

## Causality in curved spacetimes: The speed of light and gravity

Claudia de Rham<sup>\*</sup> and Andrew J. Tolley<sup>†</sup>

*Theoretical Physics, Blackett Laboratory, Imperial College, London SW7 2AZ, United Kingdom  
and CERCA, Department of Physics, Case Western Reserve University,  
10900 Euclid Avenue, Cleveland, Ohio 44106, USA*



(Received 30 July 2020; accepted 18 August 2020; published 20 October 2020)

Within the low-energy effective field theories of quantum electrodynamics and gravity, the *low-energy* speed of light or that of gravitational waves can typically be mildly superluminal in curved spacetimes. Related to this, small scattering time advances relative to the curved background can emerge from known effective field theory coefficients for photons or gravitons. We clarify why these results are not in contradiction with causality, analyticity or Lorentz invariance, and highlight various subtleties that arise when dealing with superluminalities and time advances in the gravitational context. Consistent low-energy effective theories are shown to self-protect by ensuring that any time advance and superluminality calculated within the regime of validity of the effective theory is necessarily *unresolvable*, and cannot be argued to lead to a macroscopically larger light cone. Such considerations are particularly relevant for putting constraints on cosmological and gravitational effective field theories and we provide explicit criteria to be satisfied so as to ensure causality.

DOI: [10.1103/PhysRevD.102.084048](https://doi.org/10.1103/PhysRevD.102.084048)

### I. INTRODUCTION

Causality and unitarity play a crucial role in fixing the structure of a Lorentz invariant quantum field theory as was recognized early on. This is most immediately apparent in the dispersion relation methods [1] utilized for example in the spectral representation of Källén and Lehmann [2,3] where the Fourier space Feynman propagator is recognized to be an analytic function of complex momentum squared up to a pole and right-hand branch cut. These dispersion relation methods evolved into the S-matrix analyticity program of the 1960s which—albeit in a different form—plays a crucial role today in amplitude methods. More recently, S-matrix methods have been used to put constraints on low-energy effective theories (EFTs), either through positivity bounds [4–22], demanding scattering time delays are positive [asymptotic (sub)luminality] [23–27], or related methods.

One unfortunate feature of these methods is that there is no clear way to extend them to curved spacetimes, except perhaps maximally symmetric cases; see [22,28,29] for some attempts to deal with this. The powerful analyticity properties combined with crossing symmetry are spoiled

when the background spacetime is time dependent. In the absence of clear causality constraints, one reasonable guess is to demand that relativistic causality should require the propagation speed of all degrees of freedom to be (sub)luminal. For a nongravitational theory this is largely a reasonable criterion, and indeed it is known that certain constraints from positivity bounds appear to be connected with (sub)luminality of fluctuations about different backgrounds [6].

For a gravitational theory these questions are altogether more subtle. The ability to perform field redefinitions that change the off-shell metric means that it is no longer clear which light cone to use as a reference. Due to the ambiguity of field redefinitions, causality in a gravitational theory is usually phrased for on-shell invariant quantities, such as the requirement that the tree level scattering matrix is analytic in terms of its Mandelstam variables. The relation between causality and analyticity is highlighted for example in Refs. [30–32]. But such considerations are typically of little use in understanding causality in spacetimes with significant curvature effects such as FLRW and Schwarzschild. The peculiar subtleties associated with causality in curved spacetime are well known in the context of the low-energy effective theory for quantum electrodynamics (QED) below the electron mass [33,34], where the low-energy phase and group velocity of light is known to be superluminal for certain polarization states on some curved geometries. The relation with analyticity and causality in this example has been discussed extensively in the literature, see for example Refs. [35–44].

<sup>\*</sup>c.de-rham@imperial.ac.uk

<sup>†</sup>a.tolley@imperial.ac.uk

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

From the cosmological perspective, it is known that various cosmological models such as inflation and dark energy theories can easily exhibit different speeds of propagation in different sectors. It is also known that the speed of gravitational waves (GWs) in a given low-energy EFT can be different than the luminal speed inferred from the metric from which the theory is constructed. Once again this leads to the obvious question of how causality should be set in the cosmological context. In the literature, it is often considered, by fiat, that the speed of GWs and of all other species should be (sub)luminal in an arbitrary background with respect to the metric out of which the theory is constructed. Doing this often imposes constraints on the signs of coefficients in the effective action, but as already mentioned such a criterion is not invariant under field redefinitions.

The previous criterion can sometimes be correct; however, in the canonical case of gravitational effective theories it is not necessarily the case, and indeed *demanding*  $c_s \leq 1$  for all species can sit in contradiction with those same requirements of causality and analyticity. In two recent papers [45,46] it was pointed out that integrating out matter fields typically leads to curvature operators in the low-energy EFT of gravity that can lead to a (small) superluminal low-energy speed for GWs. For the low-energy EFT of gravity, constraining the signs of the low-energy operators to be so as to entirely forbid superluminal low-energy speed, no matter how small, would lead to a criterion in contradiction with known partial high-energy completions and more generally expectations based on analyticity. Instead, we highlight that a small amount of low-energy superluminal speed does not directly imply that the support of the retarded propagator lies outside the usual light cone.

The key point is that superluminality of the low-energy speed is only in conflict with relativistic causality if it can be integrated over time to make a large macroscopic effect, i.e., the light cone of causal influence is *measurably larger* than the background geometry light cone. In solving for the retarded propagator perturbatively in the EFT expansion around the background one, this only occurs when there are *secular* effects that need to be resummed, leading to significant differences in the structure of the propagator at late times (or large distances). We will demonstrate that in those situations where a small superluminal low-energy speed arises in a given gravitational effective theory, the requirement that the EFT is under control automatically precludes any secular effects, preserving causality by ensuring that at any finite order in the convergent EFT expansion, the retarded propagator has the same causal structure as the unperturbed one. That this is the case hinges on the smallness of the superluminal speed correction that arises within gravitational effective theories. We may regard this as a “self-protection” mechanism against causality violation for consistent low-energy EFTs.

In the case of asymptotically flat geometries where an S-matrix may be defined, this discussion may be further sharpened. We may precisely define a relativistic generalization of the Eisenbud-Wigner scattering time delay  $\Delta T$  [47–50]. In particular for spherically symmetric spacetimes we may define a time delay  $\Delta T_\ell$  for each partial wave  $\ell$  via the derivative of the scattering phase shift at fixed  $\ell$

$$\Delta T_\ell = 2 \left. \frac{\partial \delta_\ell(\omega)}{\partial \omega} \right|_\ell, \quad (1.1)$$

in terms of the incident particle’s energy  $\omega$ . We may also consider the inequivalent but related time delay for fixed impact parameter  $b$

$$\Delta T_b = 2 \left. \frac{\partial \delta_\ell(\omega)}{\partial \omega} \right|_b. \quad (1.2)$$

In general relativity (GR) these time delays are the well-known “Shapiro” or “gravitational” time delays [51]. There is a long history of using positivity or boundedness<sup>1</sup> of this time delay as a means of imposing causality (see for example [52] for a review) going back to Eisenbud and Wigner and improved by Smith [49]. More recently in the context of relativistic field theories, in [23] the eikonal approximation was used to determine this scattering time delay in various low-energy effective theories. Subsequently, for example in [24–27,53,54], it was generally argued on the same lines outlined by Eisenbud and Wigner [47,48] that an overall negative time delay is a signal of causality violation and inconsistency of the low-energy effective theory. This criterion is sometimes dubbed absence of “asymptotic superluminality” [55]. The perspective of some literature is that any time delay for which there is net negative sign  $\Delta T < 0$ , i.e., a net time advance, leads to a causality violation, since it implies propagation faster than the asymptotic spacetime, regardless of magnitude [53].

In the case of weakly coupled UV completions, it was argued in Ref. [23] that apparent time advances in the low-energy effective theory could be resolved by an infinite tower of higher spins as in the case of string theory. While useful for understanding the nature of possible UV completions, from the perspective of a low-energy physicist this tower of higher spins would just show up as an infinite number of local operators. Clearly including more and more irrelevant operators in the low-energy EFT cannot change the statement that the phase shift (as computed using the leading order operators in the EFT) is of a particular sign at sufficiently low energy. Furthermore since

<sup>1</sup>In nonrelativistic quantum mechanics, the time delay can be negative, as in the case of hard sphere scattering. The original requirement of Wigner is only that the magnitude of the associated time advance is bounded [48].

the low-energy effective theory is designed to describe large distance physics, it should reliably compute the large distance behavior of the retarded propagator. Any time advance in addition to the Shapiro contribution would signal that the retarded propagator has support outside the light cone set by the background geometry. Whether or not the UV completion admits an infinite tower of spins cannot by itself resolve this tension with causality at large distances which is where the real issue lies.

Hence the import of the observation that the UV theory resolves the time advance for the low-energy physicist is merely the generic statement that the low-energy effective theory has a cutoff (set by the lowest mass in the tower for a weakly coupled UV completion), and only a *resolvable* time advance calculable within the regime of validity of the effective theory would signal a true causality violation. This criterion is of course true regardless of whether or not we consider a weakly coupled UV completion. This implies that, as far as the low-energy EFT is concerned, the resolution behind a negative sign of the phase shift in an EFT and its apparent tension with causality cannot lie in the existence of a higher-spin tower *per se* but rather must lie in the actual order of magnitude of the phase shift/time advance itself and the existence of a cutoff irrespectively of what precisely happens at that cutoff (irrespectively on whether or not it represents the onset of a tower of higher spins or whether it represents instead the mass of other heavy particles whose loops are relevant).

Moreover, the causality criterion proposed in [23] is only that the *total* time delay  $\Delta T$  is positive within the regime of validity of the effective theory  $\Delta T > 0$ . This is the essential point of the “asymptotic superluminality” condition [55]; local perturbations and corrections to the effective theory may speed up local propagation relative to the background geometry, but this is viewed as acceptable as long as they remain slower than the asymptotic geometry.

Our perspective is that this is incorrect, or at least incomplete, for two reasons: (a) strict positivity of the total time delay fails to account whether the delay is resolvable, and (b) more generally what is required is (resolvable) positivity of the EFT corrections to the time delay *relative to the background*.<sup>2</sup> This may be seen by a more careful consideration of the well-known example of QED in curved spacetime. It is well known that loop corrections from charged particles induce low-energy superluminal propagation for photons in a curved background, e.g., for a Schwarzschild background, as first noted by Drummond and Hathrell [33]. The total time delay is then

$$\Delta T = \Delta T^g + \Delta T^{\text{EFT}}, \quad (1.3)$$

where  $\Delta T^g$  is the time delay induced by the curved background spacetime, i.e., the Shapiro time delay, and  $\Delta T^{\text{EFT}}$  is in this case the Drummond-Hathrell correction from loop effects. According to [23] this is consistent with causality because it is noted that in order for the negative  $\Delta T^{\text{EFT}}$  to overpower the positive (in dimensions  $D > 4$ )  $\Delta T^g$ , it is necessary that the impact parameter is smaller than the inverse mass of the charged particle integrated out, which means the low-energy effective theory can no longer be trusted. However, this argument does not resolve the causality problem that appears to arise at larger impact parameter where  $\Delta T^{\text{EFT}}$  is negative, and its calculation can be trusted within the regime of validity of the EFT, but the overall  $\Delta T$  is positive. Any negative  $\Delta T^{\text{EFT}}$  implies that the photons are traveling superluminally relative to the background metric (accounted for by the Shapiro contribution) and this sits in clear contradiction with the fact that in the known UV completion, namely QED in curved spacetime, causality remains intact with the causal light cone defined by the background geometry and not the asymptotic Minkowski geometry. Causality in this case should then be a statement about  $\Delta T^{\text{EFT}}$  itself, and not about the full  $\Delta T$ . Yet demanding  $\Delta T^{\text{EFT}} > 0$  sits in contradiction with the well-known result of Drummond and Hathrell.

Hence the “asymptotic” causality condition of [23] fails to address what is the resolution of the apparent causality violation in the low-energy Drummond-Hathrell EFT. This case was discussed in details in Ref. [44]. As we shall demonstrate later, and as was cleanly argued in [43] for calculations in the shockwave (eikonal) limit, the real resolution of causality is that the negative Drummond-Hathrell contribution is *unresolvable*<sup>3</sup> within the effective theory, namely within the regime of validity of the EFT it remains always true that the advance is smaller than the “geometric optics” *resolution scale*

$$|\Delta T^{\text{EFT}}| \ll \omega^{-1}, \quad (1.4)$$

where  $\omega$  is the asymptotic energy of the scattering particle. Equivalently this is the statement that the EFT contribution to the scattering phase shift is bounded by unity

$$|\Delta \delta_\ell^{\text{EFT}}| \ll 1. \quad (1.5)$$

It is therefore clear that by itself the sign of the time-delay correction cannot be sufficient in determining whether or not acausality will follow; it is crucial to consider also its magnitude. Only if there is a time advance, calculable within the EFT, larger than the wavelength of the scattering

<sup>2</sup>At least in cases in which there is a clean  $M_{\text{pl}} \rightarrow \infty$  decoupling limit, and for which the Shapiro contribution is defined in the field frame for which the decoupling limit is well defined, to be discussed in Sec. III B.

<sup>3</sup>Some literature use the cutoff to define resolvability; this is not the definition we will use in what follows for reasons explained around (5.26).

state, can we infer genuine causality violation. The eikonal approximation used in [23]<sup>4</sup> relies on resumming an infinite number of ladder diagrams which are “typically” enhanced as compared to all the other diagrams. Implicitly this resummation amounts to an exponentiation of the lowest order phase shift in the partial wave expansion

$$\delta_\ell = \delta_\ell^g + \Delta\delta_\ell^{\text{EFT}} \rightarrow \frac{1}{2i} (e^{2i(\delta_\ell^g + \Delta\delta_\ell^{\text{EFT}})} - 1). \quad (1.6)$$

More precisely, it is the statement that a resummation of the  $t$ -channel exchange and its associated higher order ladder diagrams appropriately exponentiate in the manner

$$e^{2i(\delta_\ell^g + \Delta\delta_\ell^{\text{EFT}})} \rightarrow e^{2i\delta_\ell^g} \left\{ 1 + 2i\Delta\delta_\ell^{\text{EFT}} + \left( \frac{1}{2} (2i)^2 (\Delta\delta_\ell^{\text{EFT}})^2 + \begin{array}{l} \text{terms of the same order} \\ \text{in EFT expansion} \end{array} \right) \right\}.$$

Although the eikonal resummation is valid for the usual  $\delta_\ell^g$  contributions, we cannot take seriously the exponentiated  $\Delta\delta_\ell^{\text{EFT}}$  contributions as an indicator of high-energy behavior which is necessary in order to interpret them as contributing to a physical time delay. Explicit calculations in the UV theory such as those performed in [43] confirm that the eikonal resummed Drummond-Hathrell result bears no relation to the true high-energy behavior of the scattering phase shift.

We shall argue that this is a more general phenomena, applying equally for GWs and indeed to any EFT. Naive scattering time advances do occur in the low-energy effective theory of gravity, and generic EFTs in curved spacetime, arising for example in Schwarzschild spacetime from matter loops in close analogy with the QED case. However as we show later these time advances are not resolvable and are seen to satisfy

<sup>4</sup>In Ref. [23] as well as in for example the nice recent related discussion of [27] derivations are principally performed within the eikonal approximation which has the virtue of having a relatively clean interpretation in terms of Feynman diagrams as a resummation of ladder diagrams. As such, these derivations are well suited to perturbative S-matrix calculations. All of the calculations we perform here are in the related semiclassical approximation. This approximation is harder to understand in terms of a resummation of Feynman diagrams; however, it is straightforward to calculate it by applying the Wentzel-Kramers-Brillouin (WKB) approximation to the corrected Green’s functions. This is the method we will use in what follows and we refer to Appendix B for more details on the semiclassical approximation. The eikonal approximation may be obtained from a high-energy limit of the semiclassical approximation, as we outline in Appendix C. In the relativistic context the latter may be viewed as a Penrose limit of the former. Due to the close connection many of our statements apply implicitly to both the semiclassical and eikonal approximation methods.

$$\begin{aligned} & \delta_\ell(t\text{-channel}) + \sum \text{ladder diagrams} \\ &= \frac{1}{2i} (e^{2i\delta_\ell(t\text{-channel})} - 1). \end{aligned} \quad (1.7)$$

This resummation does make sense when the enhancement does actually occur, or in other words when the resulting phase shift is  $|\delta_\ell| \gtrsim 1$  so that these terms may be taken large relative to other small corrections. However, as we shall demonstrate, the validity of the low-energy effective theory actually requires  $|\Delta\delta_\ell^{\text{EFT}}| \ll 1$  in the case of QED and we should really replace the exponentiated form with

$$|\Delta T^{\text{EFT}}| \ll \omega^{-1}, \quad (1.8)$$

which is equivalent to the statement that

$$|\Delta\delta_\ell^{\text{EFT}}| \ll 1. \quad (1.9)$$

This shows up generically in the fact that local fluctuations about the background geometry allow for mildly superluminal fluctuations. Once again though, this superluminality is not resolvable, and cannot be used to argue for any causality violations. By contrast, genuinely acausal EFTs lead to resolvable time advances calculable within the regime of validity of the EFT and we give an example later. Consistent EFTs are seen then to self-protect from any macroscopic causality violation, and in this sense contain remnant information of their consistent UV completion. We stress again that the precise nature of the UV completion is immaterial to this particular part of the argument. The previous bounds and “self-protection” mechanism should indeed hold irrespectively of whether we are dealing with a weakly coupled infinite tower of spins, or a more mundane heavy loop contribution as in the Drummond-Hathrell case.

These observations have significant impact on how we put constraints on low-energy effective theories. The overly enthusiastic low-energy physicist who demands that the Wilson action should be constrained by the requirement that all fluctuations around every background should be (sub)luminal relative to the background, or similarly that the scattering time delay correction relative to the background (Shapiro/gravitational) time delay is positive may easily risk ruling out EFTs with consistent Lorentz invariant and causal UV completions. A more nuanced discussion is required that establishes whether either of these effects leads to a macroscopically observable causality violation within the regime of validity of the effective field theory. In what follows we give well-known examples of

both situations. Related discussions in the holographic context were also considered in [56].

The rest of the manuscript is organized as follows: In Sec. II we review why causality and analyticity typically require a subluminal low-energy phase velocity, while pointing out some caveats that occur in curved spacetimes. In Sec. III, we highlight subtleties that arise when dealing with superluminal low-energy speed in gravitational EFTs and the relevance of being able to take a decoupling limit where the gravitational degrees of freedom decouple before being able to restrict low-energy coefficients based on superluminal criteria. We highlight cases where macroscopic superluminalities are allowed (and even sometimes imposed) by analyticity and causality. Such cases are particularly important when attempting to restrict the allowed coefficients in cosmological EFTs. In Sec. IV we show how the small amount of superluminal low-energy speed we expect from the EFT of gravity leads to no physical propagation outside the light cone and is therefore not in contradiction with causality. The same type of arguments is then shown to apply for black hole (BH) spacetimes in Sec. V. These BH spacetimes are asymptotically flat and the connection with the sign and magnitude of the scattering phase shift can be made manifest within the EFT of gravity. The same type of arguments and absence of secular growth is also made explicit in the EFT of QED below the electron mass as highlighted in Sec. VI where we make it clear that a negative phase shift of sufficiently large magnitude to be in tension with causality can never be realized within the regime of validity of the low-energy QED EFT. This is contrary to what occurs in other EFTs where the semiclassical or eikonal approximation can remain under control for sufficiently large phase shift and hence lead to a resolvable physical time advance and be in tension with causality as illustrated in Sec. VII. We end with a summary and discussion in Sec. VIII. Appendix A provides a review of the low-energy EFT for gravity as well as the graviton dispersion relation and the direction of the RG flow. The relations between the semiclassical approximation used to compute the phase shift and time delay are reviewed in Appendix B and the relation to the eikonal approximation is outlined in Appendix C. Finally Appendix D provides useful formulas to compute the EFT corrections to the time delay relative to generic effective backgrounds.

## II. CAUSALITY AND ANALYTICITY

### A. Refractive index

#### 1. Implications from analyticity and unitarity

It is a familiar result that the speed of propagation of a wave in a medium is in general different than in vacuum. For instance for a rotational and translation invariant medium, it is sufficient to describe the propagation speed through the refractive index  $n(\omega)$  for which the phase

velocity is given by  $v_p = c/n(\omega)$ . The speed of propagation of a wavefront is determined by the front velocity  $v_f$  which is given by (from now on we set  $c = 1$ ),

$$v_f = \lim_{\omega \rightarrow \infty} 1/n(\omega). \quad (2.1)$$

Relativistic causality demands that  $v_f \leq 1$ . However this does not preclude the possibility that the low-energy phase or group velocity is superluminal. Superluminal group velocities in particular are a well-studied experimental phenomena [57–60] and do not in any way contradict causality, they rather indicate the failure of group velocity as a useful concept in a dispersive medium. This was recognized long ago by Sommerfeld and Brillouin [58] who resolved the apparent paradox between superluminal group velocities and relativity well before any experimental evidence for this phenomena.

