^2 r^2 \sim 10^{-30}$ for r of order of the Earth-Sun distance. Even on galactic scales these corrections to GR are totally irrelevant, with $m^2 r^2 \sim 10^{-17}$ for $r = 10$ kpc.

3.4.2 Limits on time variation of $G_{\rm eff}$ from Lunar Laser Ranging

The above results show that, in a static situation, the RT and RR models recover all successes of GR at short scales. As was pointed out in [157], this is not yet sufficient to guarantee that these models are viable at solar system scales. Another crucial test comes from the limit on the time variation of Newton's constant from Lunar Laser Ranging (LLR). The current observational result is $\dot{G}/G = (7.1 \pm 7.6) \times 10^{-14}\,{\rm yr}^{-1}$ [158]. This measurement is so accurate that, even if performed at the Earth-Moon scale over the last few decades, it provides significant constraints on cosmological models. Indeed, if we rewrite this limit in terms of the Hubble parameter today, using $H_0 \simeq h_0 \times (9.777752\,{\rm Gyr})^{-1}$, we get

$$\frac{\dot{G}}{G} = (0.99 \pm 1.06) \times 10^{-3} \left(\frac{0.7}{h_0}\right) H_0 \,. \tag{3.42}$$

Quite generally, in modified gravity models Newton's constant becomes time dependent on cosmological scale. The scale for the time variation today is given by H_0, so on cosmological scales one typically finds $\dot{G}/G \simeq H_0$. If, in a given modified gravity model, this result holds also down to the scale of the solar system and of the Earth-Moon system, then the bound (3.42) is violated and the model is ruled out.

In the case of the RT model, however, we have seen in eq. (3.28) that $G_{\rm eff}$ reduces to G at small scales. Therefore, it has no time dependence and the RT model satisfies trivially the LLR limit. The situation is different for the RR model (and for other modified gravity models, see appendix A). Indeed, in the RR model, for sub-horizon modes, one finds [86, 157]

$$\frac{G_{\rm eff}(t)}{G} = \left[1 - \frac{1}{3}m^2 \bar{S}(t)\right]^{-1} \left[1 + \mathcal{O}\left(\frac{1}{\hat{k}^2}\right)\right] \,, \tag{3.43}$$

where $\bar{S}(t)$ is the background cosmological solution for the auxiliary field S. This dependence on S can be traced to the term $2SG_{\mu\nu}$ in $K_{\mu\nu}$, see eq. (2.49). If one plugs here the solution for $\bar{S}(t)$ corresponding to the FRW background, one finds that $G_{\rm eff}(t)/G$ is of order H_0, and the bound (3.42) is violated. In this case one cannot appeal to non-linear screening mechanisms, since we have seen that the RR model (just as the RT model) has a smooth limit $m \to 0$, so the linearized expansion can be trusted.

Of course, the FRW metric has no direct relevance for the Earth-Moon system. The latter, just as the solar system, does not expand with the Hubble flow. However, the point is that a scalar field, such as S, that evolves on a background that interpolates between the Schwarzschild solution at short scales and the FRW solution at large distances, in general inherits a time dependence on small scales from the matching with the solution at large distances. As an extreme example, in GR one can consider the Einstein-Straus space-time, in which, inside a sphere of radius r_0, the metric is taken to be exactly the static Schwarzschild metric generated by the mass M, while in the exterior it is given by a FRW solution with energy density ρ (see e.g. [159, 160] for review). The two metrics are then matched by requiring that the induced metric on the boundary surface Σ agrees on the two sides. This fixes the matching radius r_0, that, with respect to the Schwarzschild coordinates of the interior, turns out to be given by $M = (4/3)\pi r_0^3 \rho$, where ρ is the energy density in FRW. In

this case the solution for the metric is exactly static in the interior region, so it describes a limiting case in which the cosmological expansion in the inner region is perfectly screened. Nevertheless, if one studies the propagation of a scalar field obeying the equation $\Box\phi = 0$ in this metric, one finds that the solution for the field in the inner region is time dependent [160]. This is due to the fact that we must impose a matching condition for the scalar field at the surface Σ, and in this way the field inherits a time dependence even in the inner region.

For the RR model, a detailed analysis of the solution for the scalar field S in a background that interpolates between the static solution at short distances and FRW at large distances has been performed in [161]. A useful way of studying the problem is to follow the time evolution of the auxiliary fields U and S of the RR model, starting before the epoch of structure formation. At that time the FRW metric holds everywhere, and U and S evolve with time according to the cosmological background solutions $\bar{U}(t)$ and $\bar{S}(t)$. As structures form and become non-linear, the analysis of [161] shows that the solutions for U and S remain of the form

$$U(t,\mathbf{x}) = \bar{U}(t) + \delta U(t,\mathbf{x})\,, \qquad S(t,\mathbf{x}) = \bar{S}(t) + \delta S(t,\mathbf{x})\,, \tag{3.44}$$

where $\delta U(t,\mathbf{x})$ and $\delta S(t,\mathbf{x})$ remain small perturbations of $\bar{U}(t)$ and $\bar{S}(t)$, respectively. In essence, the physical reason behind this result is that, even when structures become non-linear, e.g. in the formation of galaxies, clusters, etc., the metric perturbation Φ never become large. In non-linear structure formation are rather the second spatial derivatives of Φ that become large compared to their values in the linear regime, in particular the Laplacian of Φ, which is related to the density contrast and can become huge; however, the spatial derivatives of Φ never enter in the equations that govern the dynamics of the auxiliary fields U and S. Indeed, in a perturbed FRW metric, to first order in Φ, the explicit expression of the d'Alembertian is

$$\Box U = -(1 + 2\Phi)(\ddot{U} + 3H\dot{U}) - 4\dot{\Phi}\dot{U} + a^{-2}(1 - 2\Phi)\nabla^2 U\,, \tag{3.45}$$

so spatial derivatives of Φ do not appear. As a result, non-linear structure formation does not stop the time evolution that the auxiliary fields inherited from the earlier epoch described by a spatially homogeneous FRW solutions. Near massive bodies, the perturbations $\delta U(t,\mathbf{x})$, $\delta S(t,\mathbf{x})$ just reduce to the static solutions $U(r)$, $S(r)$ studied in the previous subsection, and remain small as long as r is larger than the Schwarzschild radius of the massive bodies (recall for instance that $U(r) = (r_S/r)\cos mr$, which is much smaller that one for $r \gg r_S$). So, in the end, at the Earth-Moon system scale, the solution for S is, with good approximation, the sum of the cosmological and static solutions, $S(t,\mathbf{x}) = \bar{S}(t) + S_{\text{static}}(r)$.[31] A study of purely static solutions misses the term $\bar{S}(t)$, because assumes from scratch that the solution is time-independent. This time dependence induces a time-dependence of the Newton's constant, such that the RR model violates the limit (3.42). This rules out the RR model. We will see in appendix A that this problem affects also other nonlocal models that were proposed in the literature.

As we have seen in eq. (3.28), in the RT model, in contrast, the effective Newton's constant on small scales reduces to G, and looses all dependence on the auxiliary fields, so it passes without problems also the LRR constraint.

---

[31]This was also shown to happen exactly in models, such as galileons or k-essence, in which a field φ has a shift symmetry $\varphi \to \varphi + \text{const}$. In this case, thanks to the shift symmetry, near the present epoch t_0 the equation of motion of the field admits an exact solution with separation of variables of the form $\varphi(t,r) = \varphi_{\text{static}}(r) + \varphi_{\text{cosmo}}(t_0) + (t - t_0)\dot{\varphi}_{\text{cosmo}}(t_0)$ [162]. See also [163] for a related example.

3.5 Tensor perturbations and modified GW propagation

Until now we have studied the cosmological consequences of the theory at the level of background evolution and scalar perturbations. We now turn to tensor perturbations, i.e. gravitational waves (GWs) propagating in FRW. We will see, following [115, 164, 165], that the RT model has striking predictions in the tensor sector, that could be detected in the near future by GW detectors.

3.5.1 Tensor perturbations in GR

Let us begin by recalling that, in GR, the evolution of tensor perturbations over FRW is governed by the equation

$$\tilde{h}_A'' + 2\mathcal{H}\tilde{h}_A' + k^2\tilde{h}_A = 16\pi G a^2 \tilde{\sigma}_A \,, \tag{3.46}$$

where $\tilde{h}_A(\eta, \mathbf{k})$ are the Fourier modes of the GW amplitude, and we use the index $A = +, \times$ to label the two polarizations. We are using now conformal time η, related as usual to cosmic time t by $dt = a(\eta)d\eta$, and $a(\eta)$ is the scale factor. In this section the prime denotes the derivative with respect to cosmic time η, and $\mathcal{H} = a'/a$. The source term $\tilde{\sigma}_A(\eta, \mathbf{k})$ is related to the helicity-2 part of the anisotropic stress tensor (see e.g. [17]). In the following we will be interested in the free propagation between source and observer, and we will set it to zero. It is convenient to introduce a field $\tilde{\chi}_A(\eta, \mathbf{k})$ from

$$\tilde{h}_A(\eta, \mathbf{k}) = \frac{1}{a(\eta)}\tilde{\chi}_A(\eta, \mathbf{k}) \,. \tag{3.47}$$

Then eq. (3.46) becomes

$$\tilde{\chi}_A'' + \left(k^2 - \frac{a''}{a}\right)\tilde{\chi}_A = 0 \,. \tag{3.48}$$

For modes well inside the horizon, such as the GWs targeted by ground-based and space-born detectors, the term $a''/a \sim 1/\eta^2$ is totally negligible with respect to k^2; for instance, for a GW with a frequency $f \sim 10^2$ Hz, as typical of ground-based interferometers, $(k\eta)^{-2} \sim (500\,\mathrm{km}/H_0^{-1})^2 \sim 10^{-41}$. We can then neglect the term a''/a in eq. (3.48), which then becomes a standard a wave equation for $\tilde{\chi}_A$, that tells us that GWs propagate at the speed of light (that we have set here equal to unity).

The factor $1/a$ in eq. (3.47) tells us how the GW amplitude decreases as it propagates across cosmological distances, from the source to the observer. For inspiraling binaries this factor combines with other factors coming from the transformation of masses and frequency from the source frame to the detector frame (see e.g. section 4.1.4 of [166]), to produce the well-known dependence of the GW amplitude $\tilde{h}_A(\eta, \mathbf{k}) \propto 1/d_L(z)$, where d_L is the luminosity distance to the source. This is the origin of the fact that coalescing binaries are 'standard sirens', i.e. their waveform allows a direct reconstruction of the luminosity distance to the source [167–179]. In GR, for a cosmological model with energy density $\rho_{\mathrm{DE}}(z)$, the relation between luminosity distance and redshift is

$$d_L(z) = \frac{1+z}{H_0}\int_0^z \frac{d\tilde{z}}{\sqrt{\Omega_M(1+\tilde{z})^3 + \Omega_R(1+\tilde{z})^4 + \rho_{\mathrm{DE}}(\tilde{z})/\rho_0}} \,. \tag{3.49}$$

Therefore, a simultaneous measurement of d_L and of the redshift z (with an electromagnetic counterpart, or the study of the $d_L - z$ relation with statistical methods) allows us to get

cosmological information. In particular, for sources at small redshift, $z \ll 1$, eq. (3.49) reduces to the Hubble law $d_L(z) \simeq H_0^{-1} z$, so from a measurement at such redshifts we can get a measurement of H_0. This has indeed been possible with the detection of the binary neutron star (BNS) coalescence GW170817, which is at a redshift $z \simeq 0.01$, and has given the value $H_0 = 70.0^{+12.0}_{-8.0}$ km s^{-1} Mpc^{-1} [180]. The detection of coalescences at higher redshift could in principle allow us to access also the DE equation of state.

3.5.2 Tensor perturbations in modified gravity

As we will see, the free propagation of tensor perturbations in the RT model is governed by an equation of the form

$$\tilde{h}''_A + 2\mathcal{H}[1 - \delta(\eta)]\tilde{h}'_A + k^2 \tilde{h}_A = 0 \,, \tag{3.50}$$

for a given function $\delta(\eta)$. It is however instructive to first work out the implications of eq. (3.50) with a generic function $\delta(\eta)$, since this equations appears in many other modified gravity models. Indeed, in a generic modified gravity model both the "friction term" $2\mathcal{H}\tilde{h}'_A$ and the term $k^2 \tilde{h}_A$ in eq. (3.46) are modified. As we will recall below, the models that modify the $k^2 \tilde{h}_A$ term predict a speed of gravity different from the speed of light. The observation of GW170817 and of the associated GRB has set a limit $|c_{\rm gw} - c|/c < O(10^{-15})$ [181], so such models are ruled out.[32] In particular, a large class of Horndeski theories and other modifications of GR have been ruled out by this limit [183–186]. It turns out that the models that survive this constraint still modify the friction term. A propagation equation of the form (3.50) was indeed first found in some scalar-tensor theories of Horndeski type [187–190] and in the RR nonlocal model [23, 164]. In [191] it was shown that it also takes place in many other Horndeski-type theories that pass the test on speed of gravity (such as $f(R)$ theories, Jordan-Brans-Dicke, galileon cosmology, etc.), in Degenerate Higher Order Scalar-Tensor (DHOST) theories, and in bigravity. Similar effects take place in theories with extra dimensions, as originally found in [192] (see also [193]), although in this case they are due to the loss of gravitons to the bulk and, in general, are not described by eq. (3.50) (see also [194] for a discussion a modified GW propagation within the effective field theory approach to dark energy, [195, 196] for general formalisms for testing gravity with GW propagation, and [197–199] for further related work in the context of scalar-tensor theories).

Let us then study first the general consequences of eq. (3.50) (we closely follow the discussion in [164, 165]). We proceed as in the GR case, except that now, to eliminate the friction term, we must introduce $\tilde{\chi}_A(\eta, \mathbf{k})$ from

$$\tilde{h}_A(\eta, \mathbf{k}) = \frac{1}{\tilde{a}(\eta)} \tilde{\chi}_A(\eta, \mathbf{k}) \,, \tag{3.51}$$

where $\tilde{a}$ now satisfies

$$\frac{\tilde{a}'}{\tilde{a}} = \mathcal{H}[1 - \delta(\eta)] \,. \tag{3.52}$$

Then we get

$$\tilde{\chi}''_A + \left(k^2 - \frac{\tilde{a}''}{\tilde{a}} \right) \tilde{\chi}_A = 0 \,. \tag{3.53}$$

---

[32]Although it could still in principle happen that there is dependence on wavenumber that allows for $c_{\rm gw} \neq c$ for modes k well below the frequencies probed by LIGO/Virgo and restore $c_{\rm gw} = c$ to sufficient accuracy at LIGO/Virgo frequencies. This could be motivated in some models [182].

Once again, inside the horizon the term $\tilde{a}''/\tilde{a}$ is totally negligible. The remaining equation,

$$\tilde{\chi}_A'' + k^2\tilde{\chi}_A = 0 \, , \tag{3.54}$$

shows that GWs still propagate at the speed of light. This is a consequence of the fact that the term $k^2\tilde{\chi}_A$ in eq. (3.50) is the same as in GR. If the coefficient of this term had been different, we would get a speed of GWs $c_{\rm gw} \neq c$.

As we see from eq. (3.51), the effect of the modified friction term is that now the amplitude of $\tilde{h}_A$ is proportional to $1/\tilde{a}$ rather than $1/a$. Then, in the propagation from the source to the observer, the amplitude is multiplied by a factor $\tilde{a}_{\rm emis}/\tilde{a}_{\rm obs} \equiv \tilde{a}(z)/\tilde{a}(0)$, instead of a factor $a_{\rm emis}/a_{\rm obs} = a(z)/a(0)$, where the labels refer to the emission time (at redshift z) and the observation time, at redshift zero, respectively. Therefore

$$\tilde{h}_A \propto \frac{\tilde{a}(z)}{\tilde{a}(0)}\frac{a(0)}{a(z)}\frac{1}{d_L(z)} = \frac{\tilde{a}(z)}{a(z)}\frac{1}{d_L(z)} \, , \tag{3.55}$$

where $d_L(z)$ is the usual notion of luminosity distance (note that, since only the ratios $\tilde{a}(z)/\tilde{a}(0)$ and $a(z)/a(0)$ enter, without loss of generality we can choose the normalizations $\tilde{a}(0) = a(0) = 1$). Equation (3.55) motivates the introduction of a 'GW luminosity distance' $d_L^{\rm gw}(z)$ [164], related to the standard luminosity distance appropriate for electromagnetic signals, that we henceforth denote by $d_L^{\rm em}(z)$, by $d_L^{\rm gw}(z) = [a(z)/\tilde{a}(z)]\, d_L^{\rm em}(z)$. Rewriting eq. (3.52) as $(\log a/\tilde{a})' = \delta(\eta)\mathcal{H}(\eta)$ and integrating, we get [164]

$$d_L^{\rm gw}(z) = d_L^{\rm em}(z)\exp\left\{-\int_0^z \frac{dz'}{1+z'}\,\delta(z')\right\} \, . \tag{3.56}$$

In modified gravity, the quantity extracted from a measurement of the GW amplitude of a coalescing binary is $d_L^{\rm gw}(z)$, rather than $d_L^{\rm em}(z)$. To avoid misunderstandings, notice that the actual distance traveled by GWs from the source to the observer is the same as the distance traveled by electromagnetic signals. Equation (3.56) is simply a convenient way of expressing the fact that, in modified gravity, the amplitude of the GW decreases in a different way during the propagation, so that, for a coalescing binary, the observed amplitude, rather than depending only on $d_L^{\rm em}(z)$ and on the inclination of the orbit, as in GR, it further depends on $\delta(z)$, in such a way that the combined dependence on $d_L^{\rm em}(z)$ and $\delta(z)$ can be reabsorbed into the quantity $d_L^{\rm gw}(z)$.[33]

3.5.3 Predictions of the RT model

We now discuss GWs in the RT model, focusing on the signal from coalescing binaries at cosmological distances.[34] First of all, notice that this model only changes the gravitational part of the action but not the matter action, so the coupling to matter is unchanged, and at the linearized level, is still given by the usual $h_{\mu\nu}T^{\mu\nu}$ coupling. Thus, the source term in eq. (3.46) is not affected. Furthermore we have seen that, at short scales, such as the distance between the two bodies in a coalescing binary, the RT model reduces to GR to huge accuracy, so there is no appreciable modification to the orbital dynamics of a binary system, and the waveform produced by a coalescing binary in the region far from the source (where

---

[33]A different effect is provided by the fact that, in brane models, a gravitational signal can travel along geodesics in the extra dimensions, while electromagnetic signals are confined to the (3+1)-dimensional brane. This can lead to delays between the arrival time of a GW and the associated electromagnetic signal [200–202].