However the front velocity is not the end of the story. The full requirement of causality is that the retarded propagator vanishes outside of the forward light cone,

$$G_{\text{ret}}(x, x') = 0, \text{ for } (x - x')^2 > 0, \text{ or } (t - t') < 0. \quad (2.2)$$

In addition to the front velocity being luminal, the latter generally requires that the refractive index is an analytic function in the upper half complex  $\omega$  plane [61]. Applying Cauchy's theorem assuming *analyticity* leads to the Kramers-Kronig relations which for future comparison is most usefully written as

$$n(\omega) = n(\infty) + \frac{2}{\pi} \int_0^\infty d\omega' \frac{\omega' \text{Im}[n(\omega')]}{(\omega')^2 - (\omega + i\epsilon)^2}. \quad (2.3)$$

A traveling wave moving in the  $z$  direction takes the form  $e^{i\omega t - in(\omega)\omega z}$  and so it is the real part  $\text{Re}[n(\omega)]$  that determines the speed, and the imaginary part the dispersion. Now in a normal medium, *unitarity* demands that the imaginary part of the refractive index is positive:  $\text{Im}[n(\omega)] \geq 0$  for real  $\omega > 0$ . At zero frequency, the real part is given more precisely as

$$\begin{aligned} \text{Re}[n(0)] &= n(\infty) + P \left[ \frac{2}{\pi} \int_0^\infty d\omega' \frac{\omega' \text{Im}[n(\omega')]}{(\omega')^2} \right] \\ &> n(\infty). \end{aligned} \quad (2.4)$$

Hence we conclude a bound on the low-energy phase velocity

$$v_p(0) < v_f. \quad (2.5)$$

Since the front velocity cannot be superluminal  $v_f \leq 1$ , it is typically inferred that the low-energy phase velocity cannot be superluminal unless we violate either (1) analyticity, or (2) unitarity.

## 2. Low-energy effective theory

This particular argument is strengthened if we imagine a situation in which the dispersive imaginary part of the refractive index is only nonzero for frequencies above some scale  $M$ , i.e.,

$$\text{Im}[n(\omega)] \approx 0, \quad 0 \leq \omega < M. \quad (2.6)$$

In this situation, there exists a low-energy effective theory valid for frequencies  $\omega \ll M$  for which the refractive index is well approximated by a Taylor series

$$n(\omega) = n(0) + \sum_{n=1}^{\infty} b_n \frac{\omega^{2n}}{M^{2n}}, \quad (2.7)$$

with leading low-energy term

$$n(0) = n(\infty) + \frac{2}{\pi} \int_M^{\infty} d\omega' \frac{\omega' \text{Im}[n(\omega')]}{(\omega')^2} > n(\infty), \quad (2.8)$$

and positive dimensionless EFT coefficients

$$b_n = M^{2n} \frac{2}{\pi} \int_M^{\infty} d\omega' \frac{\omega' \text{Im}[n(\omega')]}{(\omega')^{2n+2}} > 0. \quad (2.9)$$

The equation of motion that describes the propagation of the wave with amplitude  $\phi(t, \vec{x})$  is of the form

$$n(0)^2 \frac{\partial^2 \phi}{\partial t^2} - \nabla^2 \phi = \sum_{n=1}^{\infty} \frac{c_n}{M^{2n}} \frac{\partial^{2n+2} \phi}{\partial t^{2n+2}}, \quad (2.10)$$

for some dimensionless coefficients  $c_n$ . The higher time derivatives arise here due to the low-energy expansion and do not imply additional states. Indeed within the context of the  $1/M^2$  expansion this equation may be rewritten in the conceptually nicer form

$$\frac{\partial^2 \phi}{\partial t^2} - \frac{1}{n(0)^2} \nabla^2 \phi = \sum_{n=1}^{\infty} \frac{\tilde{c}_n}{M^{2n}} \nabla^{2n+2} \phi. \quad (2.11)$$

In this low-energy regime, there is by assumption no dispersion, and effects from high-energy physics are captured by local higher derivative operators. The leading order group velocity is the same as the phase velocity  $v_g = 1/n(0)$ . Thus low-energy sources propagate at the speed  $1/n(0)$ . If we compute the retarded propagator for  $\phi$  as a perturbative expansion in  $1/M$  then each term at finite order will vanish only outside the forward light cone defined by

$$(t - t')^2 - n(0)^2 (\vec{x} - \vec{x}')^2 > 0, \quad t - t' > 0. \quad (2.12)$$

Thus we must have  $n(0) \geq 1$ , otherwise even a tiny superluminal velocity  $n(0) = (1 + \epsilon)^{-1}$  with  $\epsilon > 0$  would

integrate up over sufficiently long periods of time to an arbitrary large increase  $\Delta x$  in the spatial size of the causal light cone from a given event  $\Delta x = \epsilon |t - t'|$ , leading to causality paradoxes. In running this argument, it is crucial that  $|t - t'|$  may be made arbitrarily large. We will see that when considering the same argument in a Friedmann-Lemaître-Robertson-Walker (FLRW) geometry or on the background of a BH, it is exactly this assumption that breaks for reasons to be explained. Furthermore it is unclear whether (2.5) holds in curved spacetimes due to the generic absence of conventional analyticity.

The fact that the situation is more subtle in curved spacetimes is well known from work on the low-energy effective theory for QED in curved spacetime [33,34] which has been discussed extensively in the literature [35–42], where it is noted that the low-energy phase velocity in a curved spacetime can be superluminal without contradicting the requirement that the front velocity is luminal. Our subsequent discussions in Secs. IV and V and Appendix A will parallel this for the speed of GWs themselves.

## B. Analyticity with gravity

Since one of our principal interests is the speed of gravity, i.e., the speed of GWs in a curved background, we would ideally repeat the argument of the previous section. A knowledge of the spectral properties of the propagator for GWs in a curved spacetime could allow us to infer concrete statements about the low-energy speed. Unfortunately a direct application of these arguments to curved spacetimes is not available since there is no requirement that analyticity should hold in general. For cosmological spacetimes, this is made transparent by the inherent time dependence of the background meaning that frequency  $\omega$  is no longer a good Fourier variable.

Fortunately all information from analyticity in Minkowski spacetime is not completely lost. Consider a particular diffeomorphism invariant low-energy gravitational theory. Analyticity constraints will impose restrictions on the form of low-energy action based on analytic scattering amplitudes in Minkowski spacetime or spectral density requirements. Since the underlying gravitational theory is diffeomorphism invariant, this can immediately be used to infer constraints on covariant operators in the effective Lagrangian which in turn have consequences around curved spacetimes. One set of arguments of this kind are reviewed in Appendix A for which we summarize the essential points here. These arguments are closely related to positivity bound arguments that apply to scattering amplitudes [4–21].

Rather than working with a dispersion relation for the refractive index (2.3), we can determine a Källén-Lehmann type dispersion relation for the gravitational wave propagator on Minkowski spacetime. Due to gauge invariance, this is most conveniently expressed as the exchange

interaction “ $TT$ -amplitude” between two conserved sources with gravitational propagator as given in (A8),

$$\begin{aligned} \Delta S_{TT} = & \int \frac{d^4k}{(2\pi)^4} \left[ \frac{Z}{2M_{\text{Pl}}^2} \frac{(|T_{\mu\nu}(k)|^2 - \frac{1}{2}|T(k)|^2)}{k^2 - i\epsilon} \right. \\ & + \frac{C_2(\mu_0)}{M_{\text{Pl}}^4} \left( |T_{\mu\nu}(k)|^2 - \frac{1}{3}|T(k)|^2 \right) + \frac{C_0(\mu_0)}{M_{\text{Pl}}^4} |T(k)|^2 \\ & + \frac{1}{M_{\text{Pl}}^4} \int_0^\infty d\mu \rho_2(\mu) \frac{(\mu - k^2)}{(\mu + \mu_0)} \frac{(|T_{\mu\nu}(k)|^2 - \frac{1}{3}|T(k)|^2)}{\mu + k^2 - i\epsilon} \\ & \left. + \frac{1}{M_{\text{Pl}}^4} \int_0^\infty d\mu \rho_0(\mu) \frac{(\mu - k^2)}{(\mu + \mu_0)} \frac{|T(k)|^2}{\mu + k^2 - i\epsilon} \right]. \quad (2.13) \end{aligned}$$

Despite being a different quantity, (2.13) is conceptually similar to (2.3) where the functions  $\rho_2(\mu)$  and  $\rho_0(\mu)$  are positive by unitarity, and are analogous to  $\text{Im}[n(\omega)]$ , and the subtraction constants  $C_2(\mu_0)$  and  $C_1(\mu_0)$  analogous to  $n(\infty)$ . As for the standard Källén-Lehmann spectral representation, the momentum space argument is an analytic function of complex momentum squared  $s = -k^2 + i\epsilon$  up to a pole at  $k^2 = 0$  and a right-hand branch cut.<sup>5</sup> In defining (2.13) we have introduced an arbitrary subtraction scale even though the result is independent of that scale. This is encoded in the renormalization group style equation (see Appendix A for derivation)

$$\mu_0 \frac{d}{d\mu_0} C_S(\mu_0) = - \int_0^\infty d\mu \rho_S(\mu) \frac{\mu_0}{(\mu + \mu_0)^2} < 0. \quad (2.14)$$

Although  $\mu_0$  is not the sliding scale or cutoff of the usual renormalization group, it encodes the same flow, and we see that unitarity demands positivity of the flow from the UV into the IR, that is

$$C_S^{\text{IR}} > C_S^{\text{UV}}. \quad (2.15)$$

While this does not constitute a proof, it certainly leads to the expectation that  $C_S^{\text{IR}} > 0$ . Attempts to prove this have been given in [62] for the case of gravity coupled to a Maxwell field via S-matrix positivity arguments.<sup>6</sup> Specifically positivity of  $C_2^{\text{IR}}$  would follow if the graviton scattering amplitude with the massless  $t$ -channel pole removed has a positive second  $s$  derivative in the forward scattering limit. Unfortunately issues with this proof have

<sup>5</sup>The branch cut lies on the real axis for  $-k^2 \geq 0$  with the physical value understood to be the limit from the upper half complex plane, hence  $s = -k^2 + i\epsilon$ .

<sup>6</sup>Since the curvature squared corrections associated with  $C_S$  can be removed with field redefinitions, the S-matrix constraints of [62] are strictly speaking applied to the  $F^4$  and  $F^2 R$  terms in the Einstein-Maxwell EFT. However, if we take the perspective that the coefficients of these operators are zero before the field redefinition, then positivity implied by the arguments of [62] would indeed infer that  $C_2^{\text{IR}} > 0$ . The Källén-Lehmann dispersive arguments are clearly weaker than the S-matrix bounds, since the former are sensitive to field frame.

recently been pointed out in [63], together with counterexamples which undermine the requirement  $C_2^{\text{IR}} > 0$ .

What does this have to do with GWs propagating in curved spacetimes? The answer is that the coefficients  $C_S^{\text{IR}}$  determine precisely the coefficients of the covariant curvature squared terms in the low-energy effective theory, and they in turn determine whether GWs travel super or subluminal. If we begin with the tree level Wilsonian effective action which includes the leading curvature squared terms expected to arise from integrating out loops and higher spin heavy states [see Eq. (A19)], the above dispersive arguments enforce positivity of the coefficients of  $R^2$  and  $R_{\mu\nu}^2 - R^2/3$  or equivalently stated positivity of the Weyl curvature squared term  $W_{\mu\nu\rho\sigma}^2$ . We may then use the resulting local equations for the low-energy effective theory to determine an effective equation for the propagation of low-energy GWs precisely analogous to (2.11). By direct analogy we can define the low-energy speed by the refractive index coefficient  $n(0)$  that relates the leading “two derivative” part of this effective equation. The central result of [45] is that the low-energy refractive index  $n(0)$  is determined at leading order by  $C_2^{\text{IR}}$ , and the sign of the latter directly determines the sub- or superluminality of the low-energy speed of propagation. In situations where the leading curvature squared terms vanish, such as Schwarzschild spacetime, it is the curvature cubed terms which determine the leading effect, as calculated in [45] and considered in detail in [46].

### III. DEALING WITH SUPERLUMINALITY IN GRAVITATIONAL SETUPS

#### A. Nongravitational criterion

Before getting to our main discussion on the speed of GWs we will review here several key points that arise when dealing with EFTs on a background that spontaneously break Lorentz invariance. Indeed, it is straightforward to write down Lorentz invariant EFTs which exhibit superluminal propagation around spontaneously broken Lorentz backgrounds. The canonical example is that of a  $P(X)$  model (ignoring gravity for now and setting the field on Minkowski), for which the nonminimal kinetic term takes the form [64]

$$\mathcal{L} = P(X) = X + \frac{a}{M^4} X^2 + \mathcal{O}(X^3/M^8) \quad (3.1)$$

with  $X = -\frac{1}{2}(\partial\phi)^2$ . For the “wrong” sign choice  $a < 0$ , this leads to superluminal propagation around a simple time-dependent background<sup>7</sup>  $\phi = \phi(t)$ . Indeed the sound speed about such a background is given by

<sup>7</sup>The same EFT with  $a < 0$  also leads to superluminal propagation around any other background that spontaneously breaks Lorentz invariance by picking up a preferred direction for  $\partial_\mu\phi$  no matter whether  $\partial_\mu\phi$  is timelike or spacelike.

$$c_s^2 = 1 - 4a \frac{\dot{\phi}^2}{M^4} + \mathcal{O}\left(\frac{\dot{\phi}^4}{M^8}\right), \quad (3.2)$$

and is superluminal for  $a < 0$ . This is connected with a violation of positivity bounds [4–6].

The departure of the speed from unity is always suppressed by factors of  $\dot{\phi}^2/M^4$  and is therefore always small within the regime of validity of the EFT. Nevertheless, even such a small correction can lead to significant macroscopic consequences. This is because in this field theory on Minkowski spacetime setting, there is no limit to how long we may wait to integrate this effect up, and even a small local effect can therefore build up to a macroscopic size. Roughly speaking, at a time  $t$ , the future light cone emanating from a given spacetime point at  $t_0$  will be larger by a radius

$$\Delta r = \int_{t_0}^t dt |c_s - 1| \sim \frac{2(-a)\dot{\phi}^2}{M^4} |t - t_0|, \quad (3.3)$$

and the smallness of  $\dot{\phi}/M^2$  may easily be compensated for by the largeness of  $|t - t_0|$  which is otherwise unrestricted. Since this distance will be macroscopically observable, which means not only will it satisfy  $\Delta r \gg M^{-1}$ , crucially it is resolvable  $\Delta r \gg \lambda$  where  $\lambda$  is the wavelength of the propagating fluctuation, we are in a situation where we may imagine violations of causality. At the very least, this would imply a violation of causality as implied by the (asymptotic) Lorentz invariant light cone.

Indeed stated this way, it is clear that *any EFT in Minkowski spacetime* that leads to *any* amount of homogeneous superluminality, no matter how small, will necessarily lead to macroscopic causality violation after a long enough time. Since there is no restriction on how long we can wait, then we must conclude that standard relativistic causality imposes the strict requirement that<sup>8</sup>

$$c_s \leq 1. \quad (3.4)$$

Our central claim is that the same argument *does not* apply to the low-energy speed in a curved spacetime or gravitational setting for a number of reasons. As we have already alluded to, in gravitational theory (in a setup that spontaneously breaks Lorentz invariance), the question is more subtle for two reasons: first, the notion of low-energy speed (i.e., any speed inferred from a low-energy effective theory) is not frame independent and second, due to the nature of the spacetime in question, it may not be possible to integrate up a small departure in speed to make a macroscopic effect. In other words, the would-be superluminality may not be resolvable. When this is the case we cannot

<sup>8</sup>This is more subtle for inhomogeneous configurations where local time advances can be canceled by neighboring local time delays to lead to a net time delay.

necessarily legitimately conclude that there is any causality violation in the regime of validity of the effective theory. In fact we will see that in the two most interesting curved spacetimes, namely FLRW and Schwarzschild, this is exactly what happens.

When computed from the leading terms in a low-energy EFT, the low-energy speed may not manifest any explicit frequency dependence. This occurs in particular if only at most second order derivative terms have been included in the low-energy EFT and any other higher derivative term has been considered as irrelevant, as was the case in the  $P(X)$  EFT considered previously in (3.1). We stress however that this is always an artifact of the truncation to the leading EFT interactions. However from the very definition of a low-energy EFT, (3.2) can only give an appropriate description for  $c_s$  at low energies, in this case at frequencies of at most  $\omega \ll M^2/|\dot{\phi}|$ . Beyond the regime of validity of the low-energy EFT, the computation of the speed is simply not valid. Even if the formula does not appear to break down mathematically at that order, it does break down physically when flowing to higher energy as higher order irrelevant operators ought to be taken into account, until ultimately the theory ought to be traded for its higher energy counterpart, explicitly including heavier modes. See Ref. [65] for a discussion.

## B. Decoupling limit

In many situations, in a given gravitational effective field theory, it may be possible to take a decoupling limit<sup>9</sup>  $M_{\text{Pl}} \rightarrow \infty$ , keeping some other interaction scale  $M$  in the system fixed, for which the helicity-2 gravitational degrees of freedom decouple from all other degrees of freedom while maintaining the interactions that arise at the lowest energy scale  $M$ . Whenever this is possible, the resulting decoupled effective theory can be analyzed from the perspective of an interacting field theory on a fixed (Minkowski or other curved) background, and in this situation the above argument (3.4) is expected to be valid. With this in mind we may then declare that *if in the field frame for which the decoupling limit is well defined*

$$\lim_{\substack{M_{\text{Pl}} \rightarrow \infty, \\ \text{fixed } M}} c_s \leq 1, \quad (3.5)$$

for all species then causality is (expected to be) satisfied. The condition on the *field frame* is crucial as we will see in Sec. III C. More precisely we will see that rate to which the effect goes to zero as  $M_{\text{Pl}} \rightarrow \infty$  is crucial, and a more refined version of this statement is the bound (3.8).

<sup>9</sup>A decoupling limit field theory should not be confused with the original Lagrangian simply evaluated on a Minkowski background. See for instance Refs. [60,66] for discussions on the physical meaning of a decoupling limit.



### 1. Superluminal speed in the decoupling limit

As a first example, we may consider the gravitational version of our canonical example in the previous Sec. III A. Promoting the previous example to a gravitational effective field theory

$$\mathcal{L} = \sqrt{-g} \left( \frac{M_{\text{Pl}}^2}{2} R + X + \frac{a}{M^4} X^2 \right), \quad (3.6)$$

it is straightforward to take the limit  $M_{\text{Pl}} \rightarrow \infty$  with  $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}/M_{\text{Pl}}$  keeping the scale  $M$  fixed. In this limit, we are left with two decoupled sectors, on one side a free massless spin-2 degree of freedom  $h_{\mu\nu}$  and on the other side an interacting  $P(X)$  scalar field theory on Minkowski identical to that considered in Sec. III A to which the usual superluminality and positivity bounds violation arguments apply.

### 2. Luminal speed in the decoupling limit

On the other hand, we may now consider a modification to the low-energy speed that is parametrically suppressed by powers of  $M^2/M_{\text{Pl}}^2$  relative to the previous effect, take for instance

$$c_s^2 = 1 + \frac{|a|\dot{\phi}^2}{M_{\text{Pl}}^2 M^2} + \dots \quad (3.7)$$

As we shall see, this is closer to the typical situation for the low-energy speed of GWs with corrections as in [45]. In this case it is not possible to take the limit  $M_{\text{Pl}} \rightarrow \infty$  and have  $c_s$  differ from unity without something else blowing up. For instance we may try to scale  $M \sim 1/M_{\text{Pl}}$  but with the understanding that  $M$  sets the scale of other irrelevant operators in the effective theory, this would inevitably lead to a break down of the low-energy EFT at arbitrarily low scales, invalidating the calculation of the speed.

Our central claim is that whenever a situation like (3.7) occurs where  $M$  is related to the cutoff of the low-energy EFT in which the speed has been computed (or to the scale of irrelevant operators), then the condition (3.5) is actually satisfied and it would then not be legitimate to demand that  $c_s \leq 1$  for all degrees of freedom away from the decoupling limit (i.e., at finite  $M_{\text{Pl}}$ ). Specifically we will see that in actual examples, the low-energy speed  $c_s$  is typically expected to be superluminal without leading to any macroscopic violation of causality. The key to this is the smallness of the effect, and this is in turn tied to the fact that the speed is luminal in the limit  $M_{\text{Pl}} \rightarrow \infty$ . As long as (3.5) continues to hold, we do not anticipate any violation of causality. More precisely we shall see that if

$$\lim_{M_{\text{Pl}} \rightarrow \infty} |c_s^2 - 1| \sim M_{\text{Pl}}^{-\alpha}, \quad (3.8)$$

again in the field frame for which the decoupling limit is well-defined, with  $\alpha \geq 2$ , there will be no macroscopic observable effects. A single  $M_{\text{Pl}}$  suppression would not be sufficient, however all gravitational induced corrections to the sound speed arise at a minimum with a  $M_{\text{Pl}}^2$  suppression (in a local theory).