[34]See [165] for a discussion of how modified GW propagation affects the ISW effect.

the $1/r$ GW behavior sets in, but still the expansion of the Universe can be neglected) is the same as in GR. In the signal received by a coalescing binary, the only difference will then come from the free propagation of the GW from the source to the observer, across cosmological distances.

The equation governing the free propagation of tensor perturbations in the RT model has been computed in [114], and is[35]

$$\tilde{h}_A'' + [2\mathcal{H} - 3\gamma\bar{V}aH_0]\tilde{h}_A' + k^2\tilde{h}_A = 0\,. \tag{3.57}$$

So the 'friction term' $2\mathcal{H}\tilde{h}_A'$ is modified with respect to GR, but the term $k^2\tilde{h}_A$ is not. Thus, first of all we see that the RT model passes the constraints from the speed of GWs. As we have mentioned, this is a non-trivial constraint that has ruled out many modified gravity theories. Equation (3.57) is of the form (3.50), with

$$\delta(\eta) = \frac{3\gamma\bar{V}(\eta)H_0}{2H(\eta)}\,, \tag{3.58}$$

where we have used $\mathcal{H} = aH$. Recall that, for the RT model in a FRW background, we have defined the auxiliary field V from $V = H_0 S_0$, where S_0 is the $\mu = 0$ component [in coordinates $(t, \mathbf{x})$] of the auxiliary four-vector field S_μ of the RT model, see eq. (3.22). Recalling the definition (3.6) of γ, we can also write eq. (3.58) as

$$\delta(\eta) = \frac{m^2\bar{S}_0(\eta)}{6H(\eta)}\,. \tag{3.59}$$

Using the numerical solution of the background evolution equation of the RT model studied in section 3.1, we can therefore immediately compute δ and $d_L^{\rm gw}/d_L^{\rm em}$, as functions of the redshift. The results are shown in figure 15 (see also [115]). These results are quite spectacular, in particular at large ΔN. For instance, for $\Delta N = 64$, at large z the ratio $d_L^{\rm gw}/d_L^{\rm em}$ tends asymptotically to a value $\simeq 1.65$, corresponding to a 65% deviation from GR, a truly huge effect. In the limit of large ΔN [exemplified here by the case $(M_{\rm infl} = 10^{16}\,{\rm GeV}, \Delta N = 100)$; as we mentioned, for $M_{\rm infl} = 10^{16}\,{\rm GeV}$ this asymptotic curve is actually reached already at $\Delta N \gtrsim 70$], at large z the ratio $d_L^{\rm gw}/d_L^{\rm em}$ reaches a value $\simeq 1.80$, i.e. a 80% deviation from GR! Similarly, at $z = 0$, $\delta(0)$, in the limit of large ΔN, saturates to a value -1.11, so, in eq. (3.50), near $z = 0$ the term $1 - \delta(0) \simeq 2.11$ is more than twice the GR value.

This is very surprising because we have seen that, for all values of ΔN, the RT model differs from ΛCDM by less that 1% at the level of background evolution (see figure 2), and by a few percent to below percent level, depending on wavenumber, for the scalar perturbations, see e.g. figures 5–14. This is indeed what allows the model to fit well the current cosmological observations. One would have then naturally guessed that also in the tensor perturbation sector the differences would be of the same order. Instead, for large ΔN, they are much bigger, a very good news for GW experiments.

It is interesting to understand better how it is possible to have large deviations in the tensor sector when the background evolution and the scalar perturbations are much closer to

---

[35]Note that eq. (5.1) of ref. [114] was written with a different definition of the auxiliary field V. Denoting by $\tilde{V}$ the definition used there, by S_η the $\mu = 0$ component of S_μ in coordinates $(\eta, \mathbf{x})$ and by S_0 the $\mu = 0$ component of S_μ in coordinates $(t, \mathbf{x})$, we have $\tilde{V} \equiv S_\eta = aS_0 = aV/H_0$, where V is the definition used here. The equation written in appendix A.1 of [114], where is described the implementation in CLASS of the perturbations of the model, are also written denoting by V the quantity that we are here calling $\tilde{V}$.
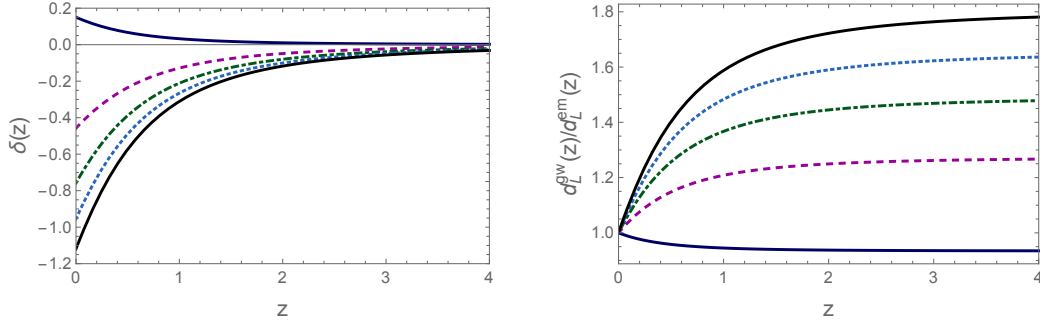
Figure 15. The functions $\delta(z)$ (left panel) and $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ (right panel), for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted) and $\Delta N = 100$ (black solid line).
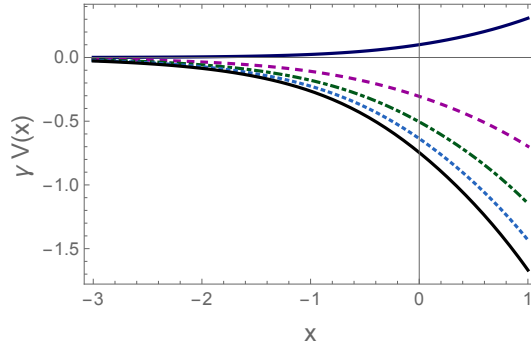

Figure 16. The quantity $\gamma \bar{V}(x)$ for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted) and $\Delta N = 100$ (black solid line), using for each model its own mean values of H_0 and Ω_M.

ΛCDM. Technically, the fact that $\delta(z=0)$ becomes of order one, and therefore, accordingly to eq. (3.50), gives a significant correction to the GR result, can be traced to the evolution of the auxiliary field $V(\eta)$, that determines $\delta(\eta)$ through eq. (3.58). The behavior of $\gamma V(\eta)$ is shown in figure 16, as a function of $x = \ln a(\eta)$. We see that, in the evolution for large ΔN, it becomes indeed of order one (in absolute value) near the present epoch. In contrast, at the level of background evolution, the deviations from the GR behavior are encoded in the DE equation of state, which, according to eqs. (3.7) and (3.18), is related to Y by $1 + w_{\rm DE} = -Y'/(3Y)$, where $Y = U - \dot{S}_0 = U - hV'$. Therefore, the deviation of $w_{\rm DE}$ from -1 is controlled by a different combination, and in particular is not sensitive to the numerical value of $\gamma V(\eta)$. Rather, in the approximation where we neglect U compared to V' [and we approximate $(hV)' \simeq hV''$], it is sensitive to V''/V'. It is therefore perfectly possible to have large deviations in the tensor perturbations even when the background evolution is close to ΛCDM. Concerning scalar perturbations, the crucial difference is that they satisfy Poisson-like equations, rather than a dynamical propagation equation involving the $\square$ operator. The modified Poisson equation in the RT model reads [see eq. (A.6) of [86]]

$$\hat{k}^2 \Phi + 3(\Phi' - \Psi) = \frac{3}{2h^2}\left[\frac{\delta\rho}{\rho_0} + \gamma\left(\delta U - h\delta V' + 2h\Psi\bar{V}' + h\Psi'\bar{V}\right)\right] . \tag{3.60}$$

Then $\gamma V(\eta)$ only affects the source term; as we have seen, it becomes of order one in the recent epoch, but, by that time, all modes of cosmological interest are well inside the horizon, so
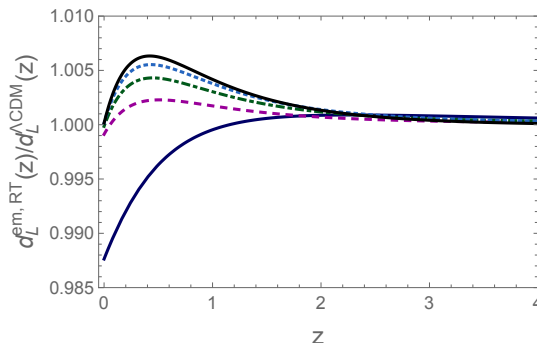
Figure 17. The ratio of $d_L^{em}(z)$ computed in the RT model to the luminosity distance of ΛCDM for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed) and $\Delta N = 64$ (cyan, dotted) and $\Delta N = 100$ (black solid line), using for each model its own mean values of H_0 and Ω_M.

$\hat{k}^2 \gg 1$ [where $\hat{k} = k/(aH)$, see eq. (3.24)] and the effect of $\gamma V(\eta)$ is by now suppressed by the small factor $1/\hat{k}^2$. Note that also in eq. (3.57), for GWs observed on ground based detectors, $\hat{k} \gg 1$. Nevertheless, the crucial difference is that eq. (3.57) is a propagation equation, involving the second time derivative of h, so that the term $k^2 h$ determines the phase of the wave, while the friction term independently determines the scaling of the amplitude with redshift. So, in this case the fact that in tensor perturbations we can have a large effect, while this does not happen for scalar perturbations, has its roots in the fact that tensor perturbations are propagating degrees of freedom, while scalar perturbations are not.

For comparing the prediction in the tensor sector of the RT model to those of ΛCDM the relevant quantity, rather than the ratio of d_L^{gw} to d_L^{em}, both computed within the RT model, is actually the ratio of d_L^{gw}, computed in the RT model, to the luminosity distance $d_L^{\Lambda CDM}$ computed in ΛCDM (for which the notion of electromagnetic and GW luminosity distance coincide), and in which, in each model, the respective mean values of the parameters H_0 and Ω_M are used. However, the results for $d_L^{gw,RT}/d_L^{\Lambda CDM}$ turn out to be practically the same as the results shown in the right panel of figure 15. This can be seen by writing

$$\frac{d_L^{gw,RT}(z)}{d_L^{\Lambda CDM}(z)} = \left(\frac{d_L^{gw,RT}(z)}{d_L^{em,RT}(z)} \right) \times \left(\frac{d_L^{em,RT}(z)}{d_L^{\Lambda CDM}(z)} \right) , \tag{3.61}$$

where, for clarity, we have denoted by $d_L^{gw,RT}(z)$ the GW luminosity distance d_L^{gw} in the RT model. The first factor on the right-hand side is the quantity that we have already shown in the right panel of figure 15. The second factor is shown in figure 17, and we see that is very close to one; in particular, for the RT model with large ΔN, it reaches at most a value of order 1.006 for $\Delta N = 100$ near $z \simeq 0.3$, and then quickly goes asymptotically to values of order 1.001. This can be understood observing that the ratio $d_L^{em,RT}(z)/d_L^{\Lambda CDM}(z)$ is determined by two factors. First, by the different mean values of H_0 and Ω_M between the RT model with the given ΔN and ΛCDM; second, by the different redshift dependence of the DE density, or, equivalently, the different DE equation of state $w_{DE}(z)$. However, we have seen in table 2 that Bayesian parameter estimation gives for the RT model values of H_0 and Ω_M very close to those of ΛCDM, particularly at large ΔN; furthermore, as discussed in [165], the change in the value of these parameters goes precisely in the direction to cancel the effect in the change of the DE equation of state. This is due to the fact that Bayesian parameter estimation in

	RT, minimal	$\Delta N = 34$	$\Delta N = 50$	$\Delta N = 64$	$\Delta N = 100$
Ξ_0	0.93	1.27	1.49	1.65	1.80
n	2.59	2.08	2.00	1.95	1.91
$\delta(0)$	0.15	-0.46	-0.76	-0.95	-1.12
$\delta(0)/(1 - \Xi_0)$	2.29	1.67	1.54	1.46	1.39

Table 5. Values of Ξ_0, n, $\delta(0) \equiv \delta(z = 0)$ and $\delta(0)/(1 - \Xi_0)$ for the RT model with various values of ΔN. The results have been obtained using for each model its own mean values for Ω_M and h_0 from table 2.

practice requires the model to fit some fixed distance scales at large redshifts, such as the scales given by the CMB peaks or by the BAO oscillations; thus, if, compared to ΛCDM, one changes $w_{\rm DE}(z)$ in the direction of giving, say, a larger (electromagnetic) luminosity distance at large redshift, H_0 and Ω_M change in the direction such that they partially compensate for this change. As a result the electromagnetic luminosity distance, particularly at moderate to large values z, changes very little. Thus, the difference in the GW luminosity distance of the RT model, compared to ΛCDM, in practice is entirely given by the effect of modified GW propagation, while the DE equation of state and the difference in H_0 and Ω_M among RT and ΛCDM have a negligible effect.

As discussed in [165], the z dependence of the ratio $d_L^{\rm gw}/d_L^{\rm em}$ is easily understood observing that, by definition, at $z \to 0$ we must have $d_L^{\rm gw}/d_L^{\rm em} \to 1$ because, if the distance to the source goes to zero, there can be no effect from modified GW propagation. At large z, $d_L^{\rm gw}/d_L^{\rm em}$ goes to a constant because, in the RT model, as in most other modified gravity model, the emergence of dark energy is a relatively recent phenomenon, so the modifications to GR, and hence the function $\delta(z)$ in eq. (3.50), go to zero at large redshifts. As a consequence, at large z the integral in eq. (3.56) saturates to a constant value. As shown in figure 18, the numerical results for $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ are extremely well fitted by the simple parametrization [165]

$$\frac{d_L^{\rm gw}(z)}{d_L^{\rm em}(z)} = \Xi_0 + \frac{1 - \Xi_0}{(1 + z)^n} \, , \tag{3.62}$$

in terms of two parameters Ξ_0 and n. This parametrization reproduces the fact that, at $z = 0$, $d_L^{\rm gw}(z)/d_L^{\rm em}(z) = 1$, while at large redshift $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ goes to a constant value Ξ_0. The index n determines the rate at which this asymptotic value is reached. The best-fit values of Ξ_0 and n are given in table 5.

Observe that the simple parametrization (3.62) reproduces the numerical results extremely well. Indeed, comparing with figure 3, we see that it works much better than the (w_0, w_a) parametrization for the equation of state. This is due to the fact that eq. (3.62) catches correctly both the $z \to 0$ limit and the large z limit.[36] The corresponding parametrization for the function $\delta(z)$ is obtained inverting eq. (3.56) to get

$$\delta(z) = -(1 + z) \frac{d}{dz} \log \left(\frac{d_L^{\rm gw}(z)}{d_L^{\rm em}(z)} \right) \, . \tag{3.63}$$

---

[36]Indeed, it was found in [191] that this parametrization fits very well the results of all other modified gravity models studied there, such as various Horndeski-type theories and DHOST theories. The only exception is given by bigravity, where it was found that, as a function of redshift, $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ has a series of oscillations due to the interaction between the two metrics.
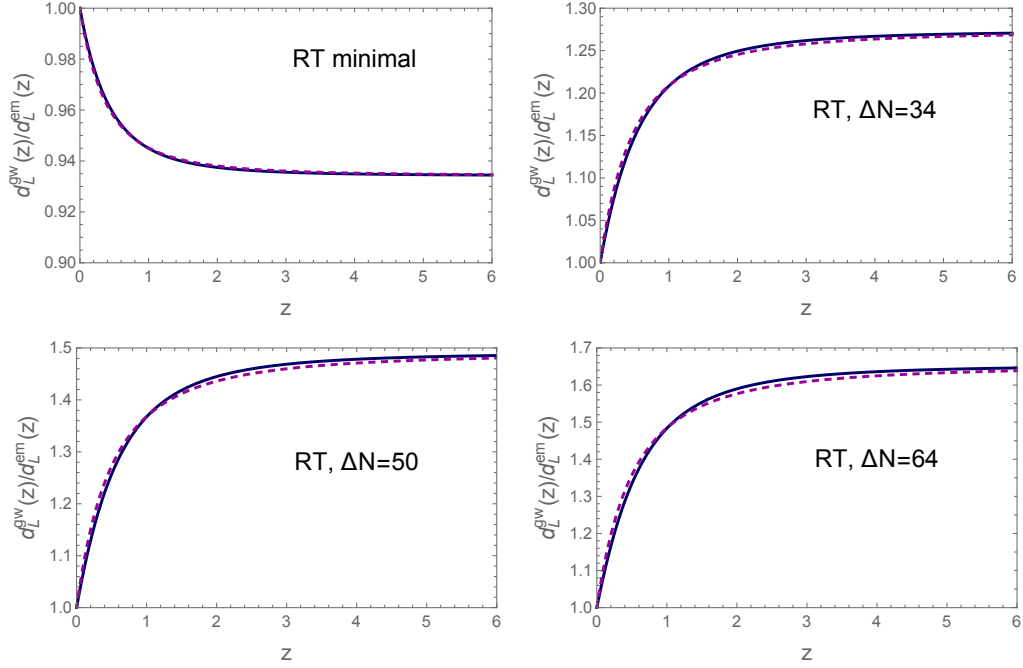
Figure 18. The function $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ from the numerical integration (blue solid line), compared with the fit (3.62) (magenta, dashed). Upper left panel: for the minimal RT model; upper right: RT with $\Delta N = 34$; lower left: RT with $\Delta N = 50$; lower right: RT with $\Delta N = 64$.
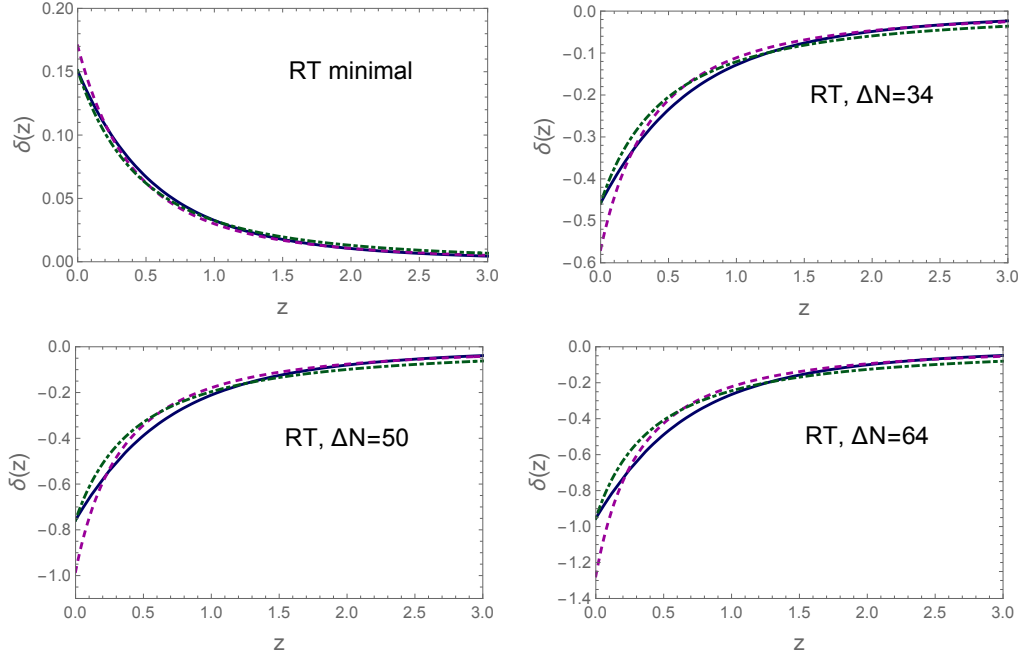

Figure 19. The function $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ from the numerical integration (blue solid line), compared to the parametrization (3.62) with the value of n obtained from the best fit to $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ (magenta, dashed), and with $n = \delta(z=0)/(1-\Xi_0)$ (green, dot-dashed). Upper left panel: for the minimal RT model; upper right: RT with $\Delta N = 34$; lower left: RT with $\Delta N = 50$; lower right: RT with $\Delta N = 64$.