### C. Macroscopic superluminality allowed by analyticity

In order to illustrate the subtleties that emerge when dealing with superluminalities in a gravitational theory, we give here an example of an effective theory for which superluminal GWs are required by analyticity. Consider the effective Lagrangian of the form

$$\mathcal{L}_{\text{EFT}} = \sqrt{-g} \left( \frac{M_{\text{Pl}}^2}{2} R + a \frac{M_{\text{Pl}}^2}{2M^4} G^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi - \frac{1}{2} (\nabla \phi)^2 + \dots \right), \quad (3.9)$$

which includes a nonminimal coupling between gravity and the scalar. It is straightforward to show that on considering perturbations around a cosmological solution sourced by a time-dependent scalar  $\phi(t)$ , the GWs propagate superluminally if  $a > 0$  and subluminally if  $a < 0$ . Furthermore this effect is a macroscopically observable one since

$$\Delta c_s^2 \sim a \frac{\dot{\phi}^2}{M^4}. \quad (3.10)$$

Surely then, since this superluminality is macroscopically observable, causality/analyticity considerations will demand that  $a < 0$ ? In fact it is straightforward to see that this is not the case and the precise opposite actually holds. Indeed, one can change frame so that the Lagrangian (3.9) exactly matches (3.6) by the simple redefinition

$$g_{\mu\nu} \rightarrow g_{\mu\nu} + a \frac{1}{2M^4} \nabla_\mu \phi \nabla_\nu \phi, \quad (3.11)$$

where standard positivity bounds following from analyticity impose that  $a > 0$  as the ‘‘causal’’ choice. In terms of the canonical normalized gravitational fluctuations  $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}/M_{\text{Pl}}$  this is

$$h_{\mu\nu} \rightarrow \tilde{h}_{\mu\nu} = h_{\mu\nu} + a \frac{M_{\text{Pl}}}{2M^4} \nabla_\mu \phi \nabla_\nu \phi. \quad (3.12)$$

This example nicely illustrates two points: (a) the ambiguity of speed under field redefinitions and (b) the importance of causality constraints being imposed in the frame in which the decoupling limit is well defined. Indeed, unlike the field frame implicit in (3.6) for which the decoupling limit  $M_{\text{Pl}} \rightarrow \infty$  is well defined, the Lagrangian (3.9) does not have a well-defined decoupling limit. Indeed in taking the limit  $M_{\text{Pl}} \rightarrow \infty$  in this frame we would have

$$\lim_{M_{\text{Pl}} \rightarrow \infty} \frac{M_{\text{Pl}}^2}{M^4} G^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi \sim \frac{M_{\text{Pl}}}{M^4} \partial \partial h \partial \phi \partial \phi \rightarrow \infty. \quad (3.13)$$

Stated equivalently, the field redefinition (3.12) blows up in the decoupling limit explaining the inequivalence of the two frames. We can make (3.12) and (3.13) finite in the limit  $M_{\text{Pl}} \rightarrow \infty$  by scaling  $M = M_{\text{Pl}}^{1/4} (\tilde{\Lambda})^{3/4}$  keeping  $\tilde{\Lambda}$  fixed, but then the effect (3.10) vanishes as  $M_{\text{Pl}} \rightarrow \infty$ .

The lesson to learn from this is that simply demanding that for a given EFT the speed of propagation of fields is (sub)luminal in a given field frame is not only unjustified, it may even explicitly violate the requirements that do come from causality. For this reason, the only safe requirement to impose on the EFT is that given in (3.5) which only applies in a field frame for which a decoupling limit exists.

## IV. COSMOLOGY IN THE EFT OF GRAVITY

### A. Speed of gravity in cosmology

Let us now focus our discussion on the specific case of cosmological spacetimes. As reviewed in Appendix A, the leading corrections to the low-energy EFT for gravity may be expressed in the form

$$\mathcal{L}^{\text{EFT}} = \sqrt{-g} \left( \frac{M_{\text{Pl}}^2}{2} R + C_{R^2}^{\text{IR}} R^2 + C_{W^2}^{\text{IR}} W_{\mu\nu\alpha\beta}^2 + C_{\text{GB}} \text{GB} \right) + \text{higher derivative terms}, \quad (4.1)$$

where GB designates the Gauss-Bonnet term and  $W$  the Weyl tensor. In addition when matter is included, we may allow for nonminimal matter curvature interactions, as for example  $RF^2$  terms in the case of Einstein-Maxwell. In order to focus on the genuine gravitational interactions, we shall not consider these nonminimal matter interactions in what follows (including them can lead to additional sources of superluminalities that can be dealt with in a more standard way). Within the low-energy EFT, one cannot determine the sign of the coefficients  $C_{R^2, W^2}^{\text{IR}}$  but as argued in Appendix A the positivity of the RG flow (A11) implies

$$C_{W^2}^{\text{IR}} > C_{W^2}^{\text{UV}}, \quad C_{R^2}^{\text{IR}} > C_{R^2}^{\text{UV}}, \quad (4.2)$$

and in what follows we shall make the *a priori* not-so-unreasonable assumption that  $C_{W^2}^{\text{IR}}$  may be positive (see Ref. [63] for a more precise discussion and potential caveats).

Given a covariant action of the local form (4.1) encoding the EFT corrections, it is straightforward to compute the corrections to the equation of motion for tensor fluctuations on a cosmological background as done in [45]. Identifying what we mean by speed is however slightly subtle since the truncated equation of motion contains higher time derivatives. These may be removed with field redefinitions and traded for space derivatives just as in the conversion from (2.10) to (2.11). After this is done, the equation of motion

for tensor GWs on FLRW may, by virtue of symmetry, be put in the following form:

$$\partial_\eta^2 h = - \sum_{n=0}^{\infty} \beta_n(\eta) k^{2n} h, \quad (4.3)$$

where we work in conformal time  $ds^2 = a(\eta)^2 (-d\eta^2 + d\vec{x}^2)$ . The previous relation may be rearranged and express in the form

$$\partial_\eta^2 h + m_{\text{eff}}^2 h + \tilde{c}_s^2(k, \eta) k^2 h = 0, \quad (4.4)$$

where  $m_{\text{eff}}^2 = \beta_0(\eta)$  is an effective mass and  $\tilde{c}_s^2(k, \eta) = \sum_{n=1}^{\infty} \beta_n(\eta) k^{2n-2}$  an effective  $k$ -dependent sound speed. We then define the low-energy sound speed to be  $c_s^2(\eta) \equiv \tilde{c}_s^2(0, \eta) = \beta_1(\eta)$ , namely the speed of propagation implied by the truncated equation

$$\partial_\eta^2 h + m_{\text{eff}}^2 h + c_s^2(\eta) k^2 h = 0. \quad (4.5)$$

Explicit calculation using (A19) gives the *low-energy* speed [45]

$$c_s^2 = 1 - \frac{16C_{W^2}\dot{H}}{M_{\text{Pl}}^2}. \quad (4.6)$$

Since the null energy condition requires  $\dot{H} < 0$ , then  $C_{W^2} > 0$  would imply that this low-energy speed is slightly superluminal. Note that we do not include the effective background-generated mass  $m_{\text{eff}}$  in consideration of the speed of propagation, because what is relevant is the causal support of the retarded propagator. If (4.5) were the exact equation this would be determined by  $c_s^2(\eta)$  alone [67]. This is in agreement with what is typically meant by the low-energy speed. If the retarded propagator for the exact equation (4.3) is determined as a perturbative expansion in spatial derivatives, with the (up to) two derivative terms (4.5) taken as the leading part, then at any finite order in perturbations the causal support for the retarded propagator will be determined by (4.5). Clearly the relevant question is, when is this a good indication of the true causal support of the exact retarded propagator?

### B. Validity of EFT in FLRW

Before proceeding, we need to address the conditions for the validity of the EFT, e.g., the validity of Eq. (4.3) to describe the evolution of GWs. In the application of the Wilsonian effective theory with cutoff  $M$  we can only describe momenta for which covariant operators are small relative to the cutoff scale, e.g.,  $\square \ll M^2$  and  $R \ll M^2$ . In the cosmological context, since Lorentz invariance is broken by the background, this means we can only use the effective theory to describe the evolution of modes in the region where

$$|\dot{H}(t)| \frac{|\vec{k}|^2}{a(t)^2} \ll M^4, \quad (4.7)$$

i.e.,  $k/a(t) \ll \Lambda_c(t)$  where  $\Lambda_c = M^2/\sqrt{-\dot{H}}$ . In the typical situation in which  $\dot{H} \sim \mathcal{O}(H^2)$  this may be stated as

$$\frac{|H(t)||\vec{k}|}{a(t)} \ll M^2. \quad (4.8)$$

Note that this scale is much higher<sup>10</sup> than that typically considered in trans-Planckian type arguments where it would be argued that the EFT breaks down when  $k \sim a(t)M$  [68,69]. The reason being is that we assume the underlying theory is Lorentz invariant, and so we require a locally Lorentz invariant combination to be comparable to  $M^2$ . In the cosmological context where only time translations are broken, we may for example decompose the Ricci tensor in the manner

$$R_{\mu\nu} = \Omega^2 g_{\mu\nu} + \kappa_\mu \kappa_\nu, \quad (4.9)$$

where  $\kappa_\mu$  is a non-normalized timelike vector (since we are dealing with cosmology here). Given an on-shell wave of momentum  $k_\mu$  for which  $k_\mu k^\mu \approx 0$ , then (4.8) is the locally Lorentz invariant bound

$$|\kappa^\mu k_\mu| \ll M^2. \quad (4.10)$$

This may be taken together with the requirement that  $\kappa_\mu \kappa^\mu \ll M^2$  and  $|\Omega| \ll M$  which require  $H^2 \ll M^2$  and  $|\dot{H}| \ll M^2$ . Indeed the argument for why (4.7) is the more general condition and not (4.8) is that de Sitter invariance in the limit  $\dot{H} \rightarrow 0$  is sufficient to ensure validity of EFT at arbitrarily on-shell high momenta.

To clarify this, let us think of a typical example EFT organized in the standard manner where all irrelevant operators are suppressed by the common scale  $M$  to the appropriate power. Schematically the effective action takes the form<sup>11</sup>

$$\begin{aligned} S_{\text{EFT}} &= \frac{M_{\text{Pl}}^2}{2} \int d^4x \sqrt{-g} \\ &\times \left( R + M^2 \sum \alpha_{ab} \left( \frac{\nabla}{M} \right)^{2a} \left( \frac{\text{Riemann}}{M^2} \right)^b \dots \right), \end{aligned} \quad (4.11)$$

<sup>10</sup>Assuming the much tighter requirement  $k \sim a(t)M$  typically considered in trans-Planckian arguments would only help with our following argumentation.

<sup>11</sup>This is, for example, the schematic form of curvature dependence in the low-energy EFT for string theory in which  $M$  is the string scale  $1/\sqrt{\alpha'}$  which is parametrically below the Planck scale [70,71].

with the usual understanding that we allow for all local scalar operators constructed out of the appropriate number of powers of the Riemann tensor and covariant derivatives in any order. Given the underlying locality and Lorentz invariance, any term in the effective dispersion relation for GWs around a curved background not of the local Lorentz invariant form  $(\omega^2 - k^2/a^2)$  will necessarily come suppressed by some power of the background curvature quantities  $H^2$ ,  $\dot{H}$  and derivatives thereof. Since these terms spontaneously break Lorentz invariance they may come together with  $k^2/a^2$  and  $\omega^2$  terms and so will naturally package in dimensionless combinations of the form

$$\frac{\dot{H}k^2/a^2}{M^4} \sim \frac{\dot{H}\omega^2}{M^4}. \quad (4.12)$$

For example these will arise from terms in (4.11) with factors of  $M^{-4}R^{\mu\nu}\nabla_\mu\nabla_\nu$ . Terms with the same number of powers of curvature, but higher powers of  $k$  such as  $\frac{\dot{H}k^4/a^4}{M^6}$  will necessarily only arise is the quasi-Lorentz invariant combination

$$\frac{\dot{H}k^2/a^2}{M^6}(\omega^2 - k^2/a^2), \quad (4.13)$$

as for example coming from terms like  $M^{-6}R^{\mu\nu}\nabla_\mu\nabla_\nu\Box$  (acting for instance on the scalar curvature). This is in essence due to index contraction, if we limit ourselves to a fixed number of powers of curvature, since Weyl is zero for FLRW, once the Ricci tensor indices have been contracted, all remaining indices must be contracted with the metric which locally takes a Lorentz invariant form. Thus schematically the effective form of the dispersion relation will be (in terms of the physical momentum  $\tilde{k} = k/a$ )

$$\begin{aligned} &\omega^2 - \tilde{k}^2 - m_{\text{eff}}^2/a^2 \\ &+ M^2 \sum \beta_{a_1 \dots a_7} \left[ \left( \frac{\omega^2 - \tilde{k}^2}{M^2} \right)^{a_1} \left( \frac{1}{M^2} \frac{d^2}{dt^2} \right)^{a_2} \left( \frac{H}{M^2} \frac{d}{dt} \right)^{a_3} \right. \\ &\times \left. \left( \frac{H^2}{M^2} \right)^{a_4} \left( \frac{\dot{H}}{M^2} \right)^{a_5} \left( \frac{\dot{H}\tilde{k}^2}{M^4} \right)^{a_6} \left( \frac{\dot{H}\omega^2}{M^4} \right)^{a_7} \right] = 0. \end{aligned} \quad (4.14)$$

The condition that the EFT remains under control requires that at a minimum the  $\beta$  corrections are small or more precisely that the  $\beta$  series is at least convergent in an asymptotic series sense. Since we are allowed arbitrary integer powers of the  $a_i$ , this will only be true if each of the dimensionless ratios in brackets are kept smaller than unity. Hence in addition to the expected requirements that the curvature remains small,  $H^2 \ll M^2$  and  $|\dot{H}| \ll M^2$ , we infer that

$$\frac{|\dot{H}|\tilde{k}^2}{M^4} = \frac{|\dot{H}|k^2/a^2}{M^4} \ll 1. \quad (4.15)$$

This implies that in the typical situation for which  $\dot{H} \sim \mathcal{O}(H^2)$ , the momentum cutoff appropriate for an on-shell state, i.e., a propagating gravitational wave, is therefore as specified in (4.8). Due to redshifting in the cosmological context of an expanding Universe, the bound (4.7) is strongest at the earliest times which is where we shall make use of it.

### C. Causality constraint

While it is known and observed that in many media the low-energy phase and group velocities may temporarily become superluminal, this is only in conflict with causality if the superluminality may be integrated up to a macroscopic effect for which the light cone is clearly larger than the Lorentzian light cone. One way to characterize this in asymptotically flat spacetimes is to ask whether there is an ‘‘asymptotic superluminality’’ [23,55]. In practice, this amounts to asking whether there can be an integrated time advance in a scattering event, which would imply that the signal from a scattering process could arrive before that of an unscattered wave—in a Lorentz invariant theory this would then be associated to some type of acausality. In the cosmological context (or any other curved geometry which is not asymptotically flat), we do not have such a clean tool, and any S-matrix calculation of this form would only be approximately valid at subhorizon scales. We can however ask, by virtue of the symmetry of the FLRW spacetime, how much larger the light cone of propagation is emanating from some event after many Hubble times. On first sight, we may imagine that even the tiniest amount of superluminality in the low-energy phase could be integrated up to some large observable effect over the entire age (or even future) of the Universe. Crucially, this is not the case as we now explain.

Let us work with the effective metric seen by GWs in the EFT of gravity, i.e., an effective metric with speed  $c_s(t)$  as in (4.6). We now consider the future light cone emanating from a spacetime event at time  $t_i$  as determined with respect to this effective metric. If  $C_{W^2} > 0$  then at a given time  $t > t_i$  this light cone is larger than the usual FLRW light cone by a radial distance  $\Delta r$

$$\Delta r(t) = a(t) \int_{t_i}^t \frac{dt'}{a(t')} (c_s(t') - 1). \quad (4.16)$$

Using (4.6) at leading order in the EFT expansion this distance is

$$\Delta r(t) = a(t) \int_{t_i}^t \frac{dt'}{a(t')} \left( -\frac{8C_{W^2}\dot{H}(t')}{M_{\text{Pl}}^2} \right) + \dots \quad (4.17)$$

In an expanding Universe, the integrand on the right-hand side is bounded by

$$\Delta r(t) \leq \frac{a(t)}{a(t_i)} \int_{t_i}^t dt' \left( -\frac{8C_{W^2}\dot{H}(t')}{M_{\text{Pl}}^2} \right) + \dots \quad (4.18)$$

This implies

$$\Delta r(t) \leq \frac{8C_{W^2}}{M_{\text{Pl}}^2} \frac{a(t)}{a(t_i)} (H(t_i) - H(t)) \leq \frac{8C_{W^2}}{M_{\text{Pl}}^2} \frac{a(t)}{a(t_i)} H(t_i), \quad (4.19)$$

given  $H(t) < H(t_i)$  for  $t > t_i$ , assuming the null energy condition is satisfied.

#### 1. Postinflation period

For a mode of a given comoving momentum  $k$ , the earliest time at which we can trust the EFT calculation of the speed is set by (4.8) to be such that [assuming that for most of cosmic history  $\dot{H} \sim \mathcal{O}(H^2)$  which is true postinflation]

$$\frac{H(t_i)}{a(t_i)} \ll \frac{M^2}{|\vec{k}|}, \quad (4.20)$$

from which we infer

$$\Delta r(t) \leq \frac{8C_{W^2}}{M_{\text{Pl}}^2} \frac{a(t)}{k} \ll \frac{4C_{W^2}M^2}{\pi M_{\text{Pl}}^2} \lambda(t), \quad (4.21)$$

where  $\lambda(t) = 2\pi a(t)/k$  is the physical wavelength. Finally the cutoff of the EFT should be at most  $M^2 \lesssim M_{\text{Pl}}^2/C_{W^2}$ , so the bound essentially becomes

$$\Delta r(t) \ll \lambda(t). \quad (4.22)$$

We recall that  $\Delta r$  represents the distance that low-energy GWs may propagate outside the light cone set by the FLRW background metric (i.e., the light cone seen by minimally coupled fields) if  $C_{W^2} > 0$ . For any GW this distance is always much less than the actual physical wavelength  $\lambda$  of the GW (if we ensure that we remain within the regime of validity of the low-energy EFT). This distance is therefore not resolvable, and if it ever were resolvable, one would not be able to trust the result as it would rely on applying the low-energy EFT beyond its regime of validity. Thus causality remains intact provided we limit ourselves to asking questions that are fully within the regime of validity of the EFT.

#### 2. Quasi-inflationary period

The situation is only slightly more subtle when there is a quasi-inflationary (or late-time acceleration) period for which  $|\dot{H}| \ll H^2$ . Consider for example a constant equation of state  $w \approx -1$  for which the scale factor takes the form  $a(t) = a(t_b)(t/t_b)^{2/(3(1+w))}$ . The additional contribution to the comoving displacement coming from an inflationary epoch  $t_e > t > t_b$  is

$$\begin{aligned}
& \int_{t_b}^{t_e} \frac{dt'}{a(t')} \left( -\frac{8C_{W^2}\dot{H}}{M_{\text{Pl}}^2} \right) \\
&= \frac{8C_{W^2}}{M_{\text{Pl}}^2} \frac{1}{(1 + \frac{3}{2}(1 + \omega))} \frac{1}{a(t_b)t_b} \left( 1 - \left( \frac{t_b}{t_e} \right)^{\frac{5+3\omega}{2(1+\omega)}} \right) \\
&\lesssim \frac{8C_{W^2}}{M_{\text{Pl}}^2} \frac{\sqrt{3(1+\omega)}/2}{(1 + \frac{3}{2}(1 + \omega))} \frac{\sqrt{-\dot{H}(t_b)}}{a(t_b)}. \quad (4.23)
\end{aligned}$$

This is further suppressed by  $\sqrt{1+\omega}$  relative to the previous estimate, and so on applying the condition (4.7) at the beginning of inflation we are led to the same conclusion (4.22).

#### D. Suppression is key

The key as to why causality is not being violated by the superluminal speed here is the smallness ( $\dot{H}/M_{\text{Pl}}^2$ ) of the effect, i.e., the gravitational suppression. To clarify this let us return to the case of a genuinely acausal example as in the “wrong” sign  $P(X)$  model with  $a = -|a|$  [64], i.e.,

$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 - \frac{|a|}{4M^4}(\partial\phi)^4 - V(\phi), \quad (4.24)$$

then the speed of propagation for the scalar about say a time-dependent background  $\phi(t)$  is given in (3.2).