Using eq. (3.62) for $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ gives

$$\delta(z) = \frac{n(1-\Xi_0)}{1-\Xi_0+\Xi_0(1+z)^n}\,. \tag{3.64}$$

Figure 19 compares the numerical result for $\delta(z)$ (blue solid line) with the fit (3.64), using the same values of Ξ_0 and n as in table 5 (magenta, dashed lines). We see that the (Ξ_0, n) parametrization provides a fit to $\delta(z)$ less good than to $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$, particularly near $z = 0$. This is due to the fact that, for $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$, the (Ξ_0, n) parametrization catches correctly both the value in $z = 0$ and the large z limit; thus, as long as $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ is smooth in between, it is natural to find a value of n such that the parametrization (3.62) performs well. In contrast, the value of $\delta(z = 0)$ is not automatically reproduced by the parametrization (3.64), and indeed we see from the figures that in this region the parametrization is not accurate. For instance, the numerical integration gives the values of $\delta(0) \equiv \delta(z = 0)$ shown in table 5, while the parametrization (3.64) would incorrectly predict $\delta(0) \simeq \{0.17, -0.57, -0.98, -1.27, -1.53\}$. Note that, with the parametrization (3.62), (3.64), we have

$$\delta(0) = n(1-\Xi_0)\,. \tag{3.65}$$

This suggests that, after having fixed Ξ_0 so to reproduce exactly the large-z behavior of $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$, rather then choosing n from a best fit to $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$, we could choose $n = \delta(0)/(1-\Xi_0)$, so that the parametrization (3.64) reproduces exactly the value of $\delta(0)$. The values obtained in this way are given in the last line of table 5. If one uses these values of n, the fit to $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ significantly degrades, but the fit to $\delta(z)$ becomes more accurate, and is shown as the green dot-dashed lines in figure 19. In general, since the directly observable quantity is $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$, it is more important to have a simple and accurate analytic representation for it, rather than for $\delta(z)$. Of course, for an accurate comparison with the data, one can also use directly the results of the numerical integration, which are obtained very quickly.

For comparison, the result for the RR model is also of the form (3.50), except that the function δ is given by

$$\delta = \frac{3\gamma\, d\bar{V}/d\log a}{2(1-3\gamma\bar{V})}\,, \tag{3.66}$$

and now $V = H_0^2 S$, where S is the auxiliary field of the RR model, defined by $U = -\Box^{-1}R$ and $S = -\Box^{-1}U$. The numerical integration then gives again a result very well fitted by eq. (3.62), with $\Xi_0 \simeq 0.97$ and $n \simeq 2.5$ [165]. However, contrary to the RT model, here the deviation from GR is only about 3%.

3.5.4 Energy density of GWs and conservation of graviton number

The fact that the GW amplitude in FRW does not scale as $1/a$ raises a question. As we will recall below, in GR the fact that in FRW $h \propto 1/a$ ensures that the GW energy density $\rho_{\rm GW}$ scales as $1/a^4$; in turn, this is consistent with an interpretation of a GW as a collection of massless graviton, whose comoving number density (i.e. number per unit volume in comoving coordinates) is conserved. Indeed, the fact that the graviton number per comoving volume is conserved means that the graviton number per physical volume scales as $1/a^3$, while the fact that the graviton is massless implies that its energy scales as $1/a$, giving overall the $1/a^4$ behavior of $\rho_{\rm GW}$. One might then wonder whether the scaling $h \propto 1/\tilde{a}$ is an indication that the (comoving) graviton number is not conserved in the RT model. We will see here that, in

fact, even in the RT model $\rho_{\rm GW}$ scales as $1/a^4$ and therefore the comoving number density of gravitons is still conserved. To this purpose, one must realize that the expression of $\rho_{\rm GW}$ in a generic modified gravity model is different from the GR expression. Let us first recall how things work in GR. We consider tensor perturbations over the FRW metric. Using conformal time, we write

$$ds^2 = a^2 \left[-d\eta^2 + \left(\delta_{ij} + h_{ij}^{\rm TT} \right) dx^i dx^j \right] . \tag{3.67}$$

It is convenient to expand the Fourier transform of $h_{ij}^{\rm TT}$ in the basis of the polarization tensors,

$$\tilde{h}_{ij}^{\rm TT}(\eta, \mathbf{k}) = \sum_{A=+,\times} e_{ij}^A(\hat{\mathbf{k}}) \tilde{h}_A(\eta, \mathbf{k}) , \tag{3.68}$$

where the polarization tensors are normalized as $e_{ij}^A(\hat{\mathbf{k}}) e_{ij}^{A'}(\hat{\mathbf{k}}) = 2\delta^{AA'}$. Expanding the Einstein-Hilbert action to second order in $h_{ij}^{\rm TT}$ one then finds (see e.g. section 21.3.4 of [17])

$$
\begin{aligned}
S_2[h] &= \frac{1}{32\pi G} \sum_A \int d^3x \, d\eta \, a^2 \left[\partial_\eta h_A \partial_\eta h_A - \partial_k h_A \partial_k h_A \right] \\
&= -\frac{1}{2} \sum_A \int d^4x \, \sqrt{-\bar{g}} \, \bar{g}^{\mu\nu} \partial_\mu \varphi_A \partial_\nu \varphi_A ,
\end{aligned} \tag{3.69}
$$

where $\bar{g}_{\mu\nu} = a^2 \eta_{\mu\nu}$ is the background FRW metric in $(\eta, \mathbf{x})$ coordinates, and

$$\varphi_A(\eta, \mathbf{x}) = \frac{1}{\sqrt{16\pi G}} \, h_A(\eta, \mathbf{x}) . \tag{3.70}$$

The action governing the two polarization amplitudes h_A is therefore the same as the curved-space action of two canonically-normalized scalar fields φ_A. The variation of the action (3.69) gives eq. (3.46) (with the left-hand side equal to zero, unless we add also the matter action). At the same time, from this action we can get the energy-momentum tensor of GWs,

$$
\begin{aligned}
t_{\mu\nu} &\equiv -\frac{2}{\sqrt{-\bar{g}}} \left\langle \frac{\delta S_2[h]}{\delta \bar{g}^{\mu\nu}} \right\rangle = \sum_A \left\langle \partial_\mu \varphi_A \partial_\nu \varphi_A - g_{\mu\nu} \frac{1}{2} g^{\rho\sigma} \partial_\rho \varphi_A \partial_\sigma \varphi_A \right\rangle \\
&= \frac{1}{16\pi G} \sum_A \left\langle \partial_\mu h_A \partial_\nu h_A - g_{\mu\nu} \frac{1}{2} g^{\rho\sigma} \partial_\rho h_A \partial_\sigma h_A \right\rangle ,
\end{aligned} \tag{3.71}
$$

where $\langle \ldots \rangle$ denotes the spatial average over several wavelengths of the GWs, or the temporal average over several periods (see e.g. section 1.4 of [166]). We denote by $t_{\eta\eta}$ the $\mu = \nu = 0$ component of $t_{\mu\nu}$ in coordinates $(\eta, \mathbf{x})$ and by $t_{tt} \equiv t_{00}$ the $\mu = \nu = 0$ component of $t_{\mu\nu}$ in coordinates $(t, \mathbf{x})$. From $t_{\eta\eta}(d\eta)^2 = t_{tt}(dt)^2$ and $dt = a \, d\eta$ if follows that $t_{00} = t_{\eta\eta}/a^2$, so eq. (3.71) gives

$$t_{00} = \frac{1}{32\pi G} \frac{1}{a^2} \sum_A \langle (\partial_\eta h_A)^2 + (\partial_i h_A)^2 \rangle . \tag{3.72}$$

On a plane wave the terms $\langle (\partial_\eta h_A)^2 \rangle$ and $\langle (\partial_i h_A)^2 \rangle$ are equal. From eq. (3.54), for wavelengths well inside the horizon, i.e. for $k\eta \gg 1$, $\tilde{\chi}_A(\eta, \mathbf{k}) \propto \sin(k\eta + \alpha)$, with α a phase. Therefore $\tilde{h}_A(\eta, \mathbf{k}) \propto \sin(k\eta + \alpha)/a(\eta)$ and, again for $k\eta \gg 1$,

$$\partial_\eta h_A(\eta, \mathbf{k}) \propto \frac{k \cos(k\eta + \alpha)}{a(\eta)} \left[1 + O\left(\frac{1}{k\eta} \right) \right] . \tag{3.73}$$

In $\langle(\partial_\eta h_A)^2\rangle$ the term $\cos^2(k\eta + \alpha)$, averaged over several periods, simply gives a factor $1/2$, so $\langle(\partial_\eta h_A)^2\rangle \propto 1/a^2$ and, from eq. (3.72), it then follows that $\rho_{\rm gw} = t_{00}$ is proportional to $1/a^4$, as indeed we expect for any form of radiation.

Let us now see how the situation changes in the RT model. The propagation equation is now given by eq. (3.50). Using eq. (3.52) we see that it can be obtained from the GR equation with the replacement $a(\eta) \to \tilde{a}(\eta)$. It can then be formally obtained from the variation of a quadratic action obtained replacing $a(\eta) \to \tilde{a}(\eta)$ in eq. (3.69), i.e. from[37]

$$S_2^{\rm RT}[h] = \frac{1}{32\pi G} \sum_A \int d^3x\, d\eta\, \tilde{a}^2 \left[\partial_\eta h_A \partial_\eta h_A - \partial_k h_A \partial_k h_A\right] . \tag{3.74}$$

Introducing an effective Newton's constant from

$$\frac{1}{\tilde{G}(\eta)} \equiv \frac{1}{G} \frac{\tilde{a}^2(\eta)}{a^2(\eta)} \tag{3.75}$$

we can rewrite eq. (3.74) as

$$S_2^{\rm RT}[h] = \sum_A \int d^3x\, d\eta\, \frac{1}{32\pi \tilde{G}(\eta)} a^2 \left[\partial_\eta h_A \partial_\eta h_A - \partial_k h_A \partial_k h_A\right] . \tag{3.76}$$

Thus, as far as tensor perturbations are concerned, at the quadratic level the RT model can be obtained from GR with the replacement $G \to \tilde{G}(\eta)$. Note that $\tilde{G}(\eta)$ plays the role of an effective Newton's constant for tensor perturbations only. As we saw in section 3.2.2, scalar perturbations are governed by a different effective Newton's constant, that we denoted as $G_{\rm eff}(\eta, k)$, and which, contrary to $\tilde{G}(\eta)$, depends also on the wavenumber k.

Repeating the above derivation of the energy-momentum tensor of GWs, eq. (3.71) becomes

$$t_{\mu\nu} = \frac{1}{16\pi \tilde{G}(\eta)} \sum_A \langle \partial_\mu h_A \partial_\nu h_A - g_{\mu\nu} \frac{1}{2} g^{\rho\sigma} \partial_\rho h_A \partial_\sigma h_A \rangle , \tag{3.77}$$

simply because the variation $\delta S_2[h]/\delta \bar{g}^{\mu\nu}$ is insensitive to the time dependence of $\tilde{G}(\eta)$. The energy density $\rho_{\rm gw} = t_{00}$ is then given by

$$\rho_{\rm gw} = \frac{1}{16\pi \tilde{G}(\eta)} \frac{1}{a^2} \sum_A \langle (\partial_\eta h_A)^2 \rangle . \tag{3.78}$$

Notice that the $1/a^2$ factor comes from the transformation from $t_{\eta\eta}$ to t_{tt}, i.e. from the relation $dt = a d\eta$. This is determined by the FRW background metric, so it still involves a rather than $\tilde{a}$. In contrast, $h_A \propto \sin(k\eta + \alpha)/\tilde{a}$ and therefore now, for $k\eta \gg 1$, $\partial_\eta h_A \propto k \cos(k\eta + \alpha)/\tilde{a}$, which replaces eq. (3.73). Again, the term $\cos^2(k\eta + \alpha)$ averages to $1/2$, so in the end the time dependence of $\rho_{\rm gw}$ is

$$\rho_{\rm gw} \propto \frac{1}{16\pi \tilde{G}(\eta)} \frac{1}{a^2 \tilde{a}^2} = \frac{1}{16\pi G} \frac{1}{a^4} . \tag{3.79}$$

Therefore, once taken into account the fact that the modification of the Einstein equations implies also a modification of the formula for the GW energy-momentum tensor, we find

---

[37]More precisely, this is the action that reproduces the linearized equations of motions of the RT model, after having substituted the auxiliary fields with their own solutions of the equations of motion. It is therefore a 'reduced' action for the h_A variables only.

that, in FRW, the GW energy density of the RT model still scales as $1/a^4$, despite modified GW propagation. Therefore, the energy density still corresponds to that of an ensemble of massless gravitons, whose number density in comoving coordinate is constant (so that the number density in physical coordinates scales as $1/a^3$) and whose energy scales as $1/a$. From the derivation, it is also clear that this result is not specific to the RT model, but holds for any modified gravity model where the equation of tensor perturbations can be written in the form (3.50). Notice also that the redshift of the graviton frequency $\omega \propto 1/a$, or of the wavelength as $\lambda \propto a$, are kinematical properties that depend only on the background metric, and are the same in GR and in the RT model. As discussed in [165], in the RR model again $\rho_{\rm GW} \propto 1/a^4$, and in this case the effective Newton constant $\tilde{G}$ for the tensor perturbations is the same as the effective Newton's constant $G_{\rm eff}$ in the scalar sector.

This result also gives useful guidance for attempts at deriving the RT model from a fundamental local theory. In particular, it rules out the possibility that the RT model could be derived from a theory with extra dimensions in which gravitons are lost to a higher-dimensional bulk, see the discussion in section 2.4.2, and rather points toward the dynamical mass generation mechanisms discussed in section 2.4.3.

3.6 Comparison with the sensitivity of current and future GW detectors

We next compare the predictions for modified GW propagation of the RT model with the sensitivities of current and future GW detectors, elaborating on the analysis in [115, 165, 203, 204] for ground-based detectors and in [191] for LISA.

3.6.1 The Advanced LIGO/Virgo/Kagra network

We first consider the network of second-generation (2G) GW detectors formed by Advanced LIGO Hanford and Livingston, Advanced Virgo, KAGRA and LIGO India (HLVKI), assumed to be all at target sensitivity. In [203] mock catalogs of binary neutron stars (BNS) detections have been produced for this network, using state-of-the art models for the cosmic star formation rate, for the extra-galactic population of neutron star binaries and for the delay between binary formation and merger [205–214], and fixing the overall normalization using the local coalescence rate estimated from the O1 LIGO observation run and the O2 LIGO/Virgo observation run [215]. Assuming a duty cycle of 80% and a network SNR threshold level $\rho_{\rm threshold} = 12$, it was found that the HLVKI network will detect between $O(60)$ and $O(80)$ BNS/yr, depending on the assumptions on star formation rate and distribution of neutron star masses. Of these, only about 1–2 events per year are expected to have a detected gamma ray burst (GRB) counterpart, assuming that Fermi-GBM can make a coincident detection and that *Swift* can slew to the combined GW/GRB error box and identify an X-ray counterpart. More electromagnetic counterparts could in principle be detected with just optical/IR/UV telescopes, without a GRB trigger, although their number is more difficult to estimate.

A sample catalog of simulated GW-GRB coincidences is given in table 6 (from table 23 of [203]) which shows 15 joint GW-GRB coincidences detected in 10 years of simulated data.[38] The first three columns of the table show the redshift of the source, which has been extracted randomly from the appropriate distribution, its luminosity distance (which, being measured from the GW signal, is in principle a GW luminosity distance, $d_L^{\rm gw}$, if we do not assume

---

[38]Such a long time span is somewhat optimistic, but, given the rate of 1–2 joint GW-GRB events per year, is necessary to build a statistically significant sample.

z	d_L^{gw} (Mpc)	Δd_L^{gw} (Mpc)	$\Delta d_L^{\text{gw}}/d_L^{\text{gw}}$	$\Delta\delta(0)$
0.029271	134.815	4.000	0.030	1.36
0.035195	157.475	5.636	0.036	1.30
0.060585	283.567	18.706	0.066	1.25
0.066283	316.373	14.509	0.046	0.84
0.071053	327.381	20.085	0.061	1.00
0.071730	342.952	16.957	0.049	0.83
0.076180	341.595	22.360	0.065	0.99
0.081819	418.469	30.238	0.072	1.00
0.088698	396.734	25.757	0.065	0.84
0.091869	402.590	34.170	0.085	1.03
0.094237	406.423	31.472	0.077	0.93
0.095288	432.996	36.423	0.084	0.99
0.099956	491.071	31.721	0.065	0.75
0.102531	461.627	36.858	0.080	0.88
0.114869	626.939	43.010	0.068	0.68

Table 6. The events in a given realization of the mock catalog of joint GW-GRB detections for the HLVKI network, over 10 yr of simulated data. The 'measured' luminosity distance is obtained from the redshift assuming ΛCDM as fiducial model, and scattering randomly the fiducial values of $d_L^{\text{gw}}(z)$ according to a Gaussian distribution with a width equal to the error $\Delta d_L^{\text{gw}}(z)$ (from ref. [203]). In the last column we give the corresponding error on the measurement of $\delta(0)$ from each single source, assuming a 1% error on the electromagnetic luminosity distance.

GR), and the expected observational error on the luminosity distance Δd_L^{gw} (which depends on the network sensitivity and on the source orbital inclination and position in the sky with respect to the network, also extracted randomly). The 'measured' value of $d_L^{\text{gw}}(z)$ is obtained from the redshift assuming ΛCDM as fiducial model, and scattering randomly this fiducial value according to a Gaussian distribution with a width equal to the error $\Delta d_L^{\text{gw}}(z)$. From the fourth column we see that, for the sources at the lowest redshifts, d_L^{gw} can be measured to $(3-4)\%$ accuracy (depending in particular on the source inclination and position in sky with respect to the network), while, for the largest redshifts in the catalog, around $z \simeq 0.1$, the accuracy on d_L^{gw} is about $(7-8)\%$.[39]

For sources at small redshift, as appropriate for the values of z in table 6, eq. (3.56) becomes

$$\frac{d_L^{\text{gw}}(z)}{d_L^{\text{em}}(z)} = 1 - z\delta(0) + \mathcal{O}(z^2)\,, \tag{3.80}$$

so in this limit we are actually sensitive to $\delta(0) \equiv \delta(z=0)$. The comparison between the predictions of a model and the data can therefore be performed without making use of any parametrization for $d_L^{\text{gw}}(z)/d_L^{\text{em}}(z)$, and simply comparing directly the predictions of the

---

[39]For comparison, GW170817 was at $z \simeq 0.01$ and its luminosity distance, as measured from the GW signal, was $d_L^{\text{gw}} = 40^{+8}_{-14}$ Mpc [216]. The corresponding value of $\Delta d_L^{\text{gw}}/d_L^{\text{gw}}$, of order 27%, is much larger than those in table 6, because it reflects the detectors sensitivities during the O2 run, while in table 6 the five detectors are taken at target sensitivity. Furthermore, the event was near a blind spot of Virgo, so Virgo could contribute to the source localization but not to the estimate of the other source parameters, so only the two LIGO detectors contributed to the estimate of d_L.

model for $\delta(0)$ with the expected error on $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$. For the RT model, with different values of ΔN, the predictions for $\delta(0)$ were given in table 5.

Observe that, in eq. (3.80), the deviation of $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ from 1 is proportional to z. Limits on $\delta(0)$ from GW170817 were obtained in [165]. In this case, given the small redshift $z = 0.01$, it is clear that one cannot obtain stringent limits, and the best result found in [165] was $\delta(0) = -7.8^{+9.7}_{-18.4}$.[40] Let us now estimate the observational error on $\delta(0)$ that could be obtained from individual detections with characteristics such as those in table 6. From eq. (3.80),

$$\Delta\delta(0) \simeq \frac{1}{z} \left[\frac{\Delta d_L^{\rm gw}}{d_L^{\rm gw}} + \frac{\Delta d_L^{\rm em}}{d_L^{\rm em}} \right] . \tag{3.81}$$

The relative error on $d_L^{\rm gw}$ is given in table 6. For the relative error on $d_L^{\rm em}$ we observe that, given a measurement of the redshift from an electromagnetic counterpart, $d_L^{\rm em}$ is in principle determined by the fiducial cosmology, and in particular, at these redshifts, by the value of H_0. From table 2, the error $\Delta H_0/H_0$ is below the 1% level (and one can imagine that this accuracy will further improve in the next few years).[41] Note also that the redshifts in table 6 are sufficiently large that the peculiar velocity of the host galaxy, typically of order $v \sim 200\,{\rm km/s}$, gives a small error on the determination of the cosmological redshift, that can be neglected. So, we assume for definiteness a relative error $\Delta d_L^{\rm em}/d_L^{\rm em} = 1\%$ for all the events shown in table 6 (since this is in any way subleading with respect to the error on $\Delta d_L^{\rm gw}/d_L^{\rm gw}$, the precise value assumed is not very important). In this way we obtain the estimates for $\Delta\delta(0)$ given in the last column of table 6. We see that the accuracy obtained from the various individual detections are quite comparable in this range of redshift, with on average slightly more accurate measurements at higher redshift, since the average increase of the observational error $\Delta d_L^{\rm gw}/d_L^{\rm gw}$ with redshift is more than compensated by the factor $1/z$ in eq. (3.81).

The error σ obtained combining the errors σ_i of the individual measurements is given as usual by $1/\sigma^2 = \sum_i 1/\sigma_i^2$. Comparing with table 5, we see that the prediction $\delta(0) \simeq -1.11$ of the RT model in the large ΔN limit could be detected at the 3σ level with about 9 BNS with counterpart, which could be collected in 6 years of data taking. Verifying the predictions for smaller values of ΔN would require more data, but in any case beyond this point one would enter in a regime where a detection of $\delta(0)$ is in principle possible already at 2G detectors. Of course, the estimate of the number of detected electromagnetic counterparts is subject to uncertainties in the modelisation of the emissions mechanism, that will hopefully be further clarified by the ongoing and future LIGO/Virgo/KAGRA observational runs. Another interesting possibility is given by the detection of NS-BH binaries. These can be seen to larger distances, because of the higher BH mass, and would therefore be very useful for testing modified GW propagation. Theoretically, it is not known whether NS-BH coalescences have a significant electromagnetic emission. Currently, a few NS-BH candidates have been reported in the O3 LIGO/Virgo run (see https://gracedb.ligo.org), but apparently no counterpart has been observed. The other option that should be explored is the possibility of

---

[40]The recently announced detection, GW190425 [217], has a redshift $z = 0.03^{+0.01}_{-0.02}$ (assuming ΛCDM). The event is classified as a NS-NS, although the possibility that one or both binary components of the system are BHs cannot be ruled out from the GW data. No counterpart has been observed to date. The event has very poor angular localization because it was confidently detected only in a single detector.

[41]Of course, a crucial issue here is the discrepancy between the value of H_0 obtained from CMB+BAO+SNe in ΛCDM (or in the RT model, which is very close) and the value from local measurements [137, 138]. Here we perform our estimates assuming the value of H_0 and ΔH_0 from CMB+BAO+SNe.

using BNS without counterparts to learn about modified GW propagation, using statistical methods, such as those based on a probabilistically assignment of the host galaxy [167] (see [174] for recent Bayesian approach tuned to 2G detectors), or on the fact that the NS mass function is relatively narrow [176, 177].

A more accurate way of estimating the sensitivity to modified GW propagation, that fully accounts for the partial degeneracies of $\delta(0)$ [or of (Ξ_0, n)], with the other cosmological parameters, and in particular with H_0, Ω_M and the dark energy equation of state [at least as described by the w_0 or (w_0, w_a) parameters], is to run a MCMC where the above catalog of mock GW detections is used in conjunction with CMB, BAO and SNa data. This has been performed in [203], where it was found that, fitting the simulated data with the parametrization (3.62) (where we set for definiteness $n \simeq 2.5$), with the above 15 mock detections the HLVKI network would determine Ξ_0 to an accuracy $\Delta\Xi_0 \simeq 0.125$. From eq. (3.65), this implies $\Delta\delta(0) \simeq 0.31$, to be compared with the value $\Delta\delta(0) \simeq 0.24$ found by combining the error of all 15 mock measurements of $\delta(0)$ in table 6 according to $1/\sigma^2 = \sum_i 1/\sigma_i^2$. Notice that, while the MCMC takes into account more accurately the degeneracies with the other cosmological parameters, which is what eventually leads to a slightly larger estimates of $\Delta\delta(0)$, its chains converge only with a sufficiently large set of mock GW events (which is the reason why in [203] was used a catalog corresponding to 10 yr of data taking). In contrast, the simpler estimate of $\Delta\delta(0)$ presented in table 6 gives an idea of the contribution of individual detections, as a function of redshift.

As already pointed out in [115], the above results also have potentially important implications for the search of the electromagnetic counterpart to a GW detection, since they imply that the actual electromagnetic luminosity distance of the source, and hence its redshift, will be different from that inferred from the GW detection assuming GR. Indeed, if the correct theory is ΛCDM, given a best-fit value value D of the luminosity distance to the source measured with GWs, the corresponding best-fit value of the redshift $z_{\Lambda\text{CDM}}$ is predicted to be given by

$$d_L^{\text{em}}(z_{\Lambda\text{CDM}}) = D \,. \tag{3.82}$$

In contrast, if the correct description of Nature is given by the RT model, the best-fit value for redshift of the source, z_{RT}, is given by

$$d_L^{\text{gw}}(z_{\text{RT}}) = D \,. \tag{3.83}$$

Combining these relations we can determine the function $z_{\text{RT}}(z_{\Lambda\text{CDM}})$. In the left panel of figure 20 we show $\Delta z \equiv z_{\text{RT}} - z_{\Lambda\text{CDM}}$ as a function of $z \equiv z_{\Lambda\text{CDM}}$, for the minimal RT model and for RT with $\Delta N = 34, 50, 64, 100$. The right panel shows the result up to $z = 3$, which is relevant for BNS at the Einstein Telescope (see section 3.6.2) and, up to $z \simeq 1$, for NS-BH binaries at the HLVKI network. The left panel provides an enlargement of the region up to $z = 0.2$, which is the range relevant for BNS at 2G detectors. At $z < 0.2$, for individual detections the difference Δz in the theoretical prediction of the redshift will be of the order of the error box induced by the observational error on d_L^{gw}, as it is clear from the fact that the observational errors on $\delta(0)$ in table 6 are of the order of the prediction of the RT model with large ΔN. For larger values of z, the difference can, however, become very significant. This plot is another way to present the prediction of the RT model, complementary to figure 15. It's relevance is particularly clear for the search of the counterpart with telescopes. For instance, from the right panel of figure 20 we see that, for a GW event for which ΛCDM would predict a redshift $z = 3$, the RT model with very large ΔN predicts that telescopes should rather search for the counterpart by targeting galaxies at $z \simeq 2$.
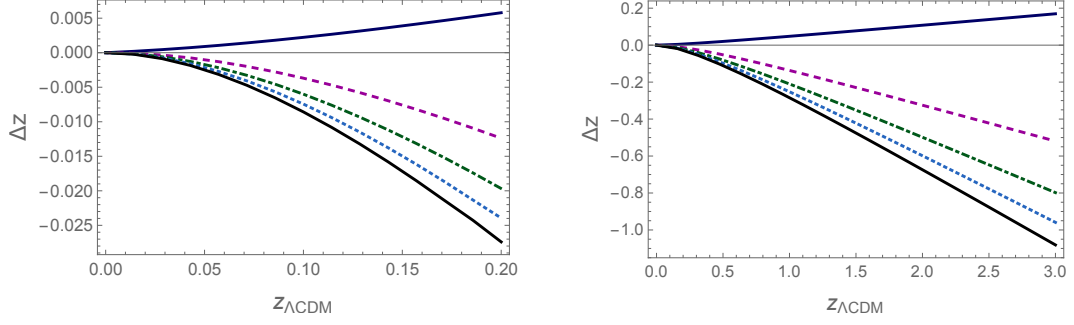
Figure 20. The change in the actual redshift of the source $\Delta z \equiv z_{\rm RT} - z_{\Lambda \rm CDM}$, compared to the ΛCDM prediction $z_{\Lambda \rm CDM}$, as a function of $z \equiv z_{\Lambda \rm CDM}$, for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed), $\Delta N = 64$ (cyan, dotted) and $\Delta N = 100$ (black solid line). Left panel: in the range $z < 0.2$, relevant for BNS at 2G detectors. Right panel: up to $z = 3$.
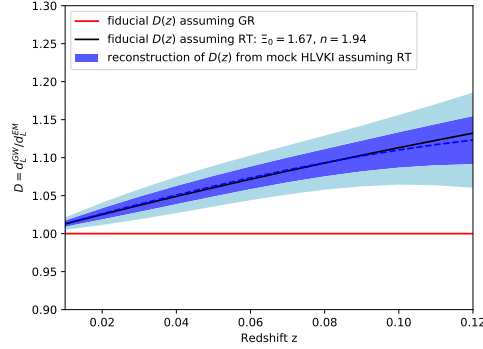
Figure 21. Gaussian process reconstruction of $d_L^{\rm gw}/d_L^{\rm em}$, using the RT model with $\Delta N = 64$ as fiducial cosmology, using GW events at the HLVKI network with GR counterpart for $d_L^{\rm gw}$, and DES supernovae for $d_L^{\rm em}$. The blue and light blue regions correspond to 68% and 95% confidence levels, respectively. From [204].

Finally, it is also interesting to see how well one can reconstruct the ratio $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ from the data, without assuming any parametrization for it, such as eq. (3.62). This can be done using the technique of gaussian processes, that allows the reconstruction of a function directly from the data. Several applications of gaussian processes in cosmology have been discussed in [179, 218–226]. In [204] this technique has been applied to the reconstruction of $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$. For $d_L^{\rm gw}$ has been used the same catalog of mock joint GW-GRB detections shown in table 6, where the GW events are detected at the HLVKI network and the GRB counterparts are observed by Fermi-GBM and Swift. For $d_L^{\rm em}(z)$ were considered simulated measurements from DES supernovae, with the data generated as in [223], with a redshift range $0.05 < z < 1.2$, and the errors on $d_L^{\rm em}$ estimated as in [227]. Figure 21 shows the result of the reconstruction; we see that, with these datasets, the prediction of the RT model with $\Delta N = 64$ (used as fiducial in the figure) is very clearly distinguished from the prediction $d_L^{\rm gw}(z)/d_L^{\rm em}(z) = 1$ of ΛCDM.

3.6.2 3G detectors: Einstein Telescope and Cosmic Explorer

We next consider third-generation (3G) ground-based interferometers currently under study, such as the Einstein Telescope (ET) in Europe [228] and Cosmic Explorer (CE) in the US [229,

230], that could start to be operative in the mid 2030s. These detectors will have the potential of exploring the Universe with GWs to truly cosmological distances, guaranteeing an extraordinary output in astrophysics, cosmology and fundamental physics [231, 232].

For instance ET, even as a single detector, with be able to detect the coalescence of compact binaries with total mass $(20 - 100) \, M_\odot$, as typical of BH-BH or BH-NS binaries, up to redshift $z \sim 20$ and higher. By comparison, in the catalog of detections from the O1 and O2 Advanced LIGO/Virgo runs, the farthest BH-BH event is at $z \simeq 0.5$ and, at final target sensitivity, 2G detectors should reach $z \simeq 1$. The corresponding rates will be of order 10^6 events per year. For binary neutron stars, ET will detect them out to $z \simeq 2 - 3$, which allows us to reach the peak of the star formation rate and therefore detect the vast majority of coalescing BNS throughout the Universe; by comparison, at final target sensitivity, 2G detectors should reach $z \simeq 0.2$. The expected rate of BNS at ET was computed in [203] using state-of-the art models for the formation and evolution of neutron star binaries, and is found to be between 6.2×10^4 and 6.9×10^4 events per year, having assumed a duty cycle of 80%. This corresponds to $(0.8 - 0.9) \times 10^5$ events in one year of actual data.[42] In order to use these BNS as standard sirens one either needs an electromagnetic counterpart, or one must use statistical methods. Here we focus on BNS with electromagnetic counterparts. We consider for definiteness mock catalog for ET, but similar estimates hold for CE.[43]

Refs. [203, 233] have estimated estimated the expected number and the redshift distribution of coincidences between GW events at ET and the electromagnetic signal observed at a GRB detector with the characteristics of the proposed THESEUS mission [234, 235], that could be in operation at the same time as 3G detectors. Depending on the assumptions made, the estimated number of joint GW-GRB detections is between $O(15)$ and $O(50)$ per year. In table 7 (from [203]) we show some properties of a sample catalog, obtained assuming 10 yr of data taking. More counterparts could be obtained from future large telescopes that will be able to monitor large regions of the sky from the radio, optical to the X-ray (see [232, 236] for discussion), although realistic estimates are difficult to obtain because they also depend on issues such as the prioritization that will be given to the follow-up of GW signals.

We can now estimate the accuracy that can be obtained on modified GW propagation from individual events such as those in table 7. In the lowest redshift bin, say $z \lesssim (0.1 - 0.2)$, we are in a situation similar to that studied above for 2G detectors, and we can use $\delta(0)$ as observable. However, now $\Delta d_L^{\rm gw}/d_L^{\rm gw}$ is below 1%, and one can also easily imagine that, by the time ET will be operative, the accuracy on H_0 will have further improved, so we assume $\Delta d_L^{\rm em}/d_L^{\rm em} = 0.5\%$. We then find that each single event with counterpart at, say, $z = 0.2$, will allow us to measure $\delta(0)$ to an accuracy of about $0.15 - 0.20$. Comparing with the predictions in table 5, we see that just one event will be sufficient to detect at 5σ the predictions of the RT model with large ΔN!

For sources at not too small redshift, we must rather use the full expression (3.62). According to table 5, we fix $n = 2$ (the precise value has limited importance for the analysis)

---

[42]Previous estimates for BNS [172] were slightly higher, $\mathcal{O}(10^5 - 10^6)$ BNS/yr. This is partly due to the fact that in [203] has been used a threshold of 12 for the network SNR, obtained by combining the three arms of ET, while previous work typically used a threshold of 8.