Although the departure from unity for the speed is small in an EFT sense, its macroscopic secular effect can be arbitrary large, even on an FLRW background. To illustrate this, let us suppose that this scalar is also the dominant source for the background expansion, then the leading order Raychaudhuri equation is  $\dot{H} = -\dot{\phi}^2/(2M_{\text{Pl}}^2) + \dots$  and so

$$c_s^2 = 1 + 8|a| \frac{(-\dot{H})M_{\text{Pl}}^2}{M^4}. \quad (4.25)$$

Assuming order unity Wilson coefficient  $a \sim \mathcal{O}(1)$ , the departure from luminality is larger by a factor of  $M_{\text{Pl}}^2/M^2$  compared to the previous example. We infer a maximal displacement of the light cone of order

$$\Delta r(t) \sim \frac{M_{\text{Pl}}^2}{M^2} \lambda(t). \quad (4.26)$$

Whenever  $M \ll M_{\text{Pl}}$  we can engineer a situation where there is an observable violation for causality with  $\Delta r(t) \gg \lambda(t)$ , justifying the inherent acausality of the wrong sign  $P(X)$  model. By comparison, in the low-energy EFT for gravity, the corrections to the speed is suppressed by an additional factor of  $M^2/M_{\text{Pl}}^2$  [or in the BH case of Sec. V by a factor  $(M/M_{\text{Pl}})^4$ ] relative to (4.25), which is precisely what makes the displacement unobservable. It is clear from (4.26) that we need at least two powers of  $M_{\text{Pl}}$  suppression to ensure the unobservability of this effect (3.8), justifying the claim made in (3.8).

#### E. Secular effects

To restate the previous results slightly differently, suppose we tried to infer the retarded Green’s function describing the response of a field to a source. From the exact equation (4.4) the momentum space retarded Green’s function may be defined by

$$[\partial_\eta^2 + m_{\text{eff}}^2 + c_s^2(k, \eta)k^2]G_{\text{ret}}(\eta, \eta') = \delta(\eta - \eta'). \quad (4.27)$$

Ideally this equation would be solved exactly; however, we only know its form within the context of an EFT expansion. The picture closest to the classical one is where we infer the propagator by means of a WKB approximation as discussed for example in [72]. The exact retarded propagator is given by

$$G_{\text{ret}}(\eta, \eta') = \theta(\eta - \eta')i[h_k(\eta)h_k^*(\eta') - h_k(\eta')h_k^*(\eta)], \quad (4.28)$$

where  $h_k(\eta)$  are the normalized “positive frequency” solutions of (4.4). If we implicitly resum the secular contribution from the sound speed, these will be built out of modes of the WKB form

$$\begin{aligned}
h_k(\eta) &\sim \frac{1}{\sqrt{2\omega_k(\eta)}} e^{\mp ik \int^\eta dt' c_s(\eta')} \\
&\sim \frac{1}{\sqrt{2\omega_k}} e^{\mp ik \int^t dt' \frac{c_s(t')}{a(t')}} \\
&= \frac{e^{\mp ik \int^t \frac{dt'}{a(t')}}}{\sqrt{2\omega_k}} [e^{\mp ik \int^t dt' \frac{c_s(t')-1}{a(t')}}]. \quad (4.29)
\end{aligned}$$

The resulting propagator will have causal support on the light cone determined by the speed  $c_s$  in the exponent of the exponentials. However, the secular resummation implicit in (4.29) only makes sense if the argument of the exponential in square brackets becomes of order unity or larger, otherwise this effect is clearly a perturbative one. However, as we have seen previously, provided we demand the EFT bound (4.7) then

$$\left| ik \int^t dt' \frac{c_s(t')-1}{a(t')} \right| \ll 1, \quad (4.30)$$

and so in the calculation of the Green’s function we may always treat the exponential perturbatively

$$h_k(\eta) \sim \frac{e^{\mp ik \int^t \frac{dt'}{a(t')}}}{\sqrt{2\omega_k}} \left[ 1 \mp ik \int^t dt' \frac{c_s(t')-1}{a(t')} + \dots \right]. \quad (4.31)$$

When computed in this manner, the resulting Green’s function will have at any finite order the same light cone structure as the FLRW background metric as determined by the leading exponential  $e^{i\vec{k}\cdot\vec{x} \mp ik \int^t \frac{dt'}{a(t')}}$ . We can only justify

the resummation of the terms that arise from expanding the exponential if those were the only terms to arise from the EFT expansion. But this is of course not the case, they represent only a subset of contributions and since their individual contribution remains perturbative we have no reason to expect that for example the term of quadratic order in  $(ik \int^t dt' \frac{c_s(t')-1}{a(t')})$  is any larger than other term that arise at the same order in the EFT expansion.

The implication of (4.7) is that since (4.30) is satisfied, it implies that (4.27) is best solved as a perturbation series defined by iterating the equation

$$\begin{aligned} & [\partial_\eta^2 + m_{\text{eff}}^2 + k^2] G_{\text{ret}}(\eta, \eta') \\ &= \delta(\eta - \eta') - (c_s^2(k, \eta) - 1) k^2 G_{\text{ret}}(\eta, \eta') \end{aligned} \quad (4.32)$$

that is

$$\begin{aligned} G_{\text{ret}}(\eta, \eta') &= G_{\text{ret}}^0(\eta, \eta') \\ &- \int_{\eta'}^{\eta} d\eta'' G_{\text{ret}}^0(\eta, \eta'') (c_s^2(k, \eta'') - 1) k^2 G_{\text{ret}}^0(\eta'', \eta') \\ &+ \dots, \end{aligned} \quad (4.33)$$

for which  $[\partial_\eta^2 + m_{\text{eff}}^2 + k^2] G_{\text{ret}}^0(\eta, \eta') = \delta(\eta - \eta')$  has the causal support of the background metric. This is legitimate as long as there is *no secular growth* in the perturbative expansion, which amounts to the requirement (4.30) which in the EFT of gravity follows from (4.7). We see that the EFT validity condition (4.7) is crucial to understanding how causality is preserved. It is the presence or absence of secular growth in the perturbative expansion that tells us whether or not the sound speed departure from unity is physical or not.

## V. BLACK HOLES IN THE EFT OF GRAVITY

### A. Speed of gravity near black holes in the EFT of gravity

As a second class of configurations that spontaneously breaks Lorentz invariance, we consider  $D = 4$  BH types of solutions and focus on static and spherically symmetric Ricci-flat vacuum configurations. This situation is not only particularly interesting phenomenologically, it also provides an explicit asymptotically flat example where S-matrix arguments can be applied. In this vacuum flat case the  $R^2$  operators in the EFT of gravity affect neither the background solution nor the propagation of GWs to first order in curvature corrections. Instead, the leading contributions arising from the dimension-6 operators of the form

$$\begin{aligned} \mathcal{L}_{\text{D6}} &= \frac{1}{M^2} [d_1 R \square R + d_2 R_{\mu\nu} \square R^{\mu\nu} + d_3 R^3 + d_4 R R_{\mu\nu} \\ &+ d_5 R R_{\mu\nu\alpha\beta} + d_6 R_{\mu\nu}^3 + d_7 R^{\mu\nu} R^{\alpha\beta} R_{\mu\nu\alpha\beta} \\ &+ d_8 R^{\mu\nu} R_{\mu\alpha\beta\gamma} R_{\nu}{}^{\alpha\beta\gamma} + d_9 R_{\mu\nu}{}^{\alpha\beta} R_{\alpha\beta}{}^{\gamma\sigma} R_{\gamma\sigma}{}^{\mu\nu} \\ &+ d_{10} R_{\mu}{}^{\alpha}{}_{\nu}{}^{\beta} R_{\alpha}{}^{\gamma}{}_{\beta}{}^{\sigma} R_{\gamma}{}^{\mu}{}_{\sigma}{}^{\nu}], \end{aligned} \quad (5.1)$$

where  $M$  is the “naïve” cutoff of the EFT. We consider the background solution to be that solution which is Schwarzschild if the  $R^3$  operators were absent, i.e., that of a corrected nonrotating black hole. In Schwarzschild coordinates, the equation of motion for the odd and even polarizations of the GWs  $h$  is the same and governed by an effective metric  $Z_{\mu\nu}$  [46]

$$Z^{\mu\nu} \mathcal{D}_\mu \mathcal{D}_\nu h + Vh = 0, \quad (5.2)$$

where the effective metric is expressed as

$$Z_{\mu\nu} dx^\mu dx^\nu = -Z_t dt^2 + Z_r^{-1} dr^2 + Z_\Omega r^2 d\Omega^2, \quad (5.3)$$

and the corrected metric functions are

$$\begin{aligned} Z_t = Z_r &= 1 - \frac{r_g}{r} + \frac{r_g}{2M^2 M_{\text{Pl}}^2 r^5} \left[ 6(4d_5 + d_8) \left( 5 \frac{r_g}{r} - 4 \frac{r_g^2}{r^2} \right) \right. \\ &- 4d_9 \left( 72 - 171 \frac{r_g}{r} + 94 \frac{r_g^2}{r^2} \right) \\ &\left. - d_{10} \left( 144 - 297 \frac{r_g}{r} + 152 \frac{r_g^2}{r^2} \right) \right] \\ &+ \mathcal{O}(MM_{\text{Pl}} r_g^2)^{-4}, \end{aligned} \quad (5.4)$$

and

$$\begin{aligned} Z_\Omega &= 1 + 72 \frac{r_g}{M^2 M_{\text{Pl}}^2 r^5} (2d_9 + d_{10}) \left( 1 - \frac{r_g}{r} \right) \\ &+ \mathcal{O}(MM_{\text{Pl}} r_g^2)^{-4}. \end{aligned} \quad (5.5)$$

Here  $r_g$  is the Schwarzschild radius of the BH solution in GR without the corrections from the EFT. In this EFT, the BH horizon  $r_g$  is slightly displaced by an amount proportional to  $(MM_{\text{Pl}} r_g^2)^{-2} \ll 1$  and is necessarily the same for every species, no matter how they couple to gravity. This is likely linked to the horizon theorem proven in [73] within the context of QED (see also [38,74]).

We can show that on the background of BH-like solutions, the speed of GWs can be both superluminal or subluminal depending on the signs of the EFT of gravity. Typically when the angular speed is subluminal, the radial speed is superluminal, as is the case if we think of this EFT as arising from integrating out a spin-1/2 field. Generically, the radial speed is given in terms of the coupling constants  $d_{9,10}$  as follows [46],

$$c_s^2(r) = 1 + \Delta c_s^2(r),$$

$$\text{with } \Delta c_s^2(r) = -144 \frac{2d_9 + d_{10}}{M^2 M_{\text{Pl}}^2 r_g^4} \left(1 - \frac{r_g}{r}\right) \frac{r_g^5}{r^5} + \mathcal{O}((MM_{\text{Pl}} r_g^2)^{-4}), \quad (5.6)$$

and is superluminal whenever  $2d_9 + d_{10} < 0$ . In the past, these types of arguments have been used to constraint EFTs. We emphasize here that this would be the wrong approach. First the choice  $2d_9 + d_{10} > 0$  would still lead to superluminalities in other configurations (e.g., superluminal angular low-energy speed), and second as in the case of the EFT of gravity in cosmological settings, the amount of low-energy superluminality is so suppressed that it can never lead to any macroscopic violation of causality as we explain below.

### B. Validity of EFT in BH spacetime

The discussion of the validity of the EFT in a BH spacetime is closely analogous to that in FLRW in Sec. IV B. One crucial difference however is that the leading order background geometry has vanishing Ricci tensor. Thus corrections to the propagation equations will be governed by the Weyl tensor and its covariant derivatives. Given an on-shell mode of momentum  $k_\mu$  with  $k_\mu k^\mu \approx 0$ , the naive highest order in  $k_\mu$  tensor we can construct that is linear in curvature is  $W_{abcd} k^a k^b k^c k^d$ , but this vanishes by virtue of the symmetries of the Weyl tensor. Hence the tensor with the highest powers of  $k$  is

$$A^a_b = W^a_{\text{bcd}} k^c k^d, \quad (5.7)$$

and by symmetry of the Weyl tensor we therefore have  $A_{ab} k^b = k^b A_{ba} = A_a^a = 0$ . In a general EFT expansion, we anticipate all scalar local operators to arise suppressed by the cutoff scale. This includes operators that are combinations of  $A_{ab}$  contracted with the metric and with itself. Hence following a reasoning identical to that in Sec. IV B, the highest possible on-shell momenta is determined by (at least) the EFT requirements

$$\text{Tr}[A^n] \ll M^{4n}, \quad (5.8)$$

for integer  $n$ . In addition, we must require the more obvious curvature requirement  $|W_{abcd} W^{abcd}| \ll M^4$  and related covariant derivative requirements.

Since the Schwarzschild metric is time translation invariant, on-shell modes are best characterized in terms of their frequency  $\omega \sim i\partial_t$ . For a transverse wave  $k_\mu = (-\omega, 0, 0, \pm\omega r^{1/2} \sin\theta / \sqrt{1 - r_g/r})$  with  $k_\mu k^\mu = 0$  the tensor  $A_{ab}$  is given by

$$A_{ab} dx^a dx^b = \frac{\omega^2 r_g}{2r^3} dt^2 - \frac{3\omega^2 r_g}{2r^3 (1 - r_g/r)^2} dr^2 + \frac{3\omega^2 r_g}{2r^3 (1 - r_g/r)} r^2 d\Omega^2, \quad (5.9)$$

and the bounds (5.8) amount to the single condition

$$\frac{r_g}{r^3} \frac{\omega^2}{(1 - r_g/r)} \ll M^4. \quad (5.10)$$

Interestingly, due to the symmetry of the spacetime, for a radial traveling wave  $k_\mu = (-\omega, \pm \frac{\omega}{(1 - r_g/r)}, 0, 0)$ , the situation is then more subtle. We then have

$$A_{ab} dx^a dx^b = -\frac{r_g}{r^3} \omega^2 \left( dt \mp \frac{1}{(1 - r_g/r)} dr \right)^2. \quad (5.11)$$

Since for such a radial mode  $A_{ab} \propto k_a k_b$ , then the conditions (5.8) are automatically satisfied. This does not mean that  $\omega$  can be arbitrarily large, but rather that we should look more closely at the higher derivative bounds such as

$$A^{ab} k^\mu \nabla_\mu A_{ab} \ll M^8, \quad (5.12)$$

and

$$((k^\mu \nabla_\mu)^p A^{ab}) ((k^\nu \nabla_\nu)^p A_{ab}) \ll M^{8+4p}, \quad (5.13)$$

as well as

$$(k^\mu \nabla_\mu)^p (W^{abcd} W_{abcd}) \ll M^{4+2p}. \quad (5.14)$$

These last two conditions are seen to be strongest in the limit  $p \rightarrow \infty$ , and amount to  $k^\mu \nabla_\mu \ll M^2$ , i.e., to the relation

$$\omega \ll M^2 r. \quad (5.15)$$

This is significantly stronger than (5.10), except near the black hole horizon, but applies only for modes with a significant radial component for which  $k^\mu \nabla_\mu = \pm\omega \partial_r + \dots$  picks out the radial dependence of the background geometry.

### C. Causality and time delay

We now focus on the case of a superluminal radial velocity where  $\Delta c_s^2(r)$  given in (5.6) is positive for  $r > r_g$ , ( $2d_9 + d_{10} < 0$ ). As a warmup let us consider the simple analysis of a radial moving trajectory. If we were to simultaneously send an outgoing radial photon and a GW with wavelength  $\lambda$  from a distance  $r_0 > r_g \gg M_{\text{Pl}}^{-1}$  just outside the BH horizon, the GW will arrive in advance

of the photon at infinity by an amount  $\Delta T_{\text{adv}}$ , whose maximal value is

$$\begin{aligned} \Delta T_{\text{adv}} &= \int_{r_0}^{\infty} \frac{dr}{(1 - \frac{r_g}{r})} - \int_{r_0}^{\infty} \frac{dr}{c_s(r)(1 - \frac{r_g}{r})} \\ &\sim \int_{r_0}^{\infty} \frac{dr \Delta c_s^2}{2(1 - \frac{r_g}{r})} \sim \frac{r_g}{M^2 M_{\text{Pl}}^2 r_0^4}. \end{aligned} \quad (5.16)$$

Remaining within the regime of validity of the EFT and being able to trust this answer requires the bound (5.15) to be satisfied at the distance  $r_0$ , leading to the requirement

$$\frac{r_g}{r_0^4} \ll M^2 \omega^{-1} \frac{r_g}{r_0^3}. \quad (5.17)$$

This leads to the upper bound

$$\Delta T_{\text{adv}} \sim \frac{r_g}{r_0^4} \frac{1}{M^2 M_{\text{Pl}}^2} \ll \underbrace{\frac{r_g}{r_0}}_{<1} \underbrace{\frac{1}{M_{\text{Pl}}^2 r_0^2}}_{\ll 1} \omega^{-1} \ll \omega^{-1}, \quad (5.18)$$

where we have used the requirement  $W^2 \ll M_{\text{Pl}}^4$ . Consequently the would-be scattering time advance is smaller than the physical wavelength of this mode and is therefore not resolvable. This result remains true regardless of how close to the horizon the initial wave starts out [so long as we only consider waves of frequencies that are within the regime of validity of the EFT at that point as dictated by (5.15)]. We see in fact that the effect described by (5.18) is even further suppressed relative to the FLRW effect by a factor  $\frac{r_g}{M_{\text{Pl}}^2 r_0^3}$  which is consistent with the fact that we are here relying on curvature-cubed (dimension-6) operators rather than curvature-squared (dimension-4) operators.

Moving beyond radial trajectories, let us consider the scattering time delay induced by the black hole background. By symmetry the retarded propagator may be determined in an angular momentum eigenbasis with angular momentum  $\ell$  for which the Eisenbud-Wigner time delay [48,50,75] as shown in Appendix B applied to the metric (5.3) is

$$\Delta T_{\ell} = 2 \int_{r_t}^{\infty} dr \left( \frac{1}{Z_t \sqrt{Z_r(Z_t^{-1} - Z_{\Omega}^{-1} \frac{b^2}{r^2})}} - 1 \right) - 2r_t, \quad (5.19)$$

where  $r_t$  is the turning point where the denominator vanishes  $r^2 Z_{\Omega} = Z_t b^2$ . As discussed in Sec. C 2 on asymptotically Schwarzschild spacetimes, this is logarithmically divergent in  $D = 4$ , but this need not concern us as the time delay correction to the usual GR Shapiro time delay is well defined

$$\begin{aligned} \Delta T_{\ell}^{\text{EFT}} &= 2 \int_{r_t}^{\infty} dr \left( \frac{1}{Z_t \sqrt{Z_r(Z_t^{-1} - Z_{\Omega}^{-1} \frac{b^2}{r^2})}} \right) \\ &\quad - 2 \int_{r_t(0)}^{\infty} dr \left( \frac{1}{f(r) \sqrt{1 - f(r) \frac{b^2}{r^2}}} \right), \end{aligned}$$

with  $f(r) = 1 - r_g/r$  and  $r_t(0)$  the usual GR turning point. Following the approach given in Appendix D at leading order in the EFT expansion this is

$$\Delta T_{\ell}^{\text{EFT}} = 2 \int_{r_t(0)}^{\infty} dr \left( \frac{1}{f(r) \sqrt{1 - f(r) \frac{b^2}{r^2}}} \right) \frac{d\delta R(r)}{dr}, \quad (5.20)$$

where the function  $\delta R$  is defined in (D7). At leading order in an  $r_g/b$  expansion is

$$\frac{d\delta R(r)}{dr} = 144 r_g (d_{10} + 2d_9) \frac{(4b^2 - r^2)}{M^2 M_{\text{Pl}}^2 b^2 r^5} + \dots, \quad (5.21)$$

and yields at leading order in  $r_g/b$  a time delay

$$\Delta T_{\ell}^{\text{EFT}} = \frac{480}{M^2 M_{\text{Pl}}^2} (d_{10} + 2d_9) \frac{r_g}{b^4}. \quad (5.22)$$

As expected when  $d_{10} + 2d_9 < 0$ , for which radial modes are superluminal, this corresponds to a time advance  $\Delta T_{\ell}^{\text{EFT}} < 0$ . However we may follow a reasoning similar to that in (5.18) to see that this time advance remains unresolvable within the EFT. Indeed even if we content ourselves with the weaker EFT condition (5.10) we would find for  $r_g/b \ll 1$

$$|\Delta T_{\ell}^{\text{EFT}}| \sim \frac{r_g}{M^2 M_{\text{Pl}}^2 b^4} \ll \underbrace{\sqrt{\frac{r_g}{b}}}_{<1} \underbrace{\frac{1}{M_{\text{Pl}}^2 b^2}}_{\ll 1} \omega^{-1} \ll \omega^{-1}, \quad (5.23)$$

and it is for this reason that causality is not being violated.

#### D. Failure of the Eikonal/semiclassical approximation

Since the contribution to the scattering phase shift is essentially  $\Delta \delta_{\ell}^{\text{EFT}} \sim 2\omega \Delta T_{\ell}^{\text{EFT}}$  the condition (5.23) is equivalent to the statement that the EFT contribution to the phase shift remains perturbatively small in the regime of validity of the EFT,

$$|\Delta \delta_{\ell}^{\text{EFT}}| \ll \sqrt{\frac{r_g}{b}} \frac{1}{M_{\text{Pl}} b^2} \ll 1. \quad (5.24)$$

Indeed not only is it smaller than unity, even if we were to take the extreme limit  $b \rightarrow \mathcal{O}(r_g)$ , it would still be parametrically suppressed by  $(M_{\text{Pl}} b)^{-2}$ . Specifically terms of



order  $(\Delta\delta_\ell^{\text{EFT}})^2$  will be of the same order as other EFT contributions that come in at the order  $(M_{\text{Pl}}b)^{-4}$ . This brings us to the essential point: it is implicit in the eikonal resummation (for example that performed in [23]) that the t-channel exchange ladder diagrams dominate over other Feynman diagrams in the perturbative expansion so that it is consistent to resum them into the exponentiated form

$$\delta_\ell(t\text{-channel}) + \text{higher order ladder diagrams} \rightarrow \frac{1}{2i} (e^{2i\delta_\ell(t\text{-channel})} - 1). \quad (5.25)$$

This is a well-justified procedure for the GR contributions which give rise to the Shapiro time delay, and indeed since the Shapiro time delay is of order  $r_g$  times a logarithmic factor, there is no problem engineering a situation for which  $\delta_\ell^g \gg 1$  or equivalently  $\Delta T^g \gg \omega^{-1}$  by having  $\omega \gg r_g^{-1}$ , which is not in conflict with any consistency requirement of pure GR.

By contrast, since the EFT contributions to the t-channel exchange are perturbatively small, it is not legitimate to resum them while neglecting other Feynman diagrams that can arise at the same order. We see that the situation of the low-energy EFT of gravity on a Schwarzschild background is closely analogous to that in FLRW. Performing an eikonal resummation of the contribution  $\delta_\ell^{\text{EFT}}$  is equivalent to the resummation of the would-be secular terms implicit in (4.29) and (4.31). In both cases this resummation is simply not justified at least as an indicator of causal support. Here, this shows up for us in the fact that the EFT correction to the scattering time delay is not resolvable within the approximation used to calculate the time delay, in analogy with the support of the FLRW light cone.

We stress that our condition for resolvability is not the same as that sometimes required in the literature, namely that the magnitude of the time delay is large in comparison to the naive cutoff, e.g.,

$$|\Delta T_\ell^{\text{EFT}}| \sim \underbrace{\sqrt{\frac{r_g}{b}}}_{<1} \underbrace{\frac{1}{bM}}_{\ll 1} \underbrace{\frac{1}{b^2 M_{\text{Pl}}^2}}_{\ll 1} M^{-1} \ll M^{-1}? \quad (5.26)$$

While we need to ensure that the scattering state/wave remains within the regime of validity of the EFT throughout its trajectory so as to be able to use the low-energy EFT to determine its time advance, nothing demands that the time advance itself should be measured within the low-energy EFT. By itself, the bound (5.26) is therefore irrelevant. Moreover,  $1/M$  is not the cutoff for time measurements within the EFT because the time delay is not a Lorentz invariant quantity. This is why we are careful in Secs. IV B and V B to identify the appropriate cutoff based on locally Lorentz invariant combinations. Resolvability here means whether it can be consistently computed in the

semiclassical/eikonal approximation, and at its heart the latter assumes frequencies and wavelengths that are large in comparison to the scales of variation of the background quantities.

## VI. QED IN CURVED SPACETIME

### A. Low-energy EFT for QED below the electron mass

It is helpful at this point to compare the above discussion with the classic case of superluminal speeds in a low-energy EFT, that of QED in curved spacetime first pointed out by Drummond and Hathrell [33] and extended in [34]. The result of [33] is particularly clean in that it does not require gravity to be dynamical, i.e., it would be obtained in a decoupling limit  $M_{\text{Pl}} \rightarrow \infty$  for fixed background curvature. Furthermore on the same background, different polarizations of light can be shown to have low-energy speeds which are both superluminal and subluminal. This gives rise to gravitational birefringence [33] and from higher order operators gravitationally induced dispersion of light [34]. The causal implications of this result for the photon have been discussed extensively in the literature [35–42].

For our present purposes it is sufficient to note that the leading effect from an effective action is of the form

$$\mathcal{L} = \sqrt{-g} \left( -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{\alpha}{M^2} R_{abcd} F^{ab} F^{cd} + \dots \right), \quad (6.1)$$

which is the relevant part of the low-energy effective action for QED on a curved spacetime where  $M$  is the electron mass. The operator  $RFF$  leads to polarization-dependent corrections to the low-energy sound speed. For a transverse-traveling wave (with momentum in the angular direction) on a Schwarzschild background, the low-energy sound speed is of the form [33]

$$c_s^2 = 1 + \frac{\beta_p}{M^2} \frac{r_g}{r^3} + \mathcal{O}\left(\frac{r_g^2}{M^4 r^6}\right), \quad (6.2)$$

where  $\beta_p$  is an order unity polarization-dependent constant. Specifically for radially polarized light  $\beta_p > 0$ , while  $\beta_p < 0$  for angular polarization. Interestingly were (6.1) the exact Lagrangian, we see that the equations for the electromagnetic field remain second order, and this speed would be the group/phase and front velocity. In practice this is not the case due to dispersion from higher order operators in the EFT not included [34]. At this level, the low-energy speed (6.2) includes no frequency dependence [none of the higher order terms  $\mathcal{O}(\frac{r_g^2}{M^4 r^6})$  in (6.2) would include any frequency dependence], yet higher-order operators that have not been included in the low-energy EFT (6.1) would affect the speed at high energy and (6.2) can certainly not be the speed of light at arbitrarily high energy.

### B. Unresolvable time advance

As noted in [33] the propagation of a photon can be understood in terms of evolution in an effective metric of the form

$$Z_{\mu\nu}dx^\mu dx^\nu = \left(1 + \frac{\beta_P r_g}{M^2 r^3}\right) \left(-f(r)dt^2 + \frac{1}{f(r)}dr^2\right) + r^2 d^2\Omega, \quad (6.3)$$

with the standard  $f(r) = (1 - \frac{r_g}{r})$ . Using the results of Appendix B for the effective metric (6.3) the naive expression for the Eisenbud-Wigner scattering time delay is

$$\Delta T_\ell = 2 \int_{r_t(\beta_P)}^\infty dr \left( \frac{1}{f(r)\sqrt{1 - \frac{b^2}{r^2}f(r)(1 + \frac{\beta_P r_g}{M^2 r^3})}} - 1 \right) - 2r_t(\beta_P), \quad (6.4)$$

where  $r_t(\beta_P)$  is the turning point as computed in the EFT with parameter  $\beta_P$ . Again in  $D = 4$  this is logarithmically divergent due to slow Coulomb  $1/r$  falloff in four dimensions but what is relevant is the extra time delay relative to scattering in the Schwarzschild metric with  $\beta_P = 0$ , i.e.,

$$\Delta T_\ell^{\text{EFT}} = 2 \int_{r_t(\beta_P)}^\infty dr \left( \frac{1}{f(r)\sqrt{1 - \frac{b^2}{r^2}f(r)(1 + \frac{\beta_P r_g}{M^2 r^3})}} \right) - 2 \int_{r_t(0)}^\infty dr \left( \frac{1}{f(r)\sqrt{1 - \frac{b^2}{r^2}f(r)}} \right), \quad (6.5)$$

which is a finite expression. As shown in Appendix D, using (D8) to first order in  $\beta_P$  this is given by

$$\Delta T_\ell^{\text{EFT}} = -\frac{2\beta_P r_g}{M^2} \int_{r_t(0)}^\infty dr \frac{1}{f(r)\sqrt{1 - \frac{b^2}{r^2}f(r)}} \times \frac{\partial}{\partial r} \left( \frac{\frac{b^2}{r^3}f(r)^3}{\frac{\partial}{\partial r}(\frac{b^2}{r^2}f(r)^3 - f(r)^2)} \right). \quad (6.6)$$

For large<sup>12</sup> impact parameter in comparison to the Schwarzschild radius  $b \gg r_g$  we have to leading order in an expansion in  $r_g/b$

<sup>12</sup>We note that the integral on the right-hand side of (6.5) does become arbitrarily large as  $r_t(0)$  approaches  $3r_g/2$  which is when  $b \rightarrow 3\sqrt{3}r_g/2$ . However this is purely an artifact of the fact that this is the peak of the effective potential (B19), and is the point at which we may easily transmit across the potential barrier. As such, the boundary conditions used for the solutions to derive (6.5) are not appropriate.

$$\Delta T_\ell^{\text{EFT}} = -\frac{2r_g\beta_P}{b^2M^2}, \quad (6.7)$$

in the large  $\ell$  limit and to leading order in  $r_g/b$ . As expected, this is a time advance for  $\beta_P > 0$ .

At the point of closest approach, the wave is purely transverse and the bound (5.10) applies. Again to leading order in  $r_g/b$ , we infer that for the EFT to remain valid at the impact parameter, the frequency should satisfy

$$\omega \ll M^2 \sqrt{\frac{b}{r_g}}. \quad (6.8)$$

Putting this together the time advance is bounded by

$$|\Delta T_\ell^{\text{EFT}}| = 2|\beta_P| \left(\frac{r_g}{b}\right) \frac{1}{bM^2} \ll \sqrt{\frac{r_g}{b}}\omega^{-1} \ll \omega^{-1}, \quad (6.9)$$

from which we again conclude that while the low-energy superluminality pointed out by Drummond and Hathrell in the context of the EFT for QED below the electron mass on a curved Schwarzschild background does technically lead to a time advance  $\Delta T_\ell < 0$ , this advance is not resolvable. In fact since in this example we know the UV completion we can calculate exactly the phase shift, and as shown in [43] in the shockwave limit the contribution to the phase shift is always small ensuring the time delay is indeed unresolvable. In addition it is shown in [43] that in the limit of high frequencies  $\lim_{\omega \rightarrow \infty} \Delta T_\ell^{\text{EFT}}(\omega) = 0$ . This confirms the underlying Lorentz invariant causality of the UV completion, but is secondary to the resolvability criterion in understanding how the low-energy EFT is consistent with causality.

### VII. CASE OF A RESOLVABLE PHYSICAL TIME ADVANCE

In the previous sections we have seen how well-known examples such as QED in curved spacetime and the EFT of gravity lead to time advances that are nevertheless unresolvable and hence not in tension with causality. To demonstrate that the time-delay analysis is not without content, we consider here an example of the opposite case, an EFT which does lead to a resolvable time advance and can therefore be concluded to be in tension with causality. For this we need to put ourselves in the situation where the bound (3.5) is violated in the decoupling limit for which it is sufficient to return to our canonical example of a  $P(X)$  scalar field minimally coupled as in (3.6) with the negative parameter  $a = -|a|$ .

In order to parallel the previous Schwarzschild discussion we consider sourcing the scalar field in the manner

$$\mathcal{L}_{\text{source}} = -\frac{\beta}{M_{\text{Pl}}} \phi T, \quad (7.1)$$

where  $T$  is the trace of the stress-energy tensor of all other matter fields present in that spacetime and  $\beta$  is a dimensionless coefficient which may be taken parametrically larger than unity. In particular if we consider a situation similar to that of the previous sections, with a Schwarzschild geometry generated by a mass  $M_*$  located at  $r = 0$ , with Schwarzschild radius  $r_g = 2M_*/M_{\text{Pl}}^2$  so that  $T$  itself is a delta function source. Ignoring for now the backreaction of the scalar field onto the geometry (i.e., working in the field theory on curved spacetime limit), the background profile  $\phi_0(r)$  for the scalar field is determined by the solution of<sup>13</sup>

$$\phi_0'(r) \left[ 1 + |a|f(r) \frac{\phi_0'(r)^2}{M^4} \right] = -\frac{\beta r_g M_{\text{Pl}}}{2r^2 f(r)}, \quad (7.2)$$

so that in the weak field regime,

$$\phi_0' = -\frac{\beta r_g M_{\text{Pl}}}{2r^2 f(r)} \left( 1 + \mathcal{O} \left( \beta^2 \frac{r_g^2 M_{\text{Pl}}^2}{r^4 M^4} \right) \right). \quad (7.3)$$

To remain within the safe region of the EFT we require  $|X| \ll M^4$  and so

$$\beta^2 \frac{r_g^2 M_{\text{Pl}}^2}{f(r) r^4 M^4} \ll 1. \quad (7.4)$$

Considering fluctuations about this background,  $\phi = \phi_0(r) + \delta\phi(t, r, \varphi)$ , the angular speed of  $\delta\phi$  waves remain

luminal while the radial speed is not only superluminal for  $a < 0$ ,

$$c_r^2 = 1 - a \frac{\beta^2 M_{\text{Pl}}^2 r_g^2}{2f(r) r^4 M^4}, \quad (7.5)$$

the departure from luminality is enhanced by a factor of  $(M_{\text{Pl}}^2/M^2)(r_g/r)$  as compared with that in QED (6.2) [and by a factor of  $(M_{\text{Pl}}^2/M^2)(M_{\text{Pl}}^2 r_g^2)(r/r_g) \gg 1$  as compared to the EFT of gravity (5.6)]. More precisely the effective metric for the fluctuations of the scalar field  $\delta\phi$  to leading order in the EFT expansion is of the form

$$Z_{\mu\nu} dx^\mu dx^\nu = \left( 1 + |a| \frac{\beta^2 M_{\text{Pl}}^2 r_g^2}{2f(r) r^4 M^4} \right) (-f(r) dt^2 + r^2 d^2\Omega) + \frac{1}{f(r)} dr^2. \quad (7.6)$$

Once again following the approach of Appendix D, the leading EFT correction to the time delay relative to the Schwarzschild metric is

$$\Delta T^{\text{EFT},P(X)} = 2 \int_{r_i(0)}^{\infty} dr \frac{1}{f(r) \sqrt{1 - \frac{b^2}{r^2} f(r)}} \frac{d\delta R(r)}{dr}, \quad (7.7)$$

where here

$$\frac{d\delta R(r)}{dr} = \frac{a\beta^2 M_{\text{Pl}}^2 r_g^2 (2b^4(3r - 5r_g)(r - r_g)^2 - 4r^6 r_g + b^2 r^3(-2r^2 + 8rr_g - 5r_g^2))}{2M^4 r^3 (b^2(2r - 5r_g)(r - r_g) + 2r^3 r_g)^2}. \quad (7.8)$$

This simplifies considerably in the limit of large impact parameter  $b \gg r_g$  to be

$$\frac{d\delta R(r)}{dr} = \frac{a\beta^2 M_{\text{Pl}}^2 r_g^2 (3b^2 - r^2)}{4b^2 M^4 r^4}, \quad (7.9)$$

which gives to leading order in  $r_g/b$  the time delay correction

$$\begin{aligned} \Delta T_\ell^{\text{EFT},P(X)} &\sim 2 \int_b^\infty dr \frac{1}{\sqrt{1 - \frac{b^2}{r^2}}} \frac{a\beta^2 M_{\text{Pl}}^2 r_g^2 (3b^2 - r^2)}{4b^2 M^4 r^4} \\ &= \frac{a\beta^2 M_{\text{Pl}}^2 r_g^2 \pi}{8b^3 M^4}, \end{aligned} \quad (7.10)$$

which as expected is a time advance for  $a < 0$ .

<sup>13</sup>We stress it is not important that there is a global solution of this equation for all  $r$ , as we are working with a truncated EFT and the solution can only be trusted in the regime given typically by (7.4).

Now turning back to establishing the regime of validity of the EFT, since the angular speed is luminal, in this case the time delay is maximal at small angular momentum. Unlike the previous discussion in Sec. V B already for a radial wave of the form  $k_\mu = (-\omega, \pm\omega f(r)^{-1}(1 + a \frac{\phi_0'^2 f}{M^4}), 0, 0)$ , the bounds (5.8) are nontrivial when the curvature of the effective metric  $Z_{\mu\nu}$  is used and leads to the requirement

$$b^7 \gg \frac{\beta^2 M_{\text{Pl}}^2 r_g^3}{M^8} \omega^2. \quad (7.11)$$

Using this bound in the time advance, we get a bound in terms of the inverse frequency

$$|\Delta T_\ell^{\text{EFT},P(X)}| \sim \frac{\beta^2 M_{\text{Pl}}^2 r_g^2}{b^3 M^4} \ll |\beta| \sqrt{M_{\text{Pl}}^2 r_g b \omega^{-1}}. \quad (7.12)$$

Remarkably we are now here in a situation where this bound can be satisfied with a resolvable time advance

$|\Delta T_\ell^{\text{EFT},P(X)}| \gg \omega^{-1}$  so long as  $M_{\text{Pl}}^2 r_g b \sim M_* b \gg 1$  which is extremely easy to satisfy. Taking the sun for instance at its Schwarzschild radius  $b \sim r_g$  would lead to a factor  $\sqrt{M_{\text{Pl}}^2 r_g b} \sim M_*/M_{\text{Pl}} \sim 10^{38}$ .

However (5.8) is only one possible EFT condition and we should also check that this effect survives for example (5.15) or (5.10). Applying the more conservative criterion (5.15) to be safe we would be led to

$$|\Delta T_\ell^{\text{EFT},P(X)}| \sim \frac{\beta^2 M_{\text{Pl}}^2 r_g^2}{b^3 M^4} \ll \beta^2 \frac{M_{\text{Pl}}^2 r_g^2}{M^2 b^2} \omega^{-1}. \quad (7.13)$$

Again we can engineer a resolvable time advance with  $|\beta| M_{\text{Pl}} r_g \gg Mb$  and (7.4) can still be satisfied for  $b \gg M^{-1}$ . The simplest way to achieve this is to make  $|\beta| \gg 1$ . Indeed, since the theory defined by (3.6) has a well-defined decoupling limit as  $M_{\text{Pl}} \rightarrow \infty$ , the time advance has no relationship to gravity per se. The coefficient  $\beta$  is the source coupling (7.1) and may equally well have been defined as  $\beta = \tilde{\beta} M_{\text{Pl}}/M$  so that the coupling (7.1) survives in the limit  $M_{\text{Pl}} \rightarrow \infty$  for fixed source mass. Then the bound (7.13) may equally well be written as

$$|\Delta T_\ell^{\text{EFT},P(X)}| \ll |\tilde{\beta}| \frac{M_*^2}{M^4 b^2} \omega^{-1}. \quad (7.14)$$

In terms of scattering amplitudes, this can be related to a phase shift of order

$$|\Delta \delta_\ell^{\text{EFT},P(X)}| \ll |\tilde{\beta}| \frac{M_*^2}{M^4 b^2}, \quad (7.15)$$

which may easily be engineered to be  $|\Delta \delta_\ell^{\text{EFT},P(X)}| \gtrsim 1$ , by taking  $M_*^2/M^2 \gg M^2 b^2$  while maintaining  $b \gg M^{-1}$  to ensure (7.4). For such a phase shift, the implicit summation that goes into the semiclassical/eikonal approximation is justified as the ladder diagrams dominate over other contributions at each order in loops. This situation is orthogonal to that of the EFT of gravity and that of QED below the electron mass and as discussed in Sec. IV D is due to the lack of suppression of the speed correction, a consequence of the fact that (3.5) is violated.

## VIII. DISCUSSION

To summarize, when dealing with a gravitational effective theory, superluminal low-energy speeds (with respect to the metric out of which it is constructed) is not only possible, but is sometimes demanded by underlying causality and analyticity criteria. Indeed this has been well known since the work of [33,34] for QED in curved spacetime [35–42], and more recently noted in [45,46] for GWs. Relativistic causality nevertheless remains intact because the causal support of the retarded propagator vanishes outside of the light cone of the metric, defined

in the field frame with a well-defined decoupling limit. This apparent contradiction is resolved in the examples of the FLRW and Schwarzschild geometries and we have discussed it by identifying the regime of validity of the effective theory, and asking—given the low-energy form of the retarded propagator—whether it is possible to influence events outside of the metric light cone? When the propagator is computed perturbatively, this can only happen when there are secular terms in the perturbative expansion which need to be resummed, and it is this resummation that extends the support of the light cone. This is exactly what happens in known pathological cases where relativistic causality is violated [6,64]. Here we have shown that in the case of EFT corrections discussed in [45,46], the condition for the validity of the EFT automatically precludes any secular behavior and hence relativistic causality is left intact, despite explicitly superluminal low-energy speeds.

With this in mind, faced with a given gravitational effective theory, how should we apply causality requirements in the absence of clean S-matrix requirements? As stated in the introduction many works in the literature simply demand that around a given background, the speed of propagation of all modes is (sub)luminal relative to the metric out of which the theory is constructed. This is a demonstrably false criterion as it is not invariant under field redefinitions as the example in Sec. III C clearly demonstrates. As noted above, the issue about field frame dependence can be partially resolved by working in a frame in which it is possible to consistently take a decoupling limit  $M_{\text{Pl}} \rightarrow \infty$  for which gravitational effects decouple (see Sec. III B for clarifications on what is meant by a decoupling limit). It is then consistent to demand that the resulting Minkowski spacetime field theory respects all of the standard causality/analyticity/positivity requirements. In particular this leads to the condition (3.5). Once this is done, this implies that all effects that lead to mild superluminalities are in this frame  $M_{\text{Pl}}$  suppressed effects.