[43]Cosmic Explorer can reach a much greater distance for BNS, up to $z \simeq 8$. However, since the peak of the star formation rate is at $z \sim 2 - 3$, most of the coalescing BNS will be seen already at the distances accessible to ET. Furthermore, beyond $z \sim 1.5 - 2$ it will be very difficult to detect an electromagnetic counterpart even with a GRB. Thus, for BNS with counterpart the estimates for CE will be basically the same as for ET. See also [203] for the prediction of BNS rates in a network with two CE and one ET detector.

redshift bin	number of joint GW-GRB events	mean redshift	mean $\Delta d_L^{\rm gw}/d_L^{\rm gw}$	$\Delta\Xi_0$
$(0 \, , \, 0.1)$	4	0.07108	0.00868	0.11
$(0.1 \, , \, 0.2)$	24	0.15001	0.01784	0.09
$(0.2 \, , \, 0.3)$	24	0.24043	0.02558	0.09
$(0.3 \, , \, 0.4)$	27	0.35355	0.03529	0.09
$(0.4 \, , \, 0.5)$	28	0.44966	0.04843	0.10
$(0.5 \, , \, 0.6)$	9	0.53785	0.05646	0.10
$(0.6 \, , \, 0.7)$	14	0.64540	0.05329	0.09
$(0.7 \, , \, 0.8)$	13	0.73793	0.05493	0.09
$(0.8 \, , \, 0.9)$	8	0.85497	0.06413	0.10
$(0.9 \, , \, 1.0)$	4	0.93702	0.06257	0.09
$(1.0 \, , \, 1.1)$	6	1.05334	0.06494	0.09
$(1.1 \, , \, 1.2)$	3	1.15162	0.06749	0.09
$(1.2 \, , \, 1.3)$	1	1.25943	0.07373	0.10
$(1.3 \, , \, 1.4)$	—	–	—	—
$(1.4 \, , \, 1.5)$	2	1.45375	0.07851	0.10
$(1.5 \, , \, 1.6)$	1	1.58407	0.07577	0.09
$(1.6 \, , \, 1.7)$	1	1.62843	0.07947	0.10

Table 7. Number of event and mean value of the observational error $\Delta d_L^{\rm gw}/d_L^{\rm gw}$ in different redshift bins, for a specific realization of the catalog of joint GW-GRB detections, assuming 10 yr of data (from [203]). In the last column we give an estimate of the error on Ξ_0 from an individual source in the given frequency bin.

and keep only Ξ_0 as the parameter labeling the predictions. From eq. (3.62),

$$\frac{\Delta d_L^{\rm gw}}{d_L^{\rm gw}} + \frac{\Delta d_L^{\rm em}}{d_L^{\rm em}} = \Delta\Xi_0 \left(1 - \frac{1}{(1+z)^n} \right) . \tag{3.84}$$

Setting for definiteness $\Delta d_L^{\rm em}/d_L^{\rm em} = 0.5\%$, for a single source with a redshift given by the third column in table 7 and a value of $\Delta d_L^{\rm gw}/d_L^{\rm gw}$ as in the fourth column, we get the accuracy $\Delta\Xi_0$ given in the last column. We see that, independently of redshift, each individual detection would provide a measurement of Ξ_0 at the $(9-10)\%$ level. Thus, the predictions for Ξ_0 given in table 5, that for large ΔN differ by the GR result by as much as 80%, could be tested at more than 5σ with just a single joint GW-GRB detection. Combining the errors on Ξ_0 from each of the 169 events in table 5 (taking into account the number of events per bin) we get an overall error $\Delta\Xi_0 \simeq 0.7\%$. This agrees with the result obtained in [203], using the full catalogs corresponding to 10 yr of data, and performing a MCMC to take into account more precisely the degeneracies with H_0, Ω_M and w_0 (and using, conservatively, the current datasets on CMB, BAO and SNe), where it was found that Ξ_0 can be determined to an accuracy of 1%. Clearly, with respect to the size of the deviation of Ξ_0 from the GR value, which can be as large as 80%, this is a remarkable precision.

Figure 22 shows the result of a gaussian process reconstruction of $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ using mock ET and DES catalogs, using as fiducial the RT model with $\Delta N = 64$. We see that,
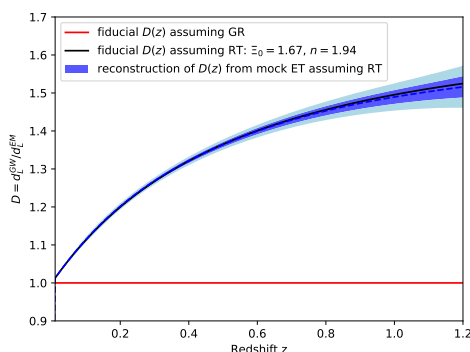
– 64 –

Figure 22. As in figure 21, using the mock ET and DES catalogs for the RT fiducial cosmology. From [204].

for a deviation from GR of this size, $d_L^{\rm gw}(z)/d_L^{\rm em}(z)$ would be reconstructed with exquisite precision, allowing not only to prove GR wrong on cosmological scales, but also to pinpoint with accuracy the properties of the alternative model, in this case the parameter ΔN that characterizes the RT model.

3.6.3 LISA

A study of the accuracy to Ξ_0 that could be obtained with the space interferometer LISA has been presented in [191]. In this case the coalescences of supermassive BH binaries plays the role of standard sirens, since they are believed to merge in a gas rich environment that is expected to power electromagnetic emission, resulting in a detectable electromagnetic counterpart. The corresponding mock catalogs were generated by using advanced models for galaxy formation and merger, and different scenarios for the seeds of the massive black holes and for the delays between galaxy merger and massive black hole merger, resulting in three main scenarios (heavy seeds and no delay, heavy seeds with a specific prescription for the delay, and light seeds due to pop III stars). For each scenario were simulated 22 catalogs corresponding to the nominal 4 yr of the LISA mission. For each catalog was then performed a quick Fisher matrix cosmological analysis assuming ΛCDM, and was then selected, for each scenario, the median catalog among all ranked 22 catalogs as the representative catalog for the corresponding astrophysical model. The corresponding catalogs had 32, 12 and 9 sources for heavy seeds and no delay, heavy seeds with delay, and light seeds, respectively. On these catalogs, were then run full MCMC to constrain Ξ_0, including again CMB, BAO and SNe to reduce the degeneracies with Ω_M, H_0 and w_0. Furthermore, two different assumptions on the accuracy of the estimate of the source redshift were considered, and denoted as 'optimistic' and 'realistic', respectively. The resulting estimate was $\Delta\Xi_0 = (1-2)\%$ with the optimistic assumption on the estimate of the source redshift, and $\Delta\Xi_0 = (2-4)\%$ with the realistic assumption. In all cases, we are again in a situation where the predictions of the RT model given in table 5, that for large ΔN differ from GR between 30% and 80%, are very clearly detected.

While the contribution of a single source cannot be obtained from a MCMC (where a large number of sources is obtained to get the convergence of the chains), it can be estimated by comparing the results from catalogs with a different number of sources. For the realistic scenario for the estimate of the source redshift, from table 2 of [191] we see that LISA
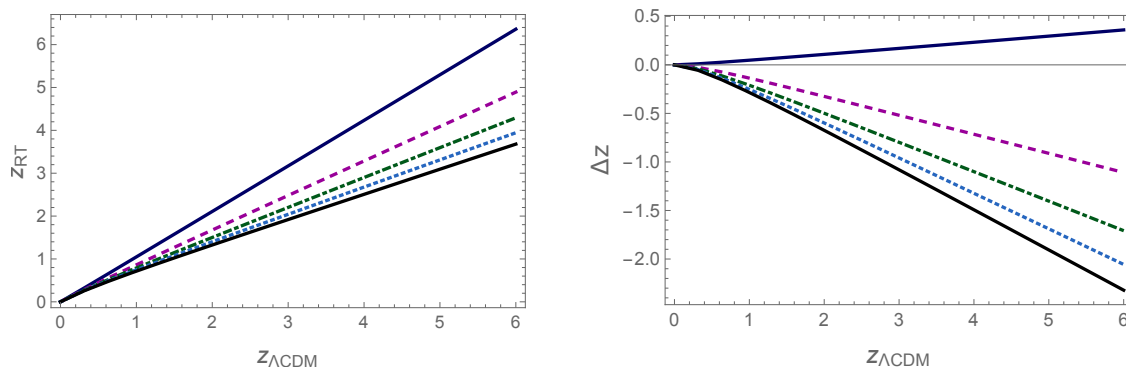
Figure 23. z_{RT} as a function of $z_{\Lambda\text{CDM}}$ (left panel) and the difference $\Delta z = z_{RT} - z_{\Lambda\text{CDM}}$, as a function of $z_{\Lambda\text{CDM}}$ (right panel) for the minimal RT model (blue solid line) and for RT with $\Delta N = 34$ (magenta, dashed), $\Delta N = 50$ (green, dot-dashed), $\Delta N = 64$ (cyan, dotted) and $\Delta N = 100$ (black solid line).

could measure Ξ_0 with an error $\Delta\Xi_0 \simeq \{0.023, 0.036, 0.044\}$ in the three catalogs containing, respectively, $N = \{32, 12, 9\}$ events. These numbers are well reproduced by [115]

$$\Delta\Xi_0 \simeq 0.13/\sqrt{N}, \tag{3.85}$$

so each SMBH event gives a measure of Ξ_0 with an average accuracy of about 13%. Using the optimistic scenario for redshift determination we rather get $\Delta\Xi_0 \simeq 0.06/\sqrt{N}$. Thus, even a single SMBH event at LISA could be sufficient to detect the effect predicted by the RT model with large ΔN.

As already pointed out in the discussion after eq. (3.83), another aspect of the effect is that, if the RT model provides the correct description of Nature, the actual redshift of the source, as determined through electromagnetic observation, would turn out to be very different from that inferred from a measurement of d_L^{gw} and interpreted using ΛCDM. For LISA this effect is particularly remarkable since, as we see from figure 12 of [191], the mock catalogs of supermassive BH binary coalescences observed at LISA include events up to $z \simeq 6$ and higher. This effect is shown in figure 23, where the left panel shows that redshift of the source predicted by the RT model as a function of the redshift predicted by ΛCDM, and the right panel shows the difference $\Delta z = z_{RT} - z_{\Lambda\text{CDM}}$, as a function of $z_{\Lambda\text{CDM}}$ (i.e. the same as figure 20, but on a range of redshifts appropriate to supermassive BH binaries at LISA). For instance, for a given measurement of the luminosity distance through GWs for which ΛCDM would predict, say, a redshift $z = 6$, the RT model with very large ΔN rather predicts that the source will be found, by electromagnetic observations, at $z \simeq 3.7$, a rather striking difference.

4 Conclusions

We have discussed in detail a modification of gravity on cosmological scales, summarizing and extending previous work by our group. The model is based on a clear and well-defined theoretical framework. Rather than introducing extra degrees of freedom, such as extra scalar, vector or tensor fields, or extra polarization for the graviton, as in typical modified gravity models, the basic idea is that long-distance modifications to the dynamics of gravity are induced by infrared quantum effects in GR itself. This means that the proper tool is no

longer the action of the theory but the corresponding quantum effective action. In quantum field theory, at a fundamental level, actions are local functionals of the fields; however, whenever the theory contains massless particles (such as the graviton in GR), or particles that are light with respect to the relevant energy scale, the corresponding quantum effective action also contains non-local terms. We have seen how, with non-local terms, we can construct a mass term for a gauge field without violating gauge invariance, and mass terms for the gravitational field that do not violate diffeomorphism invariance. Linearizing the theory over flat space, we have seen that two independent mass terms can be constructed: one for the conformal mode, and one for the transverse-traceless mode $h_{ij}^{\rm TT}$, see eq. (2.32). Our basic assumption is that the conformal mode indeed becomes massive, while $h_{ij}^{\rm TT}$ remains massless.

The idea that the conformal mode becomes massive (while $h_{ij}^{\rm TT}$ stays massless) currently has the status of a conjecture, which is difficult to verify from first principles since it involves non-perturbative physics. Nevertheless, we have seen that numerical results from Euclidean quantum gravity on the lattice and from causal dynamical triangulations, as well as analytic computations using functional renormalization group equations, give some support for the hypothesis of a dynamical mass generation. The fact that the resulting mass scale is indeed associated to the conformal mode is also suggested again by numerical results from causal dynamical triangulations, and by several arguments that show that the conformal mode is the most problematic one in the infrared. Once one assumes the validity of this conjecture, leading to the linearized action (2.33), the covariantization of this linearized nonlocal theory leads quite naturally to two different possibilities, that we have called the RT and RR models. We have seen that eventually the RR model is ruled out phenomenologically, while the RT model (2.45) has been the main focus of our paper.

We have then explored in details the observational consequences of the RT model. Constructing a model that fulfills all observational constraints and gives predictions testable in the near future is in general very difficult, as has been learned from the explicit study of several modified gravity models. Here, one should also appreciate that, once accepted the underlying assumptions spelled out above, the theory is basically fixed (apart from the two options given by the RR and RT model), and has the same number of parameters as ΛCDM, with a new mass scale m replacing the cosmological constant (plus, as we have seen for the RT model, a single constant ΔN which reflects all our ignorance on initial conditions). Thus, the theory either is consistent with observations or it doesn't. We do not have the freedom of playing with arbitrary functions, as for instance in scalar-tensor theories of the Horndeski type.

At the phenomenological level, the RT model turns out to have a number of remarkable properties:

- Its cosmological solutions, at the background level, show an accelerated expansion at the present cosmological epoch, without the need for a cosmological constant. In other words, giving a mass to the conformal mode provides a possible explanation for the observed accelerated expansion of the Universe and for the origin of dark energy.

- The fact that dark energy starts to dominate just at the present cosmological epoch is obtained by choosing a value for the mass scale m, or, more precisely, for the scale $\Lambda_{\rm RT} \sim (m_{\rm Pl} m)^{1/2}$, which is the fundamental scale that is generated dynamically. In this sense, the model does not solve the coincidence problem. However, even if $\Lambda_{\rm RT}$ cannot be predicted (just as we cannot predict the value of $\Lambda_{\rm QCD}$ in strong interactions), the required numerical value, of the order of the meV, is not particularly surprising from

the point of view of quantum field theory (contrary, e.g., to theories that introduces a fundamental mass scale of order $H_0 \sim 10^{-33}\,\mathrm{eV}$). Furthermore, such a dynamically generated mass scale is a renormalization group invariant, so there is no problem of technical naturalness.

- Scalar perturbations over the FRW background are stable and remain small during the whole cosmological evolution. This is already a non-trivial property, that has ruled out many modified gravity models. Furthermore, the scalar perturbations of the RT model are very close to those of ΛCDM, which in the end allows the model to fit current cosmological data well, while still being potentially distinguishable with future missions.

- A full MCMC analysis shows that the model (for all values of ΔN) fits CMB, BAO, SNe, measurements of $H(z)$ and structure formation data at the same level as ΛCDM.

- The model reduces to GR at small scales, without the need of invoking non-linear screening mechanisms, and therefore passes all the constraints from solar system and laboratory experiments. It furthermore complies with the limit on the time variation of the Newton's constant from Lunar Laser Ranging. As we have seen, this is in general non-trivial even when the static solution has the correct GR limit (and, indeed, it is this bound that rules out the RR model).

- The sector of cosmological tensor perturbations (i.e. GWs propagating over a FRW background) provides a great surprise. In the RT model GW propagation across cosmological distances is different from GR, so that the relation between the luminosity distance extracted from a coalescing binary and the redshift is modified, giving rise to the notion of 'GW luminosity distance' $d_L^{\mathrm{gw}}(z)$. This has been found to be common to all modified gravity models. What is remarkable for the RT model (in particular for large ΔN) is the size of the effect, that, at the redshifts accessible to future GW detectors such as third-generation ground based detectors such as Einstein Telescope and Cosmic Explorer, or the space interferometer LISA, could lead to deviations from GR as large as 80%. We have seen that, with these detectors, even the detection of a single standard siren with electromagnetic counterpart would be sufficient to detect the effect at more than 5σ. The effect is smaller at the redshifts accessible to the second-generation network made by advanced LIGO/Virgo and KAGRA, but still could be potentially within reach even at these detectors, over several years of data taking.

Acknowledgments

A Difficulties of alternative nonlocal models

A natural question is whether it is possible to construct other nonlocal models that share the good phenomenological properties of the RT model. We will see that, in fact, this is very difficult, and this will allow us to better appreciate the results presented above. To organize

the discussion, it can be useful to follow the path that actually lead to the formulation of the RT model and of other variants, and see what conditions eliminated the various alternatives (see also [22]).

A first nonlocal model associated to a mass scale was proposed, on purely phenomenological grounds, by Arkani-Hamed, Dimopoulos, Dvali, and Gabadadze [237] and consisted in modifying the Einstein equations into

$$\left(1 - \frac{m^2}{\Box}\right) G_{\mu\nu} = 8\pi G\, T_{\mu\nu}\,, \tag{A.1}$$

where m is the new mass scale.[44] This model was proposed to introduce the degravitation idea, namely the idea the vacuum energy density, even if present, does not gravitate. In fact, at least performing naively the inversion of the nonlocal operator, eq. (A.1) can be rewritten as $G_{\mu\nu} = 8\pi G\, (\Box - m^2)^{-1} \Box T_{\mu\nu}$. Therefore the low-momentum modes of $T_{\mu\nu}$ are filtered out and in particular a term in $T_{\mu\nu}$ due to a cosmological constant does not contribute.

However, a drawback of eq. (A.1) is that the energy-momentum tensor is not automatically conserved, since in curved space $\boldsymbol{\nabla}^\mu$ does not commute with $\Box$ and therefore with $\Box^{-1}$. As a consequence, the Bianchi identity $\boldsymbol{\nabla}^\mu G_{\mu\nu} = 0$ no longer ensures $\boldsymbol{\nabla}^\mu T_{\mu\nu} = 0$. In [9] it was then observed that it is possible to cure this problem by making use of the decomposition (2.41) to extract the transverse part of the tensor $\Box^{-1} G_{\mu\nu}$. One can then modify eq. (A.1) into

$$G_{\mu\nu} - m^2 \left(\Box^{-1} G_{\mu\nu}\right)^{\mathrm{T}} = 8\pi G\, T_{\mu\nu}\,, \tag{A.2}$$

and energy-momentum conservation becomes automatic. However, the cosmological evolution of this model turned out to be unstable, already at the background level [20, 84], so this model is not phenomenologically viable. This instability is due to the fact that, once written the model in local form introducing some auxiliary fields, the latter have unstable modes already during RD and MD (see in particular appendix A of [84]). Then, any small deviation from the standard FRW solution will quickly be amplified and lead to a completely different evolution, inconsistent with the observations. It was then realized in [20] that this instability is related to the action of the $\Box^{-1}$ operator on a tensor such as $G_{\mu\nu}$ or $R_{\mu\nu}$, and is absent if it acts on a scalar such as R. This led to the RT model (2.45). Trying to work out a similar model at the level of the quantum effective action, rather than of the equations of motion, led to the RR model (2.46) [21], which, at least at the level of cosmology, shared all good properties of the RT model. Its cosmological solutions were studied in detail in [21–23] (see also [238] for a different branch of solutions).

A natural generalization of the RR model is given by the quantum effective action

$$\Gamma = \frac{m_{\mathrm{Pl}}^2}{2} \int d^4x \sqrt{-g}\, \left[R - \mu_1 R \frac{1}{\Box^2} R - \mu_2 C^{\mu\nu\rho\sigma} \frac{1}{\Box^2} C_{\mu\nu\rho\sigma} - \mu_3 R^{\mu\nu} \frac{1}{\Box^2} R_{\mu\nu} \right]\,, \tag{A.3}$$

where μ_1, μ_2 and μ_3 are parameters with dimension of $(\mathrm{mass})^2$, and $C_{\mu\nu\rho\sigma}$ is the Weyl tensor. This extension was studied in [239], where it was found that the term $R^{\mu\nu}\Box^{-2}R_{\mu\nu}$ is ruled out since it gives instabilities in the cosmological evolution at the background level, again due to the behavior of the auxiliary fields. The Weyl-square term instead does not contribute

---

[44]Actually, in [237] the model was presented as a modification of GR that is acausal on cosmological scales. As we have discussed in section 2.3.3, causality is, however, preserved once this is understood as the equation of motion derived from a quantum effective action for the in-in vacuum expectation value of the metric, which automatically ensures that the Green's function in the $\Box^{-1}$ operator is the retarded one.

to the background evolution, since the Weyl tensor vanishes in FRW, and it also has well-behaved scalar perturbations. However, quite interestingly, it was ruled out by the fact that its tensor perturbations are unstable, showing the the stability of perturbations is in general a non-trivial requirement both in the scalar and in the tensor sector.

The realization that both models that survived, RT and RR, had the physical meaning of a mass for the conformal mode [81] then suggested to focus the attention on models with such a meaning. One interesting variant of the RR model is given by

$$\Gamma_{\Delta_4} = \frac{m_{\rm Pl}^2}{2} \int d^4x \sqrt{-g} \left[R - \frac{m^2}{6} R \frac{1}{\Delta_4} R \right] . \tag{A.4}$$

where Δ_4 is the Paneitz operator,

$$\Delta_4 \equiv \Box^2 + 2R^{\mu\nu}\boldsymbol{\nabla}_\mu\boldsymbol{\nabla}_\nu - \frac{2}{3}R\Box + \frac{1}{3}g^{\mu\nu}\boldsymbol{\nabla}_\mu R\boldsymbol{\nabla}_\nu , \tag{A.5}$$

and whose linearization over Minkowsli space is the same as the RT or RR models. This is the operator that enters in the conformal anomaly in four dimensions and, from the point of view of conformal invariance, is the natural generalization of the d'Alembertian from two to four dimensions. This model had a viable cosmological evolution [240], although its prediction for the equation of state of DE, $w_0 \simeq -1.31$, already seemed off with respect to the observations, as was indeed confirmed from a MCMC analysis in [23]. In any case, what definitely ruled out the model was the realization that, in the tensor sector, it predicts a speed of GWs different from the speed of light [23]. These examples show that the requirements of having a viable background evolution, stable scalar perturbations, good fit to the cosmological observations, stable tensor perturbations, and $c_{\rm gw} = c$, all provide non-trivial tests, potentially able to rule out a model.

Finally, as we mentioned in section 3.4.2, limits on the time variation of Newton's constant ruled out also the RR model [157, 161]. The detailed analysis in [161] is very general, and makes it clear that the same situation will happen in any nonlocal model in which the effective Newton's constant at short scales depends on the auxiliary fields of the theory. This therefore applies also to models of the form

$$\Gamma = \frac{m_{\rm Pl}^2}{2} \int d^4x\sqrt{-g} \left[R - \left(\frac{m^2}{\Box}\right)^n R \right] , \tag{A.6}$$

corresponding to a running of Newton's constant, that have been studied, at the level of background evolution, in [241] for $n = 1$ and in [242] for $n = 2$, where it was found that their cosmological evolution appears to be in principle viable, at least at the background level (in particular the model with $n = 1$ has an evolution very close to that of the RR model, up to the present epoch). In contrast, we have seen that the RT model passes the LLR limit because at short scales $G_{\rm eff}$ loses any dependence on the auxiliary fields and reduces to G, see eq. (3.28).

Last but not least, a different but related line of research is given by non-local models that are not associated to a mass scale, and whose development predated that of the non-local models associated to a mass scale on which we have focused. The underlying physical motivation is again that IR divergences could generate, through non-perturbative effects, the relevant nonlocal terms in the quantum effective action. The first nonlocal gravity model of this type was proposed by Wetterich [243], and was based on the quantum effective action,

$$\Gamma = \frac{m_{\rm Pl}^2}{2} \int d^4x\sqrt{-g} \left[R - \lambda R\Box^{-1}R \right] . \tag{A.7}$$

Since $\Box^{-1}R$ is dimensionless, the associated constant λ is also dimensionless. The model, however, did not produce a viable cosmological evolution [243]. Deser and Woodard [25, 27] (see [244] for review) considered a more general nonlocal model of the form

$$\Gamma_{\text{DW}} = \frac{m_{\text{Pl}}^2}{2} \int d^4x \sqrt{-g} \left[R - Rf(\Box^{-1}R) \right] , \tag{A.8}$$

with $f(X)$ a dimensionless function, which was tuned so to obtain the desired background evolution. Requiring that the cosmological evolution closely mimics that of ΛCDM leads to a function well-fitted by [245]

$$f(X) = 0.245 \left[\tanh \left(0.0350(X + 16.5) + 0.032(X + 16.5)^2 + 0.003(X + 16.5)^3 \right) - 1 \right] , \tag{A.9}$$

for $X \equiv \Box^{-1}R < 0$. To comply with solar system constraints, the proponents of the model set $f(X) = 0$ for $X > 0$. The argument suggested in refs. [27, 244] for this choice was that in a cosmological setting (where the time derivative dominates) $X \simeq (-\partial_t^2)^{-1}R$ is negative because of the minus sign in $-\partial_t^2$, while it is positive in the regime dominated by structure formation, where the spatial derivatives dominate and $X \simeq (\boldsymbol{\nabla}^2)^{-1}R$. It was however shown in [161] that this is not correct, and X is always negative, even in a static situation. Thus, the Deser-Woodard model lacks a screening mechanism and is ruled out by the comparison with observations at the solar system scale. In [246] then Deser and Woodard proposed a variant of their model constructed with a function $f(Y)$ of the variable $Y = \Box^{-1}g^{\mu\nu}\partial_\mu X \partial_\nu X$, where again $X = \Box^{-1}R$, which indeed changes sign between static and time-dependent solutions, and again postulated that $f(Y) = 0$ for $Y < 0$. The consequences of the model have not been throughly investigated, in particular stability of the solutions, etc.; however, apart from a certain convolutedness of the model, one can anticipate potential problems with LLR similar to those that ruled out the RR model.

References

[1] A. Strominger, *Lectures on the infrared structure of gravity and gauge theory*, arXiv:1703.05448 [INSPIRE].

[2] I. Antoniadis and E. Mottola, *Graviton fluctuations in de Sitter space*, *J. Math. Phys.* **32** (1991) 1037 [INSPIRE].

[3] N.C. Tsamis and R.P. Woodard, *Strong infrared effects in quantum gravity*, *Ann. Phys.* **238** (1995) 1.

[4] I. Antoniadis and E. Mottola, *4D quantum gravity in the conformal sector*, *Phys. Rev.* **D 45** (1992) 2013 [INSPIRE].

[5] I. Antoniadis, P.O. Mazur and E. Mottola, *Cosmological dark energy: prospects for a dynamical theory*, *New J. Phys.* **9** (2007) 11 [gr-qc/0612068] [INSPIRE].

[6] T.R. Taylor and G. Veneziano, *Quantum gravity at large distances and the cosmological constant*, *Nucl. Phys.* **B 345** (1990) 210 [INSPIRE].

[7] G. Dvali, *Predictive power of strong coupling in theories with large distance modified gravity*, *New J. Phys.* **8** (2006) 326 [hep-th/0610013] [INSPIRE].

[8] M. Porrati, *Higgs phenomenon for 4D gravity in Anti-de Sitter space*, *JHEP* **04** (2002) 058 [hep-th/0112166] [INSPIRE].

[9] M. Jaccard, M. Maggiore and E. Mitsou, *Nonlocal theory of massive gravity*, *Phys. Rev.* **D 88** (2013) 044033 [arXiv:1305.3034] [INSPIRE].

[10] P. Boucaud et al., *Testing Landau gauge OPE on the lattice with a < A^2 > condensate*, *Phys. Rev.* **D 63** (2001) 114003 [`hep-ph/0101302`] [InSPIRE].

[11] M.A.L. Capri et al., *A Study of the gauge invariant, nonlocal mass operator Tr integer $d^4 \times F_{\mu\nu}(D^2)^{-1}F_{\mu\nu}$ in Yang-Mills theories*, *Phys. Rev.* **D 72** (2005) 105016 [`hep-th/0510240`] [InSPIRE].

[12] D. Dudal et al., *A Refinement of the Gribov-Zwanziger approach in the Landau gauge: infrared propagators in harmony with the lattice results*, *Phys. Rev.* **D 78** (2008) 065047 [`arXiv:0806.4348`] [InSPIRE].

[13] S. Foffa, M. Maggiore and E. Mitsou, *Apparent ghosts and spurious degrees of freedom in non-local theories*, *Phys. Lett.* **B 733** (2014) 76 [`arXiv:1311.3421`] [InSPIRE].

[14] C.W. Misner, K. Thorne and J. Wheeler, *Gravitation*, Freeman, New York, U.S.A. (1973).

[15] E.E. Flanagan and S.A. Hughes, *The basics of gravitational wave theory*, *New J. Phys.* **7** (2005) 204 [`gr-qc/0501041`] [InSPIRE].

[16] M. Jaccard, M. Maggiore and E. Mitsou, *Bardeen variables and hidden gauge symmetries in linearized massive gravity*, *Phys. Rev.* **D 87** (2013) 044017 [`arXiv:1211.1562`] [InSPIRE].

[17] M. Maggiore, *Gravitational waves. Volume 2: astrophysics and cosmology*, Oxford University Press, Oxford U.K. (2018).

[18] S. Deser, *Covariant decomposition and the gravitational Cauchy problem*, *Ann. Inst. Henri Poincare* **7** (1967) 149.

[19] J.J. York, *Covariant decompositions of symmetric tensors in the theory of gravitation,"* *Ann.Inst. Henri Poincare* **21** (1974) 319.

[20] M. Maggiore, *Phantom dark energy from nonlocal infrared modifications of general relativity*, *Phys. Rev.* **D 89** (2014) 043008 [`arXiv:1307.3898`] [InSPIRE].

[21] M. Maggiore and M. Mancarella, *Nonlocal gravity and dark energy*, *Phys. Rev.* **D 90** (2014) 023005 [`arXiv:1402.0448`] [InSPIRE].

[22] M. Maggiore, *Nonlocal infrared modifications of gravity. A review*, *Fundam. Theor. Phys.* **187** (2017) 221 [`arXiv:1606.08784`].

[23] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Nonlocal gravity. Conceptual aspects and cosmological predictions*, *JCAP* **03** (2018) 002 [`arXiv:1712.07066`] [InSPIRE].

[24] N.C. Tsamis and R.P. Woodard, *Nonperturbative models for the quantum gravitational back reaction on inflation*, *Annals Phys.* **267** (1998) 145 [`hep-ph/9712331`] [InSPIRE].

[25] S. Deser and R.P. Woodard, *Nonlocal cosmology*, *Phys. Rev. Lett.* **99** (2007) 111301 [`arXiv:0706.2151`] [InSPIRE].

[26] A.O. Barvinsky, *Serendipitous discoveries in nonlocal gravity theory*, *Phys. Rev.* **D 85** (2012) 104018 [`arXiv:1112.4340`] [InSPIRE].

[27] S. Deser and R.P. Woodard, *Observational viability and stability of nonlocal cosmology*, *JCAP* **11** (2013) 036 [`arXiv:1307.6639`] [InSPIRE].

[28] S. Nojiri and S.D. Odintsov, *Modified non-local-F(R) gravity as the key for the inflation and dark energy*, *Phys. Lett.* **B 659** (2008) 821 [`arXiv:0708.0924`] [InSPIRE].

[29] S. Jhingan et al., *Phantom and non-phantom dark energy: the cosmological relevance of non-locally corrected gravity*, *Phys. Lett.* **B 663** (2008) 424 [`arXiv:0803.2613`] [InSPIRE].

[30] N.A. Koshelev, *Comments on scalar-tensor representation of nonlocally corrected gravity*, *Grav. Cosmol.* **15** (2009) 220 [`arXiv:0809.4927`] [InSPIRE].

[31] T.S. Koivisto, *Newtonian limit of nonlocal cosmology*, *Phys. Rev.* **D 78** (2008) 123505 [`arXiv:0807.3778`] [InSPIRE].

[32] T.S. Koivisto, *Cosmology of modified (but second order) gravity*, *AIP Conf. Proc.* **1206** (2010) 79 [`arXiv:0910.4097`] [ɪɴSPIRE].

[33] A. Kehagias and M. Maggiore, *Spherically symmetric static solutions in a nonlocal infrared modification of General Relativity*, *JHEP* **08** (2014) 029 [`arXiv:1401.8289`] [ɪɴSPIRE].

[34] S. Deser, J. Trubatch and S. Trubatch, *The massive spin-two field*, *Can. J. Phys.* **44** (1966) 1715.

[35] L. Alberte, A.H. Chamseddine and V. Mukhanov, *Massive gravity: resolving the puzzles*, *JHEP* **12** (2010) 023 [`arXiv:1008.5132`] [ɪɴSPIRE].

[36] R.D. Jordan, *Effective field equations for expectation values*, *Phys. Rev.* **D 33** (1986) 444 [ɪɴSPIRE].

[37] E. Calzetta and B.L. Hu, *Closed time path functional formalism in curved space-time: application to cosmological back reaction problems*, *Phys. Rev.* **D 35** (1987) 495 [ɪɴSPIRE].

[38] V. Mukhanov and S. Winitzki, *Introduction to quantum effects in gravity*, Cambridge University Press, Cambridge U.K. (2007).

[39] A.O. Barvinsky and G.A. Vilkovisky, *Beyond the Schwinger-Dewitt technique: converting loops into trees and in-in currents*, *Nucl. Phys.* **B 282** (1987) 163 [ɪɴSPIRE].

[40] M. Maggiore, *Perturbative loop corrections and nonlocal gravity*, *Phys. Rev.* **D 93** (2016) 063008 [`arXiv:1603.01515`] [ɪɴSPIRE].

[41] G. 't Hooft and M. Veltman, *One loop divergencies in the theory of gravitation*, *Ann. Poincare Phys. Theor.* **A20** (1974) 69.

[42] A.O. Barvinsky and G.A. Vilkovisky, *The generalized Schwinger-Dewitt technique in gauge theories and quantum gravity*, *Phys. Rept.* **119** (1985) 1 [ɪɴSPIRE].

[43] A.O. Barvinsky and G.A. Vilkovisky, *Covariant perturbation theory. 2: second order in the curvature. General algorithms*, *Nucl. Phys.* **B 333** (1990) 471 [ɪɴSPIRE].

[44] E.V. Gorbar and I.L. Shapiro, *Renormalization group and decoupling in curved space*, *JHEP* **02** (2003) 021 [`hep-ph/0210388`] [ɪɴSPIRE].

[45] E.V. Gorbar and I.L. Shapiro, *Renormalization group and decoupling in curved space. 2. The Standard model and beyond*, *JHEP* **06** (2003) 004 [`hep-ph/0303124`] [ɪɴSPIRE].

[46] N. Birrell and P. Davies, *Quantum fields in curved space*, Cambridge University Press, Cambridge U.K. (1982).

[47] I.L. Buchbinder, S.D. Odintsov and I.L. Shapiro, *Effective action in quantum gravity*, Institute of Physics, Bristol U.K. (1992).

[48] I.L. Shapiro, *Effective action of vacuum: semiclassical Approach*, *Class. Quant. Grav.* **25** (2008) 103001 [`arXiv:0801.0216`] [ɪɴSPIRE].

[49] J.F. Donoghue, *General relativity as an effective field theory: the leading quantum corrections*, *Phys. Rev.* **D 50** (1994) 3874 [`gr-qc/9405057`] [ɪɴSPIRE].

[50] A. Codello and R.K. Jain, *On the covariant formalism of the effective field theory of gravity and leading order corrections*, *Class. Quant. Grav.* **33** (2016) 225006 [`arXiv:1507.06308`] [ɪɴSPIRE].

[51] A. Codello and R.K. Jain, *On the covariant formalism of the effective field theory of gravity and its cosmological implications*, *Class. Quant. Grav.* **34** (2017) 035015 [`arXiv:1507.07829`] [ɪɴSPIRE].

[52] G.R. Dvali, G. Gabadadze and M. Porrati, *4D gravity on a brane in 5D Minkowski space*, *Phys. Lett.* **B 485** (2000) 208 [`hep-th/0005016`] [ɪɴSPIRE].

[53] K. Hinterbichler, *Theoretical aspects of massive gravity*, *Rev. Mod. Phys.* **84** (2012) 671 [arXiv:1105.3735] [INSPIRE].

[54] G.R. Dvali, G. Gabadadze and M. Porrati, *Metastable gravitons and infinite volume extra dimensions*, *Phys. Lett.* **B 484** (2000) 112 [hep-th/0002190] [INSPIRE].

[55] G. Gabadadze and M. Shifman, *Softly massive gravity*, *Phys. Rev.* **D 69** (2004) 124032 [hep-th/0312289] [INSPIRE].

[56] G. Dvali, S. Hofmann and J. Khoury, *Degravitation of the cosmological constant and graviton width*, *Phys. Rev.* **D 76** (2007) 084006 [hep-th/0703027] [INSPIRE].

[57] G. Dvali, G. Gabadadze and M. Shifman, *Diluting cosmological constant via large distance modification of gravity*, hep-th/0208096 [INSPIRE].

[58] G. Dvali, G. Gabadadze and M. Shifman, *Diluting cosmological constant in infinite volume extra dimensions*, *Phys. Rev.* **D 67** (2003) 044020 [hep-th/0202174] [INSPIRE].

[59] M. Porrati, *Fully covariant van Dam-Veltman-Zakharov discontinuity and absence thereof*, *Phys. Lett.* **B 534** (2002) 209 [hep-th/0203014] [INSPIRE].

[60] A. Karch and L. Randall, *Locally localized gravity*, *JHEP* **05** (2001) 008 [hep-th/0011156] [INSPIRE].

[61] A. Karch, E. Katz and L. Randall, *Absence of a VVDZ discontinuity in AdS(AdS)*, *JHEP* **12** (2001) 016 [hep-th/0106261] [INSPIRE].

[62] H.W. Hamber, *Quantum gravitation: the Feynman path integral approach*, Springer, Berlin Germany (2009).

[63] J. Ambjørn, A. Görlich, J. Jurkiewicz and R. Loll, *Nonperturbative quantum gravity*, *Phys. Rept.* **519** (2012) 127 [arXiv:1203.3591] [INSPIRE].

[64] R. Loll, *Quantum gravity from causal dynamical triangulations: a review*, *Class. Quant. Grav.* **37** (2020) 013002 [arXiv:1905.08669] [INSPIRE].

[65] H.W. Hamber, *On the gravitational scaling dimensions*, *Phys. Rev.* **D 61** (2000) 124008 [hep-th/9912246] [INSPIRE].

[66] H.W. Hamber and R.M. Williams, *Non-perturbative gravity and the spin of the lattice graviton*, *Phys. Rev.* **D 70** (2004) 124007 [hep-th/0407039] [INSPIRE].

[67] H.W. Hamber and R.M. Williams, *Nonlocal effective gravitational field equations and the running of Newton's G*, *Phys. Rev.* **D 72** (2005) 044026 [hep-th/0507017] [INSPIRE].