The next question is how do we impose causality away from the decoupling limit, at finite  $M_{\text{Pl}}$ , or in situations in which there is no clean decoupling limit? Once again, simply demanding  $c_s \leq 1$  for all perturbations around a given background is an incorrect criterion, as the known examples illustrate [33,34,45,46]. Ideally we may just appeal to the full UV theory to demonstrate that the exact retarded propagator is causal [35–42]; however, we rarely have that luxury, and furthermore in time-dependent spacetimes such as FLRW such analyticity methods are not applicable. We can however cleanly answer this equation within the low-energy effective theory itself provided that we appropriately identify its regime of validity. Indeed causality resolution should entirely lie within the purview of the low-energy EFT as by construction it gives the large distance macroscopic description of the theory, and this is

where any causality violation will become apparent. If the superluminal speed were physical, it would lead to a macroscopic consequence, specifically an enlarged support for the retarded propagator. Since at zeroth order in the EFT expansion we assume that the retarded propagator has the support implied by the usual metric light cone, and since we can structure the calculation of the corrected retarded propagator as a perturbative expansion around this, the telltale sign for a modified causal structure is secular behavior in the perturbative expansion that needs to be resummed. We have shown explicitly that in the cases considered in [45,46] this secular behavior is absent provided we restrict ourselves to effects which can be consistently calculated in the regime of validity of the EFT in question. Concretely in FLRW validity of the EFT imposes a cutoff in on-shell momenta of the form

$$\frac{|\vec{k}|}{a(t)} \ll \Lambda_c = \frac{M^2}{\sqrt{-\dot{H}}}. \quad (8.1)$$

This cutoff is parametrically higher than the invariant cutoff  $M$  due to the underlying assumption of Lorentz invariance of the UV completion.<sup>14</sup>

The entire discussion for cosmological spacetimes parallels exactly the more straightforward case of scattering in asymptotically flat/Schwarzschild geometries. There we may use the positivity of the Eisenbud-Wigner scattering time delay as a clean criterion for causality. However any computed time advance from EFT corrections can only be interpreted as a macroscopic causality violation if it is resolvable meaning if  $\Delta T_\ell^{\text{EFT}} \gtrsim \omega^{-1}$ . In all the known cases of time advances arising from consistent UV completions, the time advance is not resolvable and its associated contribution to the scattering phase shift is less than unity  $|\Delta \delta_\ell^{\text{EFT}}| \ll 1$  and  $\Delta T_\ell^{\text{EFT}} \ll \omega^{-1}$ . Hence the eikonal/semiclassical resummation, which is the parallel of the secular resummation of the retarded propagator, is not justified as an indicator of causal properties. Explicit UV completions show that the high energy behavior of the corrections to the phase shift are very different, and generically we expect  $\lim_{\omega \rightarrow \infty} \Delta T^{\text{EFT}}(\omega) \rightarrow 0$ , equivalent to the expectation that in an underlying Lorentz invariant theory  $\lim_{\omega \rightarrow \infty} c_s(\omega) \rightarrow 1$ . We stress again however that we are not using these expectations to resolve the naive causality issue with the low-energy EFT.

In using causality to constraint EFTs, in the end the issue is not just about the signs of Wilson coefficients in an effective action, but also crucially about their sizes. If a given operator induces a large superluminal effect that is both (a) nonzero in the decoupling limit, and (b) leads to secular behavior within the regime of validity, then we can safely conclude that this would violate the traditional requirements of relativistic causality as in the case of the

“wrong sign”  $P(X)$  model. However, if a given operator leads to a *small* superluminal speed which is insufficient to give rise to any secular growth in the regime of validity of the EFT, then we cannot conclude any conflict with relativistic causality. Although we have not emphasized this here, there is a strong interplay between these requirements and the application of positivity bounds in a gravitational setting as discussed in [45]. This weaker requirement of “signs” of EFT coefficients likely connects with weaker requirements for positivity bounds in the presence of a massless graviton as discussed in [63,76].

In view of the previous discussion, one may be tempted to hastily conclude that superluminalities in a low-energy EFT could potentially be allowed so long as they are bounded to be highly suppressed, along the lines of  $c_s^2 < 1 + \mathcal{O}(M_{\text{Pl}}^{-2})$  and such considerations are hence irrelevant for realistic observations. Such a conclusion would be incorrect. As emphasized in Sec. III, great care ought to be given to dealing with superluminalities in the gravitational setup and the bound (3.8) only makes sense if and in a frame where the decoupling limit  $M_{\text{Pl}} \rightarrow \infty$  can be taken appropriately. What this implies in practice depends on context and on the precise frame in which the bounds are determined, but in many EFTs for inflation and dark energy this often implies that the correct bound to be imposed on the speed of gravitational waves following from causality is in fact in the regime in which they are strictly superluminal rather than subluminal. Biasing data with a subluminal speed of gravitational wave prior can often turn out to be in direct tension with causality.

## ACKNOWLEDGMENTS

The work of A. J. T. and C. d. R. is supported by an STFC Grant No. ST/P000762/1. C. d. R. thanks the Royal Society for support at Imperial College London through a Wolfson Research Merit Award. C. d. R. is supported by the European Union’s Horizon 2020 Research Council Grant No. 724659 MassiveCosmo ERC–2016–COG and by a Simons Foundation Grant No. 555326 under the Simons Foundation’s Origins of the Universe initiative, “Cosmology Beyond Einstein’s Theory.” A. J. T. thanks the Royal Society for support at ICL through a Wolfson Research merit award.

## APPENDIX A: LOW-ENERGY EFT FOR GRAVITY

### 1. Graviton dispersion relation

The leading modification to the dispersion relation of a gravitational wave in a curved background can be determined by first inspecting the effect in Minkowski space-time. That is because diffeomorphism invariance may be used to covariantize the Minkowski answer and the ambiguities in this procedure turn out to be subleading

<sup>14</sup>Since causality is clearly very different for fundamentally Lorentz violating theories, we do not discuss them here.

corrections. With this in mind let us first determine the form of the graviton propagator or two point function in Minkowski spacetime. At the linear level this may be described by a Lorentz tensor  $h_{\mu\nu}$  which transforms under linear diffeomorphisms as  $h_{\mu\nu} \rightarrow h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$ . Although nonlinearly we cannot construct local gauge invariants for gravity, at the linear level we can by means of introducing a conserved external source  $T_{\mu\nu}$  for which  $\partial_\mu T^{\mu\nu} = 0$ , and considering the “ $TT$  amplitude” that describes the interaction between two sources (equivalently the free field connected generating function)

$$\Delta S_{TT} = \frac{i}{2M_{\text{Pl}}^2} \int d^4x \times \int d^4y T^{\mu\nu}(x) \langle 0 | \hat{T} h_{\mu\nu}(x) h_{\alpha\beta}(y) | 0 \rangle T^{\alpha\beta}(y). \quad (\text{A1})$$

Following standard arguments, causality, unitarity and Lorentz invariance fix the form of this amplitude to be

$$\Delta S_{TT} = \int \frac{d^4k}{(2\pi)^4} T^{\mu\nu}(-k) G_{\mu\nu\alpha\beta}(k) T^{\alpha\beta}(k), \quad (\text{A2})$$

where formally

$$G_{\mu\nu\alpha\beta}(k) = \frac{1}{2} \frac{Z}{M_{\text{Pl}}^2} \frac{P_{\mu\nu\alpha\beta}^{2,0}}{k^2} + \frac{1}{M_{\text{Pl}}^4} \int_0^\infty d\mu \rho_2(\mu) \frac{P_{\mu\nu\alpha\beta}^2}{\mu + k^2} + \frac{1}{M_{\text{Pl}}^4} \int_0^\infty d\mu \rho_0(\mu) \frac{P_{\mu\nu\alpha\beta}^0}{\mu + k^2}, \quad (\text{A3})$$

and where the polarization tensors are given by

$$P_{\mu\nu\alpha\beta}^{2,0} = \frac{1}{2} (\eta_{\mu\alpha} \eta_{\nu\beta} + \eta_{\nu\alpha} \eta_{\mu\beta} - \eta_{\mu\nu} \eta_{\alpha\beta}), \quad (\text{A4})$$

$$P_{\mu\nu\alpha\beta}^2 = \frac{1}{2} \left( \eta_{\mu\alpha} \eta_{\nu\beta} + \eta_{\nu\alpha} \eta_{\mu\beta} - \frac{2}{3} \eta_{\mu\nu} \eta_{\alpha\beta} \right), \quad (\text{A5})$$

$$P_{\mu\nu\alpha\beta}^0 = \eta_{\mu\nu} \eta_{\alpha\beta}. \quad (\text{A6})$$

In writing this expression we assume that this is the true propagator and so is independent of UV cutoffs or RG sliding scales used in computing loops. The first term in (A3) is the usual massless graviton pole which must arise by diffeomorphism invariance uncorrected other than by a wave function renormalization.<sup>15</sup> The second and third terms come from the exchange of intermediate massive states of spin-2 and spin-0 which may arise from either tree

<sup>15</sup>As famously noted in [77] we can have  $Z = 0$  and hence mass with gauge invariance. This is realized explicitly in the DGP model where the massless graviton mode is non-normalizable [78].

level exchange or loop corrections. Unitarity here comes from the strict positivity requirement that

$$\rho_2(\mu) > 0 \quad \text{and} \quad \rho_0(\mu) > 0. \quad (\text{A7})$$

## 2. Renormalization group without renormalization

In practice however (A3) is only valid if the integrals  $\int_{\mu_0}^\infty d\mu \rho_2(\mu) \mu^{-1}$  and  $\int_{\mu_0}^\infty d\mu \rho_0(\mu) \mu^{-1}$  converge for some finite  $\mu_0 > 0$ , i.e., provided they converge in the UV. In general they do not and their divergence is directly related to the renormalization of the curvature squared terms in the effective action. As such on dimensional grounds we expect them to be logarithmically divergent, which is borne out by explicit calculation at one-loop [45].

To deal with this we perform one subtraction defined at an arbitrary scale  $\mu_0$  to give

$$G_{\mu\nu\alpha\beta}(k) = \frac{1}{2} \frac{Z}{M_{\text{Pl}}^2} \frac{P_{\mu\nu\alpha\beta}^{2,0}}{k^2} + \frac{C_2(\mu_0)}{M_{\text{Pl}}^4} P_{\mu\nu\alpha\beta}^2 + \frac{C_0(\mu_0)}{M_{\text{Pl}}^4} P_{\mu\nu\alpha\beta}^0 + \frac{(\mu_0 - k^2)}{M_{\text{Pl}}^4} \int_0^\infty d\mu \rho_2(\mu) \frac{P_{\mu\nu\alpha\beta}^2}{(\mu + \mu_0)(\mu + k^2)} + \frac{(\mu_0 - k^2)}{M_{\text{Pl}}^4} \int_0^\infty d\mu \rho_0(\mu) \frac{P_{\mu\nu\alpha\beta}^0}{(\mu + \mu_0)(\mu + k^2)}. \quad (\text{A8})$$

Since the propagator cannot depend on the arbitrary subtraction scale  $\mu_0$  we obtain the dispersion relation analog of the renormalization group equations

$$\mu_0 \frac{d}{d\mu_0} C_S(\mu_0) = - \int_0^\infty d\mu \rho_S(\mu) \frac{\mu_0}{(\mu + \mu_0)^2}, \quad (\text{A9})$$

the right-hand side being finite if our assumption about the overall number of subtractions are correct. We stress though that  $\mu_0$  should not be confused with any UV cutoff or sliding scale used in computing loops; it is however clear that it plays a similar role. The integration constants that arise in the solution of the equation are the undetermined subtraction constants. It is natural to define the IR values of these constants at  $\mu_0 = M_{\text{IR}}^2$ , and the UV at some high-energy scale  $\mu_0 = M_{\text{UV}}^2$ . Integrating the RG equation we have

$$C_S^{\text{IR}} = C_S^{\text{UV}} + \int_0^\infty d\mu \rho_S(\mu) \frac{(M_{\text{UV}}^2 - M_{\text{IR}}^2)}{(\mu + M_{\text{IR}}^2)(\mu + M_{\text{UV}}^2)}, \quad (\text{A10})$$

and so we clearly have by unitarity

$$C_S^{\text{IR}} > C_S^{\text{UV}}. \quad (\text{A11})$$

Thus the dispersion relation demands positivity of the flow from the UV to the IR. It does not however guarantee that  $C_S^{\text{IR}} > 0$ .

### 3. Low-energy effective field theory

Let us now make an assumption similar to that described in Sec. II A that the dominant contribution to the spectral densities comes from energies for which  $\mu \geq M^2$ , where  $M$  is viewed as the cutoff of the low-energy effective theory. In other words we assume that there is a weakly coupled low-energy effective theory for which the loop contributions from light fields are small relative to the effects from heavy fields whose masses satisfy  $M_I \geq M$ . This is quite natural here for situations in which the number of heavy fields with masses greater than  $M$  is much larger than the number of light fields, precisely because each field at one-loop contributes logarithmically to  $C_S$ .

In this case the propagator may be split up as

$$G_{\mu\alpha\beta}(k) = G_{\mu\alpha\beta}^{\text{EFT}}(k) + G_{\mu\alpha\beta}^{\text{IR}}(k), \quad (\text{A12})$$

where the IR part comes exclusively from loops of light fields of masses smaller than  $M$ , and is thus parametrically suppressed if the low-energy effective theory is weakly coupled

$$G_{\mu\alpha\beta}^{\text{IR}}(k) = \frac{(\mu_0 - k^2)}{M_{\text{Pl}}^4} \int_0^{M^2} d\mu \rho_2(\mu) \frac{P_{\mu\alpha\beta}^2}{(\mu + \mu_0)(\mu + k^2)} + \frac{(\mu_0 - k^2)}{M_{\text{Pl}}^4} \int_0^{M^2} d\mu \rho_0(\mu) \frac{P_{\mu\alpha\beta}^0}{(\mu + \mu_0)(\mu + k^2)}, \quad (\text{A13})$$

and the remaining part is that which will essentially be described by the tree level low-energy effective theory

$$G_{\mu\alpha\beta}^{\text{EFT}}(k) = \frac{1}{2} \frac{Z}{M_{\text{Pl}}^2} \frac{P_{\mu\alpha\beta}^{2,0}}{k^2} + \frac{C_2(\mu_0)}{M_{\text{Pl}}^4} P_{\mu\alpha\beta}^2 + \frac{C_0(\mu_0)}{M_{\text{Pl}}^4} P_{\mu\alpha\beta}^0 + \frac{(\mu_0 - k^2)}{M_{\text{Pl}}^4} \int_{M^2}^{\infty} d\mu \rho_2(\mu) \frac{P_{\mu\alpha\beta}^2}{(\mu + \mu_0)(\mu + k^2)} + \frac{(\mu_0 - k^2)}{M_{\text{Pl}}^4} \int_{M^2}^{\infty} d\mu \rho_0(\mu) \frac{P_{\mu\alpha\beta}^0}{(\mu + \mu_0)(\mu + k^2)}. \quad (\text{A14})$$

The IR part  $G_{\mu\alpha\beta}^{\text{IR}}(k)$  is generically nonlocal since it includes loops of light fields, in particular those of the graviton itself. For example if  $\rho_{0,2}(\mu)$  are approximately constant over the range  $0 \leq \mu < M^2$  then we have approximately

$$G_{\mu\alpha\beta}^{\text{IR}}(k) \approx \frac{1}{M_{\text{Pl}}^4} \rho_2(0) P_{\mu\alpha\beta}^2 \ln \left( \frac{\mu_0(M^2 + k^2)}{k^2(M^2 + \mu_0)} \right) + \frac{1}{M_{\text{Pl}}^4} \rho_0(0) P_{\mu\alpha\beta}^0 \ln \left( \frac{\mu_0(M^2 + k^2)}{k^2(M^2 + \mu_0)} \right). \quad (\text{A15})$$

By contrast  $G_{\mu\alpha\beta}^{\text{EFT}}(k)$  is local when viewed at energies  $|k^2| \ll M^2$ . In other words we may perform a standard EFT expansion in the form

$$G_{\mu\alpha\beta}^{\text{EFT}}(k) \approx \frac{1}{2} \frac{1}{M_{\text{Pl}}^2} Z \frac{P_{\mu\alpha\beta}^{2,0}}{k^2} + \frac{C_2(\mu_0)}{M_{\text{Pl}}^4} P_{\mu\alpha\beta}^2 + \frac{C_0(\mu_0)}{M_{\text{Pl}}^4} P_{\mu\alpha\beta}^0 + \sum_{n=1}^{\infty} \frac{(\mu_0 - k^2)^n}{M_{\text{Pl}}^4} \int_{M^2}^{\infty} d\mu \rho_2(\mu) \frac{P_{\mu\alpha\beta}^2}{(\mu + \mu_0)^{n+2}} + \sum_{n=1}^{\infty} \frac{(\mu_0 - k^2)^n}{M_{\text{Pl}}^4} \int_{M^2}^{\infty} d\mu \rho_0(\mu) \frac{P_{\mu\alpha\beta}^0}{(\mu + \mu_0)^{n+2}}. \quad (\text{A16})$$

Furthermore taking  $\mu_0 = M_{\text{IR}}^2 \approx 0$ , then in the region  $M_{\text{IR}}^2 \ll |k^2| \ll M^2$  we may approximate this as

$$G_{\mu\alpha\beta}^{\text{EFT}}(k) \approx \frac{1}{2} \frac{1}{M_{\text{Pl}}^2} Z \frac{P_{\mu\alpha\beta}^{2,0}}{k^2} + \frac{C_2^{\text{IR}}}{M_{\text{Pl}}^4} P_{\mu\alpha\beta}^2 + \frac{C_0^{\text{IR}}}{M_{\text{Pl}}^4} P_{\mu\alpha\beta}^0 + \sum_{n=1}^{\infty} \frac{(-k^2)^n}{M_{\text{Pl}}^4} \int_{M^2}^{\infty} d\mu \rho_2(\mu) \frac{P_{\mu\alpha\beta}^2}{\mu^{n+2}} + \sum_{n=1}^{\infty} \frac{(-k^2)^n}{M_{\text{Pl}}^4} \int_{M^2}^{\infty} d\mu \rho_0(\mu) \frac{P_{\mu\alpha\beta}^0}{\mu^{n+2}}. \quad (\text{A17})$$

This is the standard form of the a tree level effective theory description of the propagator as a local derivative expansion, and is the direction analog of (2.7).

### 4. 1PI and Wilson effective action

In performing the decomposition (A12), we are implicitly assuming that the light loops are computed through a unitarity cut method, consistent with the dispersion relation. In short, rather than computing the loop process through standard means, we compute its imaginary part, and then infer its remaining contribution through its dispersion relation. With this proviso, we then recognize that  $G_{\mu\alpha\beta}^{\text{EFT}}(k)$  will be the two point function computed from the Wilsonian effective action valid below the scales  $|k^2| < M^2$ , and  $G_{\mu\alpha\beta}(k) = G_{\mu\alpha\beta}^{\text{EFT}}(k) + G_{\mu\alpha\beta}^{\text{IR}}(k)$  will be the result of the 1PI effective action. As always this split is arbitrary here depending on the subtraction scale  $\mu_0$  which hence we take as some IR scale  $\mu_0 = M_{\text{IR}}^2$ .

Assuming we only compute loops of heavy fields (not the graviton itself), then the 1PI effective action and Wilsonian effective actions will be diffeomorphism invariant. We may thus write a local covariant action which can reproduce the propagator  $G_{\mu\alpha\beta}^{\text{EFT}}(k)$  which is found to be

$$\mathcal{L}^{\text{EFT}} = \sqrt{-g} \left( \frac{M_{\text{Pl}}^2}{2} R + C_0^{\text{IR}} R^2 + C_2^{\text{IR}} \left( R_{\mu\nu} R^{\mu\nu} - \frac{1}{3} R^2 \right) \right) + \text{higher derivative terms}. \quad (\text{A18})$$

The higher derivative terms are relevant for BH solutions, but these cannot be inferred from our above argument and must be computed explicitly as done in [45,46]. However

they do have the virtue of being prescription independent. We can also add to this action a Gauss-Bonnet term which cannot be inferred from our calculation. Indeed up to a Gauss-Bonnet term (A19) may equivalently be written as

$$\mathcal{L}^{\text{EFT}} = \sqrt{-g} \left( \frac{M_{\text{Pl}}^2}{2} R + C_{R^2}^{\text{IR}} R^2 + C_{W^2}^{\text{IR}} W_{\mu\nu\alpha\rho}^2 + C_{\text{GB}} \text{GB} \right) + \text{higher derivative terms}, \quad (\text{A19})$$

where

$$C_{W^2} = \frac{1}{2} C_2 \quad C_{R^2} = C_0. \quad (\text{A20})$$

The virtue of writing this covariantly is that assuming that matter remains minimally coupled to this metric then we may infer the effect of these additional corrections on other backgrounds such as FLRW. The positivity of the flow (A11) then implies

$$C_{W^2}^{\text{IR}} > C_{W^2}^{\text{UV}}, \quad C_{R^2}^{\text{IR}} > C_{R^2}^{\text{UV}}. \quad (\text{A21})$$

The covariant form of the 1PI effective action being nonlocal is more complicated (see [79–87]), but it is sufficient to note that the following covariant expression reproduces the desired  $TT$  amplitude and is consistent with the Wilsonian effective action

$$\begin{aligned} \mathcal{L}^{\text{1PI}} = & \sqrt{-g} \left( \frac{M_{\text{Pl}}^2}{2} R + C_2(\mu_0) \left( R_{\mu\nu} R^{\mu\nu} - \frac{1}{3} R^2 \right) + C_0(\mu_0) R^2 + C_{\text{GB}}(\mu_0) R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma}^* \right. \\ & + R^{\mu\nu} \int_0^\infty d\mu \rho_2(\mu) \frac{(\mu_0 + \square)}{(\mu + \mu_0)(\mu - \square)} \left( R_{\mu\nu} - \frac{1}{3} g_{\mu\nu} R \right) + R \int_0^\infty d\mu \rho_0(\mu) \frac{(\mu_0 + \square)}{(\mu + \mu_0)(\mu - \square)} R \\ & \left. + R^{\mu\nu\rho\sigma} \int_0^\infty d\mu \rho_{\text{GB}}(\mu) \frac{(\mu_0 + \square)}{(\mu + \mu_0)(\mu - \square)} R_{\mu\nu\rho\sigma}^* \right) + \dots \end{aligned} \quad (\text{A22})$$

The last term is a nonlocal extension of the Gauss-Bonnet whose coefficient cannot be inferred from the arguments made so far. In particular we cannot assume that  $\rho_{\text{GB}}(\mu) > 0$ . As a special case, a nonlocal action of this form is used for example in [88] for the specific form of  $\rho_2(\mu)$  and  $\rho_1(\mu)$  that arises from one-loop integrals of massive and massless states. We stress again that despite appearances, (A22) is independent of the arbitrary subtraction scale  $\mu_0$ .