[68] H.W. Hamber and R. Toriumi, *Inconsistencies from a running cosmological constant*, *Int. J. Mod. Phys.* **D 22** (2013) 1330023 [arXiv:1301.6259] [INSPIRE].

[69] H.W. Hamber, *Scaling exponents for lattice quantum gravity in four dimensions*, *Phys. Rev.* **D 92** (2015) 064017 [arXiv:1506.07795] [INSPIRE].

[70] H.W. Hamber, *Vacuum condensate picture of quantum gravity*, *Symmetry* **11** (2019) 87 [arXiv:1707.08188] [INSPIRE].

[71] B. Knorr and F. Saueressig, *Towards reconstructing the quantum effective action of gravity*, *Phys. Rev. Lett.* **121** (2018) 161304 [arXiv:1804.03846] [INSPIRE].

[72] J. Polchinski, *Renormalization and effective lagrangians*, *Nucl. Phys.* **B 231** (1984) 269 [INSPIRE].

[73] C. Wetterich, *Exact evolution equation for the effective potential*, *Phys. Lett.* **B 301** (1993) 90 [arXiv:1710.05815] [INSPIRE].

[74] M. Reuter and F. Saueressig, *Quantum Einstein gravity*, *New J. Phys.* **14** (2012) 055022 [arXiv:1202.2274] [INSPIRE].

[75] C. Wetterich, *Graviton fluctuations erase the cosmological constant*, *Phys. Lett.* **B 773** (2017) 6 [arXiv:1704.08040] [InSPIRE].

[76] B. Knorr, *Infinite order quantum-gravitational correlations*, *Class. Quant. Grav.* **35** (2018) 115005 [arXiv:1710.07055] [InSPIRE].

[77] T.R. Morris, *Renormalization group properties in the conformal sector: towards perturbatively renormalizable quantum gravity*, *JHEP* **08** (2018) 024 [arXiv:1802.04281] [InSPIRE].

[78] C. Wetterich, *Infrared limit of quantum gravity*, *Phys. Rev.* **D 98** (2018) 026028 [arXiv:1802.05947] [InSPIRE].

[79] M. Reuter and H. Weyer, *Quantum gravity at astrophysical distances?*, *JCAP* **12** (2004) 001 [hep-th/0410119] [InSPIRE].

[80] J.A. Dietz, T.R. Morris and Z.H. Slade, *Fixed point structure of the conformal factor field in quantum gravity*, *Phys. Rev.* **D 94** (2016) 124014 [arXiv:1605.07636] [InSPIRE].

[81] M. Maggiore, *Dark energy and dimensional transmutation in R^2 gravity*, arXiv:1506.06217 [InSPIRE].

[82] M.B. Einhorn and D.R.T. Jones, *Naturalness and dimensional transmutation in classically scale-invariant gravity*, *JHEP* **03** (2015) 047 [arXiv:1410.8513] [InSPIRE].

[83] D. Tong and C. Turner, *Quantum dynamics of supergravity on $R^3 \times S^1$*, *JHEP* **12** (2014) 142 [arXiv:1408.3418] [InSPIRE].

[84] S. Foffa, M. Maggiore and E. Mitsou, *Cosmological dynamics and dark energy from nonlocal infrared modifications of gravity*, *Int. J. Mod. Phys.* **A 29** (2014) 1450116 [arXiv:1311.3435] [InSPIRE].

[85] S.X. Tian and Z.-H. Zhu, *Revisiting scalar and tensor perturbations in a nonlocal gravity*, *Phys. Rev.* **D 100** (2019) 124059 [arXiv:2001.00772] [InSPIRE].

[86] Y. Dirian, S. Foffa, N. Khosravi, M. Kunz and M. Maggiore, *Cosmological perturbations and structure formation in nonlocal infrared modifications of general relativity*, *JCAP* **06** (2014) 033 [arXiv:1403.6068] [InSPIRE].

[87] E. Belgacem, G. Cusin, S. Foffa, M. Maggiore and M. Mancarella, *Stability issues of nonlocal gravity during primordial inflation*, *Int. J. Mod. Phys.* **A 33** (2018) 1850007 [*Erratum ibid.* **A 33** (2018) 1892005] [arXiv:1610.05664] [InSPIRE].

[88] M. Chevallier and D. Polarski, *Accelerating universes with scaling dark matter*, *Int. J. Mod. Phys.* **D 10** (2001) 213 [gr-qc/0009008] [InSPIRE].

[89] E.V. Linder, *Exploring the expansion history of the universe*, *Phys. Rev. Lett.* **90** (2003) 091301 [astro-ph/0208512] [InSPIRE].

[90] S. Nesseris and S. Tsujikawa, *Cosmological perturbations and observational constraints on nonlocal massive gravity*, *Phys. Rev.* **D 90** (2014) 024070 [arXiv:1402.4613] [InSPIRE].

[91] L. Amendola, M. Kunz and D. Sapone, *Measuring the dark side (with weak lensing)*, *JCAP* **04** (2008) 013 [arXiv:0704.2421] [InSPIRE].

[92] S.F. Daniel et al., *Testing general relativity with current cosmological data*, *Phys. Rev.* **D 81** (2010) 123508 [arXiv:1002.1962] [InSPIRE].

[93] C. Deffayet, *Cosmology on a brane in Minkowski bulk*, *Phys. Lett.* **B 502** (2001) 199 [hep-th/0010186] [InSPIRE].

[94] C. Deffayet, G.R. Dvali and G. Gabadadze, *Accelerated universe from gravity leaking to extra dimensions*, *Phys. Rev.* **D 65** (2002) 044023 [astro-ph/0105068] [InSPIRE].

[95] M.A. Luty, M. Porrati and R. Rattazzi, *Strong interactions and stability in the DGP model*, *JHEP* **09** (2003) 029 [hep-th/0303116] [InSPIRE].

[96] A. Nicolis and R. Rattazzi, *Classical and quantum consistency of the DGP model*, *JHEP* **06** (2004) 059 [hep-th/0404159] [ɪɴSPIRE].

[97] D. Gorbunov, K. Koyama and S. Sibiryakov, *More on ghosts in DGP model*, *Phys. Rev.* **D 73** (2006) 044016 [hep-th/0512097] [ɪɴSPIRE].

[98] C. Charmousis, R. Gregory, N. Kaloper and A. Padilla, *DGP specteroscopy*, *JHEP* **10** (2006) 066 [hep-th/0604086] [ɪɴSPIRE].

[99] K. Izumi, K. Koyama and T. Tanaka, *Unexorcized ghost in DGP brane world*, *JHEP* **04** (2007) 053 [hep-th/0610282] [ɪɴSPIRE].

[100] C. de Rham and G. Gabadadze, *Generalization of the Fierz-Pauli Action*, *Phys. Rev.* **D 82** (2010) 044020 [arXiv:1007.0443] [ɪɴSPIRE].

[101] C. de Rham, G. Gabadadze and A.J. Tolley, *Resummation of massive gravity*, *Phys. Rev. Lett.* **106** (2011) 231101 [arXiv:1011.1232] [ɪɴSPIRE].

[102] S.F. Hassan and R.A. Rosen, *Resolving the ghost problem in non-linear massive gravity*, *Phys. Rev. Lett.* **108** (2012) 041101 [arXiv:1106.3344] [ɪɴSPIRE].

[103] G. D'Amico et al., *Massive cosmologies*, *Phys. Rev.* **D 84** (2011) 124046 [arXiv:1108.5231] [ɪɴSPIRE].

[104] S.F. Hassan and R.A. Rosen, *Bimetric gravity from ghost-free massive gravity*, *JHEP* **02** (2012) 126 [arXiv:1109.3515] [ɪɴSPIRE].

[105] F. Koennig et al., *Stable and unstable cosmological models in bimetric massive gravity*, *Phys. Rev.* **D 90** (2014) 124014 [arXiv:1407.4331] [ɪɴSPIRE].

[106] M. Lagos and P.G. Ferreira, *Cosmological perturbations in massive bigravity*, *JCAP* **12** (2014) 026 [arXiv:1410.0207] [ɪɴSPIRE].

[107] G. Cusin, R. Durrer, P. Guarato and M. Motta, *Gravitational waves in bigravity cosmology*, *JCAP* **05** (2015) 030 [arXiv:1412.5979] [ɪɴSPIRE].

[108] Y. Akrami et al., *Bimetric gravity is cosmologically viable*, *Phys. Lett.* **B 748** (2015) 37 [arXiv:1503.07521] [ɪɴSPIRE].

[109] G. Cusin, R. Durrer, P. Guarato and M. Motta, *Inflationary perturbations in bimetric gravity*, *JCAP* **09** (2015) 043 [arXiv:1505.01091] [ɪɴSPIRE].

[110] A. Schmidt-May and M. von Strauss, *Recent developments in bimetric theory*, *J. Phys.* **A 49** (2016) 183001 [arXiv:1512.00021] [ɪɴSPIRE].

[111] G. Cusin, R. Durrer, P. Guarato and M. Motta, *A general mass term for bigravity*, *JCAP* **04** (2016) 051 [arXiv:1512.02131] [ɪɴSPIRE].

[112] M. Ishak, *Testing general relativity in cosmology*, *Living Rev. Rel.* **22** (2019) 1 [arXiv:1806.10122] [ɪɴSPIRE].

[113] Y. Dirian et al., *Non-local gravity and comparison with observational datasets*, *JCAP* **04** (2015) 044 [arXiv:1411.7692] [ɪɴSPIRE].

[114] Y. Dirian et al., *Non-local gravity and comparison with observational datasets. II. Updated results and Bayesian model comparison with ΛCDM*, *JCAP* **05** (2016) 068 [arXiv:1602.03558] [ɪɴSPIRE].

[115] E. Belgacem et al., *Nonlocal gravity and gravitational-wave observations*, *JCAP* **11** (2019) 022 [arXiv:1907.02047] [ɪɴSPIRE].

[116] D. Blas, J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) II: approximation schemes*, *JCAP* **07** (2011) 034 [arXiv:1104.2933] [ɪɴSPIRE].

[117] E. Bellini et al., *Comparison of Einstein-Boltzmann solvers for testing general relativity*, *Phys. Rev.* **D 97** (2018) 023520 [arXiv:1709.09135] [ɪɴSPIRE].

[118] https://github.com/AndreasFinke/class_public

[119] https://github.com/dirian/class_public/tree/nonlocal

[120] Y. Dirian, *Changing the Bayesian prior: absolute neutrino mass constraints in nonlocal gravity*, *Phys. Rev.* **D 96** (2017) 083513 [arXiv:1704.04075] [INSPIRE].

[121] M.C. Gonzalez-Garcia, M. Maltoni, J. Salvado and T. Schwetz, *Global fit to three neutrino mixing: critical look at present precision*, *JHEP* **12** (2012) 123 [arXiv:1209.3023] [INSPIRE].

[122] PLANCK collaboration, *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [arXiv:1502.01589] [INSPIRE].

[123] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, arXiv:1807.06209 [INSPIRE].

[124] PLANCK collaboration, *Planck 2018 results. V. CMB power spectra and likelihoods*, arXiv:1907.12875 [INSPIRE].

[125] PLANCK collaboration, *Planck 2018 results. VIII. Gravitational lensing*, arXiv:1807.06210 [INSPIRE].

[126] D.M. Scolnic et al., *The complete light-curve sample of spectroscopically confirmed SNe Ia from Pan-STARRS1 and cosmological constraints from the combined pantheon sample*, *Astrophys. J.* **859** (2018) 101 [arXiv:1710.00845] [INSPIRE].

[127] SDSS collaboration, *Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples*, *Astron. Astrophys.* **568** (2014) A22 [arXiv:1401.4064] [INSPIRE].

[128] F. Beutler et al., *The 6dF galaxy survey: baryon acoustic oscillations and the local Hubble Constant*, *Mon. Not. Roy. Astron. Soc.* **416** (2011) 3017 [arXiv:1106.3366].

[129] A.J. Ross et al., *The clustering of the SDSS DR7 main Galaxy sample — I. A 4 per cent distance measure at $z = 0.15$*, *Mon. Not. Roy. Astron. Soc.* **449** (2015) 835 [arXiv:1409.3242] [INSPIRE].

[130] BOSS collaboration, *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, *Mon. Not. Roy. Astron. Soc.* **470** (2017) 2617 [arXiv:1607.03155] [INSPIRE].

[131] A. Lewis and S. Bridle, *Cosmological parameters from CMB and other data: A Monte Carlo approach*, *Phys. Rev.* **D 66** (2002) 103511 [astro-ph/0205436] [INSPIRE].

[132] A. Lewis, *Efficient sampling of fast and slow cosmological parameters*, *Phys. Rev.* **D 87** (2013) 103529 [arXiv:1304.4473] [INSPIRE].

[133] R.M. Neal, *Taking bigger metropolis steps by dragging fast variables*, math/0502099.

[134] C. Cartis, J. Fiala, B. Marteau and L. Roberts, *Improving the flexibility and robustness of model-based derivative-free optimization solvers*, *A.C.M. Trans. Math. Softw.* **45** (2019) 32.

[135] C. Cartis, L. Roberts and O. Sheridan-Methven, *Escaping local minima with derivative-free methods: a numerical investigation*, arXiv:1812.11343.

[136] M.J. Powell, *The bobyqa algorithm for bound constrained optimization without derivatives*, Cambridge NA Report NA2009/06 (2009).

[137] A.G. Riess et al., *Large Magellanic Cloud Cepheid standards provide a 1% foundation for the determination of the Hubble constant and stronger evidence for physics beyond ΛCDM*, *Astrophys. J.* **876** (2019) 85 [arXiv:1903.07603] [INSPIRE].

[138] K.C. Wong et al., *H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes*, arXiv:1907.04869 [INSPIRE].

[139] V. Poulin, K.K. Boddy, S. Bird and M. Kamionkowski, *Implications of an extended dark energy cosmology with massive neutrinos for cosmological tensions*, *Phys. Rev.* **D 97** (2018) 123504 [arXiv:1803.02474] [INSPIRE].

[140] K. Aylor et al., *Sounds discordant: classical distance ladder & ΛCDM-based determinations of the cosmological sound horizon*, *Astrophys. J.* **874** (2019) 4 [arXiv:1811.00537] [INSPIRE].

[141] M. Moresco et al., *Constraining the time evolution of dark energy, curvature and neutrino properties with cosmic chronometers*, *JCAP* **12** (2016) 039 [arXiv:1604.00183] [INSPIRE].

[142] S. Nesseris, D. Sapone and S. Sypsas, *Evaporating primordial black holes as varying dark energy*, *Phys. Dark Univ.* **27** (2020) 100413 [arXiv:1907.05608].

[143] L. Kazantzidis and L. Perivolaropoulos, *Evolution of the $f\sigma_8$ tension with the Planck15/ΛCDM determination and implications for modified gravity theories*, *Phys. Rev.* **D 97** (2018) 103503 [arXiv:1803.01337] [INSPIRE].

[144] F. Beutler et al., *The 6dF galaxy survey: $z \approx 0$ measurement of the growth rate and σ_8*, *Mon. Not. Roy. Astron. Soc.* **423** (2012) 3430 [arXiv:1204.4725].

[145] A. Oka et al., *Simultaneous constraints on the growth of structure and cosmic expansion from the multipole power spectra of the SDSS DR7 LRG sample*, *Mon. Not. Roy. Astron. Soc.* **439** (2014) 2515 [arXiv:1310.2820] [INSPIRE].

[146] L. Samushia et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measuring growth rate and geometry with anisotropic clustering*, *Mon. Not. Roy. Astron. Soc.* **439** (2014) 3504 [arXiv:1312.4899] [INSPIRE].

[147] C. Blake et al., *The WiggleZ dark energy survey: joint measurements of the expansion and growth history at $z < 1$*, *Mon. Not. Roy. Astron. Soc.* **425** (2012) 405 [arXiv:1204.3674].

[148] S. de la Torre et al., *The VIMOS Public Extragalactic Redshift Survey (VIPERS). Galaxy clustering and redshift-space distortions at $z = 0.8$ in the first data release*, *Astron. Astrophys.* **557** (2013) A54 [arXiv:1303.2622] [INSPIRE].

[149] EUCLID THEORY WORKING GROUP collaboration, *Cosmology and fundamental physics with the Euclid satellite*, *Living Rev. Rel.* **16** (2013) 6 [arXiv:1206.1225] [INSPIRE].

[150] Y.-S. Song et al., *Complementarity of weak lensing and peculiar velocity measurements in testing general relativity*, *Phys. Rev.* **D 84** (2011) 083523 [arXiv:1011.2106] [INSPIRE].

[151] M. Ishak et al., *Modified gravity and dark energy models beyond $w(z)CDM$ testable by LSST*, arXiv:1905.09687 [INSPIRE].

[152] A.I. Vainshtein, *To the problem of nonvanishing gravitation mass*, *Phys. Lett.* **39B** (1972) 393 [INSPIRE].

[153] C. Deffayet, G.R. Dvali, G. Gabadadze and A.I. Vainshtein, *Nonperturbative continuity in graviton mass versus perturbative discontinuity*, *Phys. Rev.* **D 65** (2002) 044026 [hep-th/0106001] [INSPIRE].

[154] N. Arkani-Hamed, H. Georgi and M.D. Schwartz, *Effective field theory for massive gravitons and gravity in theory space*, *Annals Phys.* **305** (2003) 96 [hep-th/0210184] [INSPIRE].

[155] K. Koyama, G. Niz and G. Tasinato, *Analytic solutions in non-linear massive gravity*, *Phys. Rev. Lett.* **107** (2011) 131101 [arXiv:1103.4708] [INSPIRE].

[156] K. Koyama, G. Niz and G. Tasinato, *Strong interactions and exact solutions in non-linear massive gravity*, *Phys. Rev.* **D 84** (2011) 064033 [arXiv:1104.2143] [INSPIRE].

[157] A. Barreira et al., *Nonlinear structure formation in nonlocal gravity*, *JCAP* **09** (2014) 031 [arXiv:1408.1084] [INSPIRE].

[158] F. Hofmann and J. Müller, *Relativistic tests with lunar laser ranging*, *Class. Quant. Grav.* **35** (2018) 035015 [INSPIRE].

[159] M. Carrera and D. Giulini, *On the influence of global cosmological expansion on the dynamics and kinematics of local systems*, arXiv:0810.2712 [INSPIRE].

[160] W.G. Laarakkers and E. Poisson, *Radiative falloff in Einstein-Straus space-time*, *Phys. Rev.* **D 64** (2001) 084008 [gr-qc/0105016] [INSPIRE].

[161] E. Belgacem, A. Finke, A. Frassino and M. Maggiore, *Testing nonlocal gravity with Lunar Laser Ranging*, *JCAP* **02** (2019) 035 [arXiv:1812.11181] [INSPIRE].

[162] E. Babichev, C. Deffayet and G. Esposito-Farese, *Constraints on shift-symmetric scalar-tensor theories with a vainshtein mechanism from bounds on the time variation of G*, *Phys. Rev. Lett.* **107** (2011) 251102 [arXiv:1107.1569] [INSPIRE].

[163] S. Tsujikawa, *Lunar laser ranging constraints on nonminimally coupled dark energy and standard sirens*, *Phys. Rev.* **D 100** (2019) 043510 [arXiv:1903.07092] [INSPIRE].

[164] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Gravitational-wave luminosity distance in modified gravity theories*, *Phys. Rev.* **D 97** (2018) 104066 [arXiv:1712.08108] [INSPIRE].

[165] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Modified gravitational-wave propagation and standard sirens*, *Phys. Rev.* **D 98** (2018) 023510 [arXiv:1805.08731] [INSPIRE].

[166] M. Maggiore, *Gravitational waves. Volume 1: theory and experiments*, Oxford Master Series in Physics, Oxford University Press, Oxford U.K. (2007).

[167] B.F. Schutz, *Determining the Hubble constant from gravitational wave observations*, *Nature* **323** (1986) 310 [INSPIRE].

[168] N. Dalal, D.E. Holz, S.A. Hughes and B. Jain, *Short grb and binary black hole standard sirens as a probe of dark energy*, *Phys. Rev.* **D 74** (2006) 063006 [astro-ph/0601275] [INSPIRE].

[169] C.L. MacLeod and C.J. Hogan, *Precision of Hubble constant derived using black hole binary absolute distances and statistical redshift information*, *Phys. Rev.* **D 77** (2008) 043512 [arXiv:0712.0618] [INSPIRE].

[170] S. Nissanke et al., *Exploring short gamma-ray bursts as gravitational-wave standard sirens*, *Astrophys. J.* **725** (2010) 496 [arXiv:0904.1017] [INSPIRE].

[171] C. Cutler and D.E. Holz, *Ultra-high precision cosmology from gravitational waves*, *Phys. Rev.* **D 80** (2009) 104009 [arXiv:0906.3752] [INSPIRE].

[172] B.S. Sathyaprakash, B.F. Schutz and C. Van Den Broeck, *Cosmography with the Einstein telescope*, *Class. Quant. Grav.* **27** (2010) 215006 [arXiv:0906.4151] [INSPIRE].

[173] W. Zhao, C. Van Den Broeck, D. Baskaran and T.G.F. Li, *Determination of dark energy by the Einstein telescope: comparing with CMB, BAO and SNIa observations*, *Phys. Rev.* **D 83** (2011) 023005 [arXiv:1009.0206] [INSPIRE].

[174] W. Del Pozzo, *Inference of the cosmological parameters from gravitational waves: application to second generation interferometers*, *Phys. Rev.* **D 86** (2012) 043011 [arXiv:1108.1317] [INSPIRE].

[175] A. Nishizawa, K. Yagi, A. Taruya and T. Tanaka, *Cosmology with space-based gravitational-wave detectors — dark energy and primordial gravitational waves*, *Phys. Rev.* **D 85** (2012) 044047 [arXiv:1110.2865] [INSPIRE].

[176] S.R. Taylor, J.R. Gair and I. Mandel, *Hubble without the Hubble: cosmology using advanced gravitational-wave detectors alone*, *Phys. Rev.* **D 85** (2012) 023535 [arXiv:1108.5161] [INSPIRE].

[177] S.R. Taylor and J.R. Gair, *Cosmology with the lights off: standard sirens in the Einstein Telescope era*, *Phys. Rev.* **D 86** (2012) 023502 [arXiv:1204.6739] [INSPIRE].

[178] N. Tamanini et al., *Science with the space-based interferometer eLISA. III: probing the expansion of the Universe using gravitational wave standard sirens*, *JCAP* **04** (2016) 002 [arXiv:1601.07112] [InSPIRE].

[179] R.-G. Cai and T. Yang, *Estimating cosmological parameters by the simulated data of gravitational waves from the Einstein Telescope*, *Phys. Rev.* **D 95** (2017) 044024 [arXiv:1608.08008] [InSPIRE].

[180] LIGO SCIENTIFIC collaboration et al., *A gravitational-wave standard siren measurement of the Hubble constant*, *Nature* **551** (2017) 85 [arXiv:1710.05835] [InSPIRE].

[181] LIGO SCIENTIFIC collaboration et al., *Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A*, *Astrophys. J.* **848** (2017) L13 [arXiv:1710.05834] [InSPIRE].

[182] C. de Rham and S. Melville, *Gravitational rainbows: LIGO and dark energy at its cutoff*, *Phys. Rev. Lett.* **121** (2018) 221101 [arXiv:1806.09417] [InSPIRE].

[183] P. Creminelli and F. Vernizzi, *Dark energy after GW170817 and GRB170817A*, *Phys. Rev. Lett.* **119** (2017) 251302 [arXiv:1710.05877] [InSPIRE].

[184] J. Sakstein and B. Jain, *Implications of the neutron star merger GW170817 for cosmological scalar-tensor theories*, *Phys. Rev. Lett.* **119** (2017) 251303 [arXiv:1710.05893] [InSPIRE].

[185] J.M. Ezquiaga and M. Zumalacárregui, *Dark energy after GW170817: dead ends and the road ahead*, *Phys. Rev. Lett.* **119** (2017) 251304 [arXiv:1710.05901] [InSPIRE].

[186] T. Baker et al., *Strong constraints on cosmological gravity from GW170817 and GRB 170817A*, *Phys. Rev. Lett.* **119** (2017) 251301 [arXiv:1710.06394] [InSPIRE].

[187] I.D. Saltas, I. Sawicki, L. Amendola and M. Kunz, *Anisotropic stress as a signature of nonstandard propagation of gravitational waves*, *Phys. Rev. Lett.* **113** (2014) 191101 [arXiv:1406.7139] [InSPIRE].

[188] L. Lombriser and A. Taylor, *Breaking a dark degeneracy with gravitational waves*, *JCAP* **03** (2016) 031 [arXiv:1509.08458] [InSPIRE].

[189] S. Arai and A. Nishizawa, *Generalized framework for testing gravity with gravitational-wave propagation. II. Constraints on Horndeski theory*, *Phys. Rev.* **D 97** (2018) 104038 [arXiv:1711.03776] [InSPIRE].

[190] L. Amendola, I. Sawicki, M. Kunz and I.D. Saltas, *Direct detection of gravitational waves can measure the time variation of the Planck mass*, *JCAP* **08** (2018) 030 [arXiv:1712.08623] [InSPIRE].

[191] LISA COSMOLOGY WORKING GROUP collaboration, *Testing modified gravity at cosmological distances with LISA standard sirens*, *JCAP* **07** (2019) 024 [arXiv:1906.01593] [InSPIRE].

[192] C. Deffayet and K. Menou, *Probing gravity with spacetime sirens*, *Astrophys. J.* **668** (2007) L143 [arXiv:0709.0003] [InSPIRE].

[193] K. Pardo, M. Fishbach, D.E. Holz and D.N. Spergel, *Limits on the number of spacetime dimensions from GW170817*, *JCAP* **07** (2018) 048 [arXiv:1801.08160] [InSPIRE].

[194] J. Gleyzes, D. Langlois and F. Vernizzi, *A unifying description of dark energy*, *Int. J. Mod. Phys.* **D 23** (2015) 1443010 [arXiv:1411.3712] [InSPIRE].

[195] A. Nishizawa, *Generalized framework for testing gravity with gravitational-wave propagation. I. Formulation*, *Phys. Rev.* **D 97** (2018) 104037 [arXiv:1710.04825] [InSPIRE].

[196] A. Garoffolo et al., *Gravitational waves and geometrical optics in scalar-tensor theories*, arXiv:1912.08093 [InSPIRE].

[197] E.V. Linder, *No slip gravity*, *JCAP* **03** (2018) 005 [arXiv:1801.01503] [InSPIRE].

[198] C. Dalang and L. Lombriser, *Limitations on standard sirens tests of gravity from screening*, *JCAP* **10** (2019) 013 [arXiv:1906.12333] [INSPIRE].

[199] R. D'Agostino and R.C. Nunes, *Probing observational bounds on scalar-tensor theories from standard sirens*, *Phys. Rev.* **D 100** (2019) 044041 [arXiv:1907.05516] [INSPIRE].

[200] D.J.H. Chung and K. Freese, *Can geodesics in extra dimensions solve the cosmological horizon problem?*, *Phys. Rev.* **D 62** (2000) 063513 [hep-ph/9910235] [INSPIRE].

[201] R.R. Caldwell and D. Langlois, *Shortcuts in the fifth dimension*, *Phys. Lett.* **B 511** (2001) 129 [gr-qc/0103070] [INSPIRE].

[202] L. Visinelli, N. Bolis and S. Vagnozzi, *Brane-world extra dimensions in light of GW170817*, *Phys. Rev.* **D 97** (2018) 064039 [arXiv:1711.06628] [INSPIRE].

[203] E. Belgacem et al., *Cosmology and dark energy from joint gravitational wave-GRB observations*, *JCAP* **08** (2019) 015 [arXiv:1907.01487] [INSPIRE].

[204] E. Belgacem, S. Foffa, M. Maggiore and T. Yang, *Gaussian processes reconstruction of modified gravitational wave propagation*, *Phys. Rev.* **D 101** (2020) 063505 [arXiv:1911.11497] [INSPIRE].

[205] T. Regimbau et al., *A Mock data challenge for the Einstein gravitational-wave telescope*, *Phys. Rev.* **D 86** (2012) 122001 [arXiv:1201.3563] [INSPIRE].

[206] T. Regimbau, D. Meacher and M. Coughlin, *Second Einstein telescope mock science challenge: detection of the gravitational-wave stochastic background from compact binary coalescences*, *Phys. Rev.* **D 89** (2014) 084046 [arXiv:1404.1134] [INSPIRE].

[207] T. Regimbau et al., *Revisiting coincidence rate between gravitational wave detection and short gamma-ray burst for the advanced and third generation*, *Astrophys. J.* **799** (2015) 69 [arXiv:1410.2739] [INSPIRE].

[208] D. Meacher et al., *Mock data and science challenge for detecting an astrophysical stochastic gravitational-wave background with Advanced LIGO and Advanced Virgo*, *Phys. Rev.* **D 92** (2015) 063002 [arXiv:1506.06744] [INSPIRE].

[209] D. Meacher et al., *Second Einstein telescope mock data and science challenge: low frequency binary neutron star data analysis*, *Phys. Rev.* **D 93** (2016) 024018 [arXiv:1511.01592] [INSPIRE].

[210] T. Regimbau et al., *Digging deeper: Observing primordial gravitational waves below the binary black hole produced stochastic background*, *Phys. Rev. Lett.* **118** (2017) 151105 [arXiv:1611.08943] [INSPIRE].

[211] LIGO SCIENTIFIC, VIRGO collaboration, *GW170817: implications for the stochastic gravitational-wave background from compact binary coalescences*, *Phys. Rev. Lett.* **120** (2018) 091101 [arXiv:1710.05837] [INSPIRE].

[212] E. Vangioni et al., *The impact of star formation and gamma-ray burst rates at high redshift on cosmic chemical evolution and reionization*, *Mon. Not. Roy. Astron. Soc.* **447** (2015) 2575 [arXiv:1409.2462] [INSPIRE].

[213] P. Madau and M. Dickinson, *Cosmic star formation history*, *Ann. Rev. Astron. Astrophys.* **52** (2014) 415 [arXiv:1403.0007] [INSPIRE].

[214] S. Vitale, W.M. Farr, K. Ng and C.L. Rodriguez, *Measuring the star formation rate with gravitational waves from binary black holes*, *Astrophys. J.* **886** (2019) L1 [arXiv:1808.00901] [INSPIRE].

[215] LIGO SCIENTIFIC, VIRGO collaboration, *GWTC-1: a gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs*, *Phys. Rev.* **X 9** (2019) 031040 [arXiv:1811.12907] [INSPIRE].

[216] LIGO Scientific, Virgo collaboration, *GW170817: observation of gravitational waves from a binary neutron star inspiral*, *Phys. Rev. Lett.* **119** (2017) 161101 [arXiv:1710.05832] [inSPIRE].

[217] LIGO Scientific, Virgo collaboration, *GW190425: observation of a compact binary coalescence with total mass* $\sim 3.4 M_\odot$, arXiv:2001.01761 [inSPIRE].

[218] M. Seikel, C. Clarkson and M. Smith, *Reconstruction of dark energy and expansion dynamics using Gaussian processes*, *JCAP* **06** (2012) 036 [arXiv:1204.2832] [inSPIRE].

[219] M. Seikel, S. Yahya, R. Maartens and C. Clarkson, *Using $H(z)$ data as a probe of the concordance model*, *Phys. Rev.* **D 86** (2012) 083001 [arXiv:1205.3431] [inSPIRE].

[220] S. Yahya et al., *Null tests of the cosmological constant using supernovae*, *Phys. Rev.* **D 89** (2014) 023503 [arXiv:1308.4099] [inSPIRE].

[221] V.C. Busti, C. Clarkson and M. Seikel, *Evidence for a lower value for H_0 from cosmic chronometers data?*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) 11 [arXiv:1402.5429] [inSPIRE].

[222] V.C. Busti and C. Clarkson, *Dodging the dark matter degeneracy while determining the dynamics of dark energy*, *JCAP* **05** (2016) 008 [arXiv:1505.01821] [inSPIRE].

[223] T. Yang, Z.-K. Guo and R.-G. Cai, *Reconstructing the interaction between dark energy and dark matter using Gaussian Processes*, *Phys. Rev.* **D 91** (2015) 123533 [arXiv:1505.04443] [inSPIRE].

[224] R.-G. Cai, Z.-K. Guo and T. Yang, *Null test of the cosmic curvature using $H(z)$ and supernovae data*, *Phys. Rev.* **D 93** (2016) 043517 [arXiv:1509.06283] [inSPIRE].

[225] R.-G. Cai, Z.-K. Guo and T. Yang, *Dodging the cosmic curvature to probe the constancy of the speed of light*, *JCAP* **08** (2016) 016 [arXiv:1601.05497] [inSPIRE].

[226] R.-G. Cai, N. Tamanini and T. Yang, *Reconstructing the dark sector interaction with LISA*, *JCAP* **05** (2017) 031 [arXiv:1703.07323] [inSPIRE].

[227] J.P. Bernstein et al., *Supernova simulations and strategies for the dark energy survey*, *Astrophys. J.* **753** (2012) 152 [arXiv:1111.1969] [inSPIRE].

[228] M. Punturo et al., *The Einstein telescope: a third-generation gravitational wave observatory*, *Class. Quant. Grav.* **27** (2010) 194002 [inSPIRE].

[229] S. Dwyer et al., *Gravitational wave detector with cosmological reach*, *Phys. Rev.* **D 91** (2015) 082001 [arXiv:1410.0612] [inSPIRE].

[230] D. Reitze et al., *Cosmic explorer: the U.S. contribution to gravitational-wave astronomy beyond LIGO*, *Bull. Am. Astron. Soc.* **51** (2019) 035 [arXiv:1907.04833] [inSPIRE].

[231] B. Sathyaprakash et al., *Scientific objectives of Einstein telescope*, *Class. Quant. Grav.* **29** (2012) 124013 [*Erratum ibid.* **30** (2013) 079501] [arXiv:1206.0331] [inSPIRE].

[232] M. Maggiore et al., *Science case for the Einstein telescope*, arXiv:1912.02622 [inSPIRE].

[233] THESEUS collaboration, *THESEUS: a key space mission concept for multi-messenger astrophysics*, *Adv. Space Res.* **62** (2018) 662 [arXiv:1712.08153] [inSPIRE].

[234] THESEUS collaboration, *The THESEUS space mission concept: science case, design and expected performances*, *Adv. Space Res.* **62** (2018) 191 [arXiv:1710.04638] [inSPIRE].

[235] G. Stratta, L. Amati, R. Ciolfi and S. Vinciguerra, *THESEUS in the era of multi-messenger astronomy*, *Mem. Soc. Ast. It.* **89** (2018) 205 [arXiv:1802.01677] [inSPIRE].

[236] B.S. Sathyaprakash et al., *Multimessenger universe with gravitational waves from binaries*, arXiv:1903.09277 [inSPIRE].

[237] N. Arkani-Hamed, S. Dimopoulos, G. Dvali and G. Gabadadze, *Nonlocal modification of gravity and the cosmological constant problem*, hep-th/0209227 [INSPIRE].

[238] H. Nersisyan, Y. Akrami, L. Amendola, T.S. Koivisto and J. Rubio, *Dynamical analysis of $R\frac{1}{\Box^2}R$ cosmology: impact of initial conditions and constraints from supernovae*, *Phys. Rev.* **D 94** (2016) 043531 [arXiv:1606.04349] [INSPIRE].

[239] G. Cusin, S. Foffa, M. Maggiore and M. Mancarella, *Nonlocal gravity with a Weyl-square term*, *Phys. Rev.* **D 93** (2016) 043006 [arXiv:1512.06373] [INSPIRE].

[240] G. Cusin, S. Foffa, M. Maggiore and M. Mancarella, *Conformal symmetry and nonlinear extensions of nonlocal gravity*, *Phys. Rev.* **D 93** (2016) 083008 [arXiv:1602.01078] [INSPIRE].

[241] V. Vardanyan, Y. Akrami, L. Amendola and A. Silvestri, *On nonlocally interacting metrics and a simple proposal for cosmic acceleration*, *JCAP* **03** (2018) 048 [arXiv:1702.08908] [INSPIRE].

[242] L. Amendola, N. Burzilla and H. Nersisyan, *Quantum gravity inspired nonlocal gravity model*, *Phys. Rev.* **D 96** (2017) 084031 [arXiv:1707.04628] [INSPIRE].

[243] C. Wetterich, *Effective nonlocal Euclidean gravity*, *Gen. Rel. Grav.* **30** (1998) 159 [gr-qc/9704052] [INSPIRE].

[244] R.P. Woodard, *Nonlocal models of cosmic acceleration*, *Found. Phys.* **44** (2014) 213 [arXiv:1401.0254] [INSPIRE].

[245] C. Deffayet and R.P. Woodard, *Reconstructing the distortion function for nonlocal cosmology*, *JCAP* **08** (2009) 023 [arXiv:0904.0961] [INSPIRE].

[246] S. Deser and R.P. Woodard, *Nonlocal Cosmology II — Cosmic acceleration without fine tuning or dark energy*, *JCAP* **06** (2019) 034 [arXiv:1902.08075] [INSPIRE].