## APPENDIX B: SEMICLASSICAL PHASE SHIFT AND TIME DELAY

### 1. Langer approach

The scattering phase shift  $\delta_\ell$  in the semiclassical (WKB) approximation for scattering in a spherically symmetric background is easily computed and we sketch the essential result here. We will assume that the equation of motion of the propagating degrees of freedom may be put in the form of a scalar field living on an effective metric  $Z_{\mu\nu}$  as is the case in all the discussed examples. Although generically this equation will have an effective mass, this mass term makes a negligible contribution to the scattering phase shift for high frequencies; consequently it is sufficient to consider a massless scalar for the purposes of our discussion. To accommodate the behavior of modes with  $\ell$  small we follow the approach of Langer [89]. The fluctuations of the effective scalar in a  $D$ -dimensional spherically symmetric background expressed in the form

$$Z_{\mu\nu} dx^\mu dx^\nu = -Z_t dt^2 + Z_r^{-1} dr^2 + r^2 Z_\Omega d\Omega_{D-2}^2 \quad (\text{B1})$$

can be expressed in terms of  $D-2$  dimensional generalization of spherical harmonics, i.e., eigenstates of  $\nabla_{D-2}^2 = -\ell(\ell + (D-3))$  and satisfies a wave equation for a given  $\ell$

$$\begin{aligned} \mathcal{Z}_t \omega^2 \phi_\ell + \frac{1}{r^{D-2}} \frac{\partial}{\partial r} \left( r^{D-2} \mathcal{Z}_r \frac{\partial}{\partial r} \phi_\ell \right) \\ - \mathcal{Z}_\Omega \frac{\ell(\ell + D - 3)}{r^2} \phi_\ell = 0, \end{aligned} \quad (\text{B2})$$

where  $\mathcal{Z}_t = \sqrt{Z_t Z_r^{-1}} Z_\Omega^{(D-2)/2} Z_t^{-1}$ ,  $\mathcal{Z}_r = \sqrt{Z_t Z_r^{-1}} Z_\Omega^{(D-2)/2} Z_r$  and  $\mathcal{Z}_\Omega = \sqrt{Z_t Z_r^{-1}} Z_\Omega^{(D-2)/2} Z_\Omega^{-1}$ .

We will initially assume that  $Z_{\mu\nu}$  has no singularity, and no horizon. In this case a naive application of the WKB approximation to the equation in this form will result in an expression that does poorly for low  $\ell$ , although gives the correct classical phase shift at large  $\ell$ . The origin of this problem as first noted by Langer [89] is that the scattering process here is defined on the line  $r \geq 0$ , and the behavior of the solutions near  $r = 0$  is not well approximated by WKB. This problem is easily resolved performing a coordinate transformation  $r = e^\rho$  which maps the origin at  $r = 0$  to  $\rho = -\infty$ . The correct asymptotic solution near  $\rho = -\infty$  is now the exponentially decaying WKB solution. To proceed we change variables



$$\begin{aligned} & \mathcal{Z}_t \omega^2 \phi_\ell + e^{-(D-1)\rho} \frac{\partial}{\partial \rho} \left( e^{(D-3)\rho} \mathcal{Z}_r \frac{\partial}{\partial \rho} \phi_\ell \right) \\ & - \mathcal{Z}_\Omega \ell(\ell + D - 3) e^{-2\rho} \phi_\ell = 0, \end{aligned} \quad (\text{B3})$$

and then define  $\phi_\ell = e^{-(D-3)\rho/2} (\mathcal{Z}_r)^{-1/2} \chi_\ell$  which puts the equation in the canonical form

$$\frac{\partial^2 \chi_\ell}{\partial \rho^2} = -W_\ell(\rho) \chi_\ell(\rho), \quad (\text{B4})$$

where

$$\begin{aligned} W_\ell(\rho) = & \frac{1}{\mathcal{Z}_r} e^{2\rho} [\mathcal{Z}_t \omega^2 - \mathcal{Z}_\Omega \ell(\ell + D - 3) e^{-2\rho}] \\ & - \frac{(D-3)^2}{4} + \frac{1}{4} \left( \frac{d\mathcal{Z}_r}{d\rho} \right)^2 - \frac{1}{2} (D-3) \frac{d\mathcal{Z}_r}{d\rho} + \frac{d^2 \mathcal{Z}_r}{d\rho^2}. \end{aligned}$$

In the usual case for which  $D = 4$  and  $\mathcal{Z}_r = \mathcal{Z}_t = \mathcal{Z}_\Omega = 1$  this gives

$$W_\ell^{D=4, \mathcal{Z}=1} = e^{2\rho} \omega^2 - \ell(\ell + 1) - \frac{1}{4} = e^{2\rho} \omega^2 - \left( \ell + \frac{1}{2} \right)^2. \quad (\text{B5})$$

As pointed out by Langer, this corresponds to using the standard WKB formula with the replacement  $\ell(\ell + 1) \rightarrow (\ell + 1/2)^2$  which is relevant at low  $\ell$ .

The turning point, i.e., what is interpreted classically as the point of closest approach, is defined by  $W_\ell(\rho_t) = 0$ . For  $\rho < \rho_t$  we have  $W_\ell < 0$  and the desired WKB solution is one that decays exponentially as  $\rho \rightarrow -\infty$

$$\chi_\ell \approx \frac{\bar{\chi}}{(-W_\ell)^{1/4}} e^{-\int_\rho^{\rho_t} \sqrt{-W_\ell} d\rho}, \quad (\text{B6})$$

for some normalization constant  $\bar{\chi}$ . Using the WKB matching formula this matches onto  $\rho > \rho_t$ ,

$$\begin{aligned} \chi_\ell & \approx \frac{\bar{\chi}}{(W_\ell)^{1/4}} \sin \left( \int_{\rho_t}^{\rho} \sqrt{W_\ell} d\rho + \frac{\pi}{4} \right) \\ & \approx \frac{\bar{\chi}}{(W_\ell)^{1/4}} \sin \left( \int_{r_t}^r \sqrt{\frac{1}{\mathcal{Z}_r} \left( \mathcal{Z}_t \omega^2 - \mathcal{Z}_\Omega \frac{(\ell + (D-3)/2)^2}{r^2} - \frac{1}{4r^2} \mathcal{Z}_r \beta_R(r) \right)} dr + \frac{\pi}{4} \right), \end{aligned} \quad (\text{B7})$$

where  $r_t$  is the turning point expressed in terms of  $r$  and

$$\beta_R(r) = (D-3)^2 \left( 1 - \frac{\mathcal{Z}_\Omega}{\mathcal{Z}_r} \right) - r^2 \frac{(\frac{\partial \mathcal{Z}_r}{\partial r})^2}{(\mathcal{Z}_r)^2} + 2 \frac{((D-2)r \frac{\partial \mathcal{Z}_r}{\partial r} + r^2 \frac{\partial^2 \mathcal{Z}_r}{\partial r^2})}{\mathcal{Z}_r}. \quad (\text{B8})$$

In the idealized case in which all components  $\mathcal{Z}_I$  asymptote to unity faster than  $1/r$ , then the scattering phase shifts are determined by requiring that this solution has the asymptotic form

$$\chi_\ell \propto (e^{2i\delta_\ell} e^{i\omega r} + e^{i\pi(D-2)/2} e^{i\pi\ell} e^{-i\omega r}). \quad (\text{B9})$$

Performing the comparison we obtain the standard WKB formula for the partial wave phase shifts

$$\delta_\ell(\omega) = \int_{r_t}^{\infty} dr \left( \sqrt{\frac{1}{\mathcal{Z}_r} \left( \mathcal{Z}_t \omega^2 - \mathcal{Z}_\Omega \frac{(\ell + (D-3)/2)^2}{r^2} - \frac{1}{4r^2} \mathcal{Z}_r \beta_R(r) \right)} - \omega \right) - \omega r_t + \frac{\pi}{2} (\ell + (D-3)/2). \quad (\text{B10})$$

The total Eisenbud-Wigner time delay for each partial wave is then given by

$$\Delta T_\ell = 2 \frac{\partial \delta_\ell(\omega)}{\partial \omega} \quad (\text{B11})$$

$$= 2 \int_{r_t}^{\infty} dr \left( \frac{\mathcal{Z}_t \omega}{\sqrt{\mathcal{Z}_r (\mathcal{Z}_t \omega^2 - \mathcal{Z}_\Omega \frac{(\ell + (D-3)/2)^2}{r^2} - \frac{1}{4r^2} \mathcal{Z}_r \beta_R(r))}} - 1 \right) - 2r_t. \quad (\text{B12})$$

This simplifies in the limit of large  $\ell$ , fixed apparent impact parameter  $b = (\ell + (D-3)/2)/\omega$ , so that  $\beta_R(r)$  may be neglected and we find

$$\Delta T_\ell = 2 \int_{r_t}^{\infty} \left( \frac{Z_t}{\sqrt{Z_r(Z_t - Z_\Omega \frac{b^2}{r^2})}} - 1 \right) dr - 2r_t, \quad (\text{B13})$$

$$= 2 \int_{r_t}^{\infty} \left( \frac{Z_t}{\sqrt{Z_r(Z_t - Z_\Omega \frac{b^2}{r^2})}} \right) dr - 2 \int_b^{\infty} \frac{1}{\sqrt{1 - \frac{b^2}{r^2}}} dr, \quad (\text{B14})$$

which is the ‘‘classical’’ time-delay result for a particle moving along a null geodesic in the metric  $Z_{\mu\nu}$ .

## 2. Dealing with a horizon

To deal with spacetimes with a horizon, it is necessary to modify the Langer transformation and instead of taking  $r = 0$  to  $-\infty$ , we map the horizon to  $-\infty$ . This is achieved by defining an analog of the tortoise coordinates for which the two-dimensional  $r, t$  metric is conformally flat, i.e., for which

$$Z_t^{-1/2} Z_r^{-1/2} dr = d\hat{r}, \quad (\text{B15})$$

so that the metric takes the form  $ds^2 = Z_t(-dt^2 + d\hat{r}^2) + Z_\Omega r^2 d\Omega^2$ . The integration constant in this change of variables can be fixed at infinity for  $D > 4$  as

$$\hat{r} = r - \int_r^{\infty} dr (Z_t^{-1/2} Z_r^{-1/2} - 1). \quad (\text{B16})$$

For  $D = 4$  there is a logarithmic divergence in this integral and an arbitrary finite comparison scale must be chosen has is explained in section C 2. Then with the choice  $\phi_\ell = Z_\Omega^{-(D-2)/4} r^{-(D-2)/2} \chi_\ell$ , the wave equation becomes

$$\frac{d^2 \chi_\ell}{d\hat{r}^2} + \omega^2 \chi_\ell - V_{\text{eff}}(\hat{r}) \chi_\ell = 0, \quad (\text{B17})$$

with the effective potential

$$V_{\text{eff}}(\hat{r}) = \frac{Z_t}{Z_\Omega r^2} \ell(\ell + (D-3)) + \frac{1}{\gamma} \frac{d^2 \gamma}{d\hat{r}^2}, \quad (\text{B18})$$

with  $\gamma = Z_\Omega^{(D-2)/4} r^{(D-2)/2}$ . In the familiar case of the  $D = 4$  Schwarzschild solution this is

$$V_{\text{eff}}(\hat{r}) = \left(1 - \frac{r_g}{r}\right) \left(\frac{\ell(\ell+1)}{r^2} + \frac{r_g}{r^3}\right). \quad (\text{B19})$$

The asymptotic form of the mode functions is in these new coordinates conventional

$$\chi_\ell \propto (e^{2i\hat{r}} e^{i\omega\hat{r}} - (-1)^\ell e^{-i\omega\hat{r}}), \quad (\text{B20})$$

which gives the phase shift

$$\begin{aligned} \hat{\delta}_\ell(\omega) &= \int_{\hat{r}_t}^{\infty} d\hat{r} \left( \sqrt{(\omega^2 - V_{\text{eff}}(\hat{r}))} - \omega \right) - \omega \hat{r}_t \\ &+ \frac{\pi}{2} (\ell + (D-3)/2). \end{aligned} \quad (\text{B21})$$

In deriving this we assume that  $r_t$  is sufficiently larger than the peak of the potential  $V_{\text{eff}}(\hat{r})$ , which for example in  $D = 4$  occurs at approximately  $r = 3r_g/2$  for large  $\ell$ , so that there is sufficient barrier that (B6) is approximately the solution inside the potential barrier. Clearly as the turning point approaches the top of the barrier we must take appropriate consideration of the absorbed waves [90,91]. Translated back into the original coordinates this is

$$\hat{\delta}_\ell(\omega) = \int_{r_t}^{\infty} dr \left( \sqrt{\frac{1}{Z_r} \left( Z_t \omega^2 - Z_\Omega \frac{\ell(\ell + (D-3))}{r^2} - \frac{1}{4r^2} Z_r \beta_H(r) \right)} - \omega \right) - \omega r_t + \frac{\pi}{2} (\ell + (D-3)/2), \quad (\text{B22})$$

which gives a time delay

$$\Delta T_\ell = 2 \int_{r_t}^{\infty} dr \left( \frac{Z_t \omega}{\sqrt{Z_r \left( Z_t \omega^2 - Z_\Omega \frac{\ell(\ell + (D-3))}{r^2} - \frac{1}{4r^2} Z_r \beta_H(r) \right)}} - 1 \right) - 2r_t, \quad (\text{B23})$$

where now

$$\beta_H(r) = \frac{4r^2}{Z_r Z_t} \frac{1}{\gamma} \frac{d^2 \gamma}{d\hat{r}^2}. \quad (\text{B24})$$

This expression is to be compared with (B12). The two differ only in the subleading semiclassical contribution which accounts for the different boundary conditions describing the two different physical situations.

### APPENDIX C: EIKONAL AS A LIMIT OF SEMICLASSICAL PHASE SHIFTS

Since many discussions of causality in effective field theories are phrased in the eikonal or shockwave (Penrose limit) approximation (see for example Refs. [23,27,92]), it is worth showing here that the eikonal approximation can be obtained straightforwardly as a limiting case of the semiclassical approximation, and hence the latter may be regarded as more general. We begin with the phase shift relevant to the 2D conformally flat coordinates  $\hat{r}$  (B21). We split the potential in the form of the usual centrifugal potential plus corrections

$$V_{\text{eff}}(\hat{r}) = \frac{b^2 \omega^2}{\hat{r}^2} + U_{\text{eff}}(\hat{r}), \quad (\text{C1})$$

with  $b = (\ell + (D-3)/2)\omega^{-1}$ . The eikonal approximation corresponds to assuming a high energy limit  $\omega^2 \gg |U_{\text{eff}}(\hat{r})|$  so that we may treat  $U_{\text{eff}}(\hat{r})$  perturbatively. In the present relativistic context this limit is more subtle than it is in nonrelativistic quantum mechanics since  $U_{\text{eff}}(\hat{r})$  itself scales with  $\omega^2$ . Nevertheless its different radial dependence ensures there is always a regime in which we may imagine  $\omega^2 \gg U_{\text{eff}}(\hat{r})$ . Naively we can just perturb the square root in (B21); however, this becomes problematic since the point of closest approach  $\hat{r}_t$  is itself dependent on the potential  $U_{\text{eff}}(\hat{r})$  and a naive expansion will lead to an ill-defined expression. The solution is to use the relation between  $b$  and  $\hat{r}_t$

$$\frac{b^2 \omega^2}{\hat{r}_t^2} + U_{\text{eff}}(\hat{r}_t) = \omega^2, \quad (\text{C2})$$

to rewrite (B21) in the form

$$\hat{\delta}_\ell(\omega) = \int_{\hat{r}_t}^{\infty} d\hat{r} \left( \sqrt{\left( \omega^2 - \frac{\hat{r}_t^2 \omega^2}{\hat{r}^2} - \left( U_{\text{eff}}(\hat{r}) - U_{\text{eff}}(\hat{r}_t) \frac{\hat{r}_t^2}{\hat{r}^2} \right) \right)} - \omega \right) - \omega \hat{r}_t + \frac{\pi}{2} (\ell + (D-3)/2), \quad (\text{C3})$$

which admits a well-defined expansion to any order, for which the first order term in  $U_{\text{eff}}(\hat{r})$  is

$$\begin{aligned} \hat{\delta}_\ell(\omega) &= \int_{\hat{r}_t}^{\infty} d\hat{r} \left( \sqrt{\left( \omega^2 - \frac{\hat{r}_t^2 \omega^2}{\hat{r}^2} \right)} - \omega \right) - \omega \hat{r}_t + \frac{\pi}{2} (\ell + (D-3)/2) \\ &\quad - \frac{1}{2} \int_{\hat{r}_t}^{\infty} d\hat{r} \frac{1}{\sqrt{\omega^2 - \frac{\hat{r}_t^2 \omega^2}{\hat{r}^2}}} \left( U_{\text{eff}}(\hat{r}) - U_{\text{eff}}(\hat{r}_t) \frac{\hat{r}_t^2}{\hat{r}^2} \right) + \dots \end{aligned} \quad (\text{C4})$$

$$= \frac{\pi}{2} \omega (b - \hat{r}_t) - \frac{1}{2} \int_{\hat{r}_t}^{\infty} d\hat{r} \frac{1}{\sqrt{\omega^2 - \frac{\hat{r}_t^2 \omega^2}{\hat{r}^2}}} \left( U_{\text{eff}}(\hat{r}) - U_{\text{eff}}(\hat{r}_t) \frac{\hat{r}_t^2}{\hat{r}^2} \right). \quad (\text{C5})$$

Now from (C2)

$$b - \hat{r}_t = \frac{\hat{r}_t^2 \omega^{-2}}{b + \hat{r}_t} U_{\text{eff}}(\hat{r}_t) \approx \frac{\hat{r}_t}{2\omega^2} U_{\text{eff}}(\hat{r}_t), \quad (\text{C6})$$

and so substituting in (C5) we have

$$\hat{\delta}_\ell(\omega) = -\frac{1}{2\omega} \int_{\hat{r}_t}^{\infty} d\hat{r} \frac{1}{\sqrt{1 - \frac{\hat{r}_t^2}{\hat{r}^2}}} U_{\text{eff}}(\hat{r}) + \dots \quad (\text{C7})$$

To put this in a recognizable form we define  $\hat{r}^2 = \hat{r}_t^2 + z^2$  and notice that to this order we may replace  $\hat{r}_t$  with the apparent impact parameter  $b$  to give

$$\hat{\delta}_\ell(\omega) \approx \hat{\delta}_\ell^{\text{eik}}(\omega) = -\frac{1}{4\omega} \int_{-\infty}^{\infty} dz U_{\text{eff}}(\sqrt{b^2 + z^2}), \quad (\text{C8})$$

which is clearly the eikonal result. This same expression may be derived in a more covariant manner by recognizing that in the classical limit the phase shift  $\delta$  is one-half the difference between the action of a relativistic particle in the curved spacetime  $Z_{\mu\nu}$  and in Minkowski spacetime with the same asymptotic momenta. Writing the action for a massless relativistic particle in phase space form

$$S = \int_{-\infty}^{\infty} d\tau \left( p_\mu \frac{dx^\mu}{d\tau} - \frac{1}{2} Z^{\mu\nu} p_\mu p_\nu \right), \quad (\text{C9})$$

then perturbing around Minkowski we have

$$\begin{aligned}\hat{\delta}^{\text{eik}} &\approx \frac{1}{4} \int_{-\infty}^{\infty} d\tau (\delta Z_{\mu\nu} p^\mu p^\nu) \\ &= \frac{1}{2} \int_b^\infty dr \frac{d\tau}{dr} (\delta Z_{\mu\nu} p^\mu p^\nu),\end{aligned}\quad (\text{C10})$$

where the momentum  $p^\mu$  and velocity are their solutions in Minkowski spacetime written in radial coordinates, i.e.,  $\frac{dr}{d\tau} = \omega\sqrt{1-b^2/r^2}$ ,  $p^t = \omega$ ,  $p^r = \omega\sqrt{1-b^2/r^2}$ ,  $p^\theta = \omega b/r^2$  which for a spherically symmetric background in 2D conformally flat coordinates

$$\begin{aligned}\hat{\delta}^{\text{eik}} &= -\frac{1}{4\omega} \int_{-\infty}^{\infty} dz \frac{\omega^2 b^2}{b^2 + z^2} \\ &\quad \times \left( \delta Z_t \left( \sqrt{b^2 + z^2} \right) - \delta Z_\Omega \left( \sqrt{b^2 + z^2} \right) \right),\end{aligned}\quad (\text{C11})$$

which is consistent with (C8) given  $U_{\text{eff}}(\hat{r}) = \frac{b^2 \omega^2}{\hat{r}^2} (\delta Z_t(\hat{r}) - \delta Z_\Omega(\hat{r}))$  for large  $\ell$  from (B18).

### 1. Shapiro time delay in $D > 4$

To illustrate the applicability of the above formula let us compute the  $D$  dimensional Shapiro delay. We begin with the Schwarzschild-Tangherlini metric in  $D$  dimensions

$$ds^2 = -f(r)dt^2 + f(r)^{-1}dr^2 + r^2 d\Omega_{D-2}^2, \quad (\text{C12})$$

with  $f(r) = 1 - \mu/r^{D-3}$  and  $\mu$  related to the physical source mass  $M_*$  by

$$\mu = \frac{16\pi GM_*}{(D-2)\Omega_{D-2}}, \quad (\text{C13})$$

with  $\Omega_{D-2} = 2\pi^{(D-1)/2}/\Gamma[(D-1)/2]$ . We define the 2D conformal (tortoise) coordinate  $\hat{r}$  via

$$\begin{aligned}\hat{r} &= r - \int_r^\infty dr \left( \frac{1}{f(r)} - 1 \right) \\ &= r - \frac{1}{(D-4)} \frac{\mu}{r^{D-4}} + \dots\end{aligned}\quad (\text{C14})$$

hence to leading order in  $\mu$  the metric is

$$\begin{aligned}ds^2 &\approx \left( 1 - \frac{\mu}{\hat{r}^{D-3}} \right) (-dt^2 + d\hat{r}^2) \\ &\quad + \left( 1 + \frac{2}{(D-4)} \frac{\mu}{\hat{r}^{D-2}} \right) \hat{r}^2 d^2\Omega_{D-2},\end{aligned}\quad (\text{C15})$$

from which we infer the effective potential

$$U_{\text{eff}}(\hat{r}) = -\frac{(D-2)}{(D-4)} b^2 \omega^2 \frac{\mu}{\hat{r}^{D-1}}. \quad (\text{C16})$$

Hence the leading order contribution to the phase shift from (C8) is

$$\delta_\ell(\omega) = \frac{(D-2)}{(D-4)} b^2 \omega \frac{\mu}{4} \int_{-\infty}^{\infty} dz \frac{1}{(b^2 + z^2)^{(D-1)/2}} \quad (\text{C17})$$

$$\begin{aligned}&= \frac{(D-2)}{(D-4)} b^2 \omega \frac{\mu}{4} b^{4-D} \sqrt{\pi} \frac{\Gamma[\frac{D-4}{2}]}{\Gamma[\frac{D-1}{2}]} \\ &= \frac{GM_* \omega}{\pi^{(D-4)/2} b^{D-4}} \Gamma\left[\frac{D-4}{2}\right].\end{aligned}\quad (\text{C18})$$

Denoting the Mandelstam invariant  $s = -p_{\text{total}}^2 = (M_* + \omega)^2 - \omega^2 \approx 2M_*\omega$  for large  $\omega$  this is

$$\delta_\ell(\omega) = \frac{sG}{2\pi^{(D-4)/2} b^{D-4}} \Gamma\left[\frac{D-4}{2}\right]. \quad (\text{C19})$$

In this Lorentz invariant form we may easily translate it into the boosted Penrose limit form considered for example in [23],<sup>16</sup> and so we see the shockwave calculation reproduced precisely the leading term in the semiclassical expansion. Given the phase shift we may now define the time delay for fixed impact parameter  $b$  as

$$\Delta T_b = 2 \left. \frac{\partial \delta_\ell(\omega)}{\partial \omega} \right|_b = \frac{2M_* G}{\pi^{(D-4)/2} b^{D-4}} \Gamma\left[\frac{D-4}{2}\right], \quad (\text{C20})$$

which is the conventional classical Shapiro delay. However, quantum mechanically what is more meaningful is the delay for fixed  $\ell$ , which is the usual definition of the Eisenbud-Wigner time delay for partial waves

$$\Delta T_\ell = 2 \left. \frac{\partial \delta_\ell(\omega)}{\partial \omega} \right|_\ell = \frac{2M_* G(D-3)}{\pi^{(D-4)/2} b^{D-4}} \Gamma\left[\frac{D-4}{2}\right]. \quad (\text{C21})$$

Both expressions are finite and positive and the former plays a crucial role in the discussion in [23].

An alternative derivation of this result is to consider

$$\Delta T_\ell = 2 \int_{r_i}^\infty \frac{dr}{f(r)\sqrt{1-f(r)\frac{b^2}{r^2}}} - 2 \int_b^\infty \frac{dr}{\sqrt{1-\frac{b^2}{r^2}}}, \quad (\text{C22})$$

and following the approach described in Appendix D to leading order in an expansion in  $G$  this is

$$\Delta T_\ell \approx 2 \int_b^\infty dr \frac{1}{\sqrt{1-\frac{b^2}{r^2}}} \frac{\mu}{2b^2 r^{D-3}} [3b^2(D-4) - 2r^2(D-6)] \quad (\text{C23})$$

$$\begin{aligned}&= \frac{b^{4-D}(D-2)\mu\sqrt{\pi}\Gamma[\frac{D-4}{2}]}{2\Gamma[\frac{D-3}{2}]} = \frac{2GM_*(D-3)}{\pi^{(D-4)/2} b^{D-4}} \Gamma\left[\frac{D-4}{2}\right],\end{aligned}\quad (\text{C24})$$

<sup>16</sup>Note our definition of  $\delta_\ell(\omega)$  differs by a factor of 2.

which agrees with (C21). Strictly speaking the integral (C23) is only convergent for  $D > 6$  but the final result is finite and correct for  $D > 4$ .

## 2. Shapiro time delay in $D=4$

The result (B13) only applies if the effective metric asymptotes to Minkowski faster than  $1/r$  to ensure convergence of the integral. In four-dimensional GR this is never the case and (B13) is divergent, a result of the

massless nature of the graviton. This dimensionality dependence may be related to that of Schwarzschild causality established in [93]. The resolution is to account for the attractive Coulomb distortion of the asymptotic wave functions by modifying (B9) to

$$\chi_\ell \propto (e^{2i\delta'_\ell} e^{i\omega r + i\omega \alpha \ln(2\omega r)} - (-1)^\ell e^{-i\omega r - i\omega \alpha \ln(2\omega r)}), \quad (\text{C25})$$

which replaces (B10) with the finite phase shift

$$\delta'_\ell(\omega) = \lim_{R \rightarrow \infty} \left[ \int_{r_t}^R dr \sqrt{\frac{1}{Z_r} \left( Z_t \omega^2 - Z_\Omega \frac{(\ell + 1/2)^2}{r^2} - \frac{1}{4r^2} Z_r \beta_R(r) \right)} + \frac{\pi}{2} (\ell + 1/2) - \omega R - \omega \alpha \ln(2\omega R) \right] \quad (\text{C26})$$

for regular geometries. For geometries with a horizon, the analog of (B22) is replaced by [91]

$$\hat{\delta}_\ell(\omega) = \lim_{R \rightarrow \infty} \left[ \int_{r_t}^R d\hat{r} \left( \sqrt{\frac{1}{Z_r} \left( Z_t \omega^2 - Z_\Omega \frac{\ell(\ell + 1)}{r^2} - \frac{1}{4r^2} Z_r \beta_H(r) \right)} - \omega \right) - \omega r_t + \frac{\pi}{2} \left( \ell + \frac{1}{2} \right) - \omega \alpha \ln(2\omega R) \right], \quad (\text{C27})$$

where the Coulomb distortion scale  $\alpha$  is defined by

$$\alpha = \lim_{r \rightarrow \infty} \left[ \frac{r}{2} \left( \frac{Z_t}{Z_r} - 1 \right) \right], \quad (\text{C28})$$

which in the standard case of an asymptotically Schwarzschild geometry is the Schwarzschild radius  $\alpha = r_g$ . In particular evaluating (C27) on the Schwarzschild metric to leading order in an expansion in  $r_g/b$  gives [91]

$$\begin{aligned} \hat{\delta}_\ell(\omega) &= -\omega r_g \ln(\ell + 1/2) \\ &\quad - \frac{1}{2} \omega r_g \left( 1 + \frac{1}{(\ell + 1/2)^2} \right) + \dots \end{aligned} \quad (\text{C29})$$

Unlike the result in dimensions  $D > 4$  (C23), this corresponds to a time advance

$$\Delta T_\ell = -r_g \left( 2 \ln(\ell + 1/2) + \left( 1 + \frac{1}{(\ell + 1/2)^2} \right) \right). \quad (\text{C30})$$

We should not interpret this as any violation of causality though. Firstly, even in nonrelativistic quantum mechanics time advances do occur, as is well known in the example of scattering from a hard sphere where, for a sphere of radius  $a$ , the scattered wave of speed  $v$  is reflected at a time  $a/v$  before a free wave would reach the center at  $r = 0$  and so is reduced in travel time by  $2a/v$ . For this reason, Wigner's original causality condition for  $S$ -wave scattering is stated as [48,52]

$$\Delta T_{\ell=0} \geq -2a/v, \quad (\text{C31})$$

up to fluctuations at the scale  $\omega^{-1}$ . The implication is that unlike in higher dimensions, the black hole acts like scattering off of a hard sphere with an  $\ell$ -dependent radius of order  $r_g$ . However this is also at the scale of the ambiguity in the definition of the phase due to the divergence from the Coulombic behavior, i.e., the logarithmic term included in (C27) and so we should be careful not to read too much into this. Indeed in defining the tortoise coordinates analogous to (C14) we are faced with a logarithmic divergence which must be cut off at a scale  $C$

$$\begin{aligned} \hat{r} &= r - \int_r^C dr \frac{1}{1 - r_g/r} \\ &= r - r_g \ln((r - r_g)/(C - r_g)). \end{aligned} \quad (\text{C32})$$

In a pure Schwarzschild metric,  $C$  is usually fixed to be  $2r_g$  but in a spacetime which is only asymptotically Schwarzschild there is no requirement for this particular choice. The inherent logarithmic ambiguity in the  $\hat{r}$  coordinate translates into the same ambiguity in the phase shift and hence time delay. As such we will content ourselves with determining time delays relative to the asymptotic Schwarzschild metric for which the ambiguity related to the definition of the phase shift cancels out. This corresponds to the classical criterion [55]. It is noteworthy that these issues are entirely avoided in higher dimensions. Another way to understand these subtleties is to introduce an IR regulator through an effective mass  $\mu$  by replacing the Coulombic  $r_g/r$  with a Yukawa form  $r_g e^{-\mu r}/r$ . This renders the naive phase shift definition (B10) and (B22) finite, and gives a time delay logarithmically sensitive to the IR cutoff. This divergence will cancel in considering

time delay differences as we do throughout, allowing us to take the limit  $\mu \rightarrow 0$ .

Alternatively, from a classical point of view we can regulate the divergence by asking for the time delay for a trajectory that begins and ends at finite radii  $r_b$  and  $r_e$  respectively relative to the same result in Minkowski spacetime with impact parameter  $b$

$$\Delta T_\ell = \int_{r_t}^{r_e} dr \frac{\mathcal{Z}_t}{\sqrt{\mathcal{Z}_r(\mathcal{Z}_t - \mathcal{Z}_\Omega \frac{b^2}{r^2})}} - \int_{r_t}^{r_e} dr \frac{1}{\sqrt{1 - \frac{b^2}{r^2}}} + \int_{r_t}^{r_b} dr \frac{\mathcal{Z}_t}{\sqrt{\mathcal{Z}_r(\mathcal{Z}_t - \mathcal{Z}_\Omega \frac{b^2}{r^2})}} - \int_{r_t}^{r_b} dr \frac{1}{\sqrt{1 - \frac{b^2}{r^2}}}. \quad (\text{C33})$$

with  $f(r) = 1 - r_g/r$  which to first order in  $r_g/r_t$  is<sup>17</sup>

$$\Delta T_\ell^g = r_g \ln \left( \frac{r_e + \sqrt{r_e^2 - r_t^2}}{r_t} \right) + r_g \ln \left( \frac{r_b + \sqrt{r_b^2 - r_t^2}}{r_t} \right) + r_g \sqrt{\frac{r_e - r_t}{r_e + r_t}} + r_g \sqrt{\frac{r_b - r_t}{r_b + r_t}} + (r_e^2 - r_t^2)^{1/2} + (r_b^2 - r_t^2)^{1/2} - (r_e^2 - b^2)^{1/2} - (r_b^2 - b^2)^{1/2} + \mathcal{O}(r_g^2/r_t). \quad (\text{C35})$$

#### APPENDIX D: TIME-DELAY CORRECTIONS

We are generally interested in corrections to the time delay from EFT corrections to the effective background geometry and hence consider the geometry (B1) with now  $Z_{\mu\nu} \rightarrow Z_{\mu\nu} + \delta Z_{\mu\nu}$ . In terms of the corrected factors  $\mathcal{Z}_I \rightarrow \mathcal{Z}_I + \delta \mathcal{Z}_I$  as defined below the wave equation (B2), the correction to the time delay is defined by

$$\Delta T_\ell^{\text{EFT}} = 2 \int_{r_t + \delta r_t}^{\infty} dr \left( \frac{\mathcal{Z}_t + \delta \mathcal{Z}_t}{\sqrt{(\mathcal{Z}_r + \delta \mathcal{Z}_r)(\mathcal{Z}_t + \delta \mathcal{Z}_t - (\mathcal{Z}_\Omega + \delta \mathcal{Z}_\Omega) \frac{b^2}{r^2})}} \right) - 2 \int_{r_t}^{\infty} dr \left( \frac{\mathcal{Z}_t}{\sqrt{\mathcal{Z}_r(\mathcal{Z}_t - \mathcal{Z}_\Omega \frac{b^2}{r^2})}} \right), \quad (\text{D1})$$

where  $r_t$  is the turning point in GR [where  $\mathcal{Z}_t(r_t)r_t^2 = \mathcal{Z}_\Omega(r_t)b^2$ ] and  $r_t + \delta r_t$  is the turning point associated with the effective metric  $Z_{\mu\nu} + \delta Z_{\mu\nu}$ .

For simplicity, we shall denote the integrand in GR as  $\mathcal{A}$  and that in the EFT as  $\mathcal{A} + \delta \mathcal{A}$ ,

$$\mathcal{A}(r) = \frac{\mathcal{Z}_t}{\sqrt{\mathcal{Z}_r(\mathcal{Z}_t - \mathcal{Z}_\Omega \frac{b^2}{r^2})}} \quad (\text{D2})$$

$$\delta \mathcal{A}(r) = -\frac{1}{2} \frac{\mathcal{A}^3}{\mathcal{Z}_t^3} \left( \mathcal{Z}_t(\mathcal{Z}_t \delta \mathcal{Z}_r - \mathcal{Z}_r \delta \mathcal{Z}_t) - (\mathcal{Z}_t \mathcal{Z}_\Omega \delta \mathcal{Z}_r - 2 \mathcal{Z}_r \mathcal{Z}_\Omega \delta \mathcal{Z}_t + \mathcal{Z}_r \mathcal{Z}_t \delta \mathcal{Z}_\Omega) \frac{b^2}{r^2} \right). \quad (\text{D3})$$

To determine the correction  $\Delta T_\ell^{\text{EFT}}$ , we perform in each integral a distinct coordinate transformation  $r \rightarrow \rho$ . For the second integral in (D1) we perform a change of variable defined as  $r = R(\rho)$  so that the relation (D4) below is

Evaluating this for a Schwarzschild geometry we have the well-known Shapiro time delay written in terms of coordinate time  $t$  [51]

$$\Delta T_\ell^g = \int_{r_t}^{r_e} dr \left( \frac{1}{f(r) \sqrt{1 - \frac{b^2}{r^2} f(r)}} \right) + \int_{r_t}^{r_b} dr \left( \frac{1}{f(r) \sqrt{1 - \frac{b^2}{r^2} f(r)}} \right) - (r_e^2 - b^2)^{1/2} - (r_b^2 - b^2)^{1/2}, \quad (\text{C34})$$

<sup>17</sup>This result is more often quoted as the total time signaling time back and forth between  $r_b$  and  $r_e$  in the proper time of an observer at  $r_b$  which in our notation is  $\Delta \tau_{b \rightarrow e \rightarrow b} = 2(1 - r_g/r_b)((r_e^2 - b^2)^{1/2} + (r_b^2 - b^2)^{1/2} + \Delta T_\ell^g)$ .

satisfied and for the first integral in (D1) we perform a change of variable  $r = R(\rho) + \delta R(\rho)$  defined by

$$I_2 : 1 - \frac{1}{\rho^2} = \mathcal{A}^{-2}(r)|_{r=R(\rho)}, \quad (\text{D4})$$

$$I_1 : 1 - \frac{1}{\rho^2} = (\mathcal{A}(r) + \delta \mathcal{A}(r))^{-2}|_{r=R(\rho) + \delta R(\rho)}, \quad (\text{D5})$$

in terms of which the time delay correction is

$$\Delta T_{\ell}^{\text{EFT}} = 2 \int_1^{\infty} \frac{d\rho}{\sqrt{1-1/\rho^2}} \left( \frac{d\delta R(\rho)}{d\rho} \right). \quad (\text{D6})$$

To first order in  $\delta Z_{\mu\nu}$ , the function  $\delta R(\rho)$  is determined by perturbing (D5), giving

$$\delta R(\rho) = - \left. \frac{\delta \mathcal{A}(r)}{\mathcal{A}'(r)} \right|_{r=R(\rho)}. \quad (\text{D7})$$

Substituting this expression back into (D6) and using (D4) to change variables back to  $r$  gives

$$\Delta T_{\ell}^{\text{EFT}} = -2 \int_{r_i}^{\infty} dr \mathcal{A}(r) \frac{d}{dr} \left( \frac{\delta \mathcal{A}(r)}{\mathcal{A}'(r)} \right). \quad (\text{D8})$$

- 
- [1] H. Nussenzveig, *Causality and Dispersion Relations* (Academic Press, New York, 1972), Vol. 95.
- [2] G. Källén, On the definition of the renormalization constants in quantum electrodynamics, *Helv. Phys. Acta* **25**, 417 (1952).
- [3] H. Lehmann, On the properties of propagation functions and renormalization constants of quantized fields, *Nuovo Cimento* **11**, 342 (1954).
- [4] T. N. Pham and T. N. Truong, Evaluation of the derivative quartic terms of the meson chiral Lagrangian from forward dispersion relation, *Phys. Rev. D* **31**, 3027 (1985).
- [5] B. Ananthanarayan, D. Toublan, and G. Wanders, Consistency of the chiral pion pion scattering amplitudes with axiomatic constraints, *Phys. Rev. D* **51**, 1093 (1995).
- [6] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, Causality, analyticity and an IR obstruction to UV completion, *J. High Energy Phys.* **10** (2006) 014.
- [7] B. Bellazzini, Softness and amplitudes' positivity for spinning particles, *J. High Energy Phys.* **02** (2017) 034.
- [8] C. Cheung and G. N. Remmen, Positivity of Curvature-Squared Corrections in Gravity, *Phys. Rev. Lett.* **118**, 051601 (2017).
- [9] J. Bonifacio, K. Hinterbichler, and R. A. Rosen, Positivity constraints for pseudolinear massive spin-2 and vector Galileons, *Phys. Rev. D* **94**, 104001 (2016).
- [10] C. de Rham, S. Melville, A. J. Tolley, and S.-Y. Zhou, Positivity bounds for scalar field theories, *Phys. Rev. D* **96**, 081702 (2017).
- [11] C. de Rham, S. Melville, A. J. Tolley, and S.-Y. Zhou, Massive Galileon positivity bounds, *J. High Energy Phys.* **09** (2017) 072.
- [12] C. de Rham, S. Melville, A. J. Tolley, and S.-Y. Zhou, UV complete me: Positivity bounds for particles with spin, *J. High Energy Phys.* **03** (2018) 011.
- [13] B. Bellazzini, F. Riva, J. Serra, and F. Sgarlata, Beyond Positivity Bounds and the Fate of Massive Gravity, *Phys. Rev. Lett.* **120**, 161101 (2018).
- [14] C. de Rham, S. Melville, A. J. Tolley, and S.-Y. Zhou, Positivity bounds for massive spin-1 and spin-2 fields, *J. High Energy Phys.* **03** (2019) 182.
- [15] C. de Rham, S. Melville, and A. J. Tolley, Improved positivity bounds and massive gravity, *J. High Energy Phys.* **04** (2018) 083.
- [16] C. Zhang and S.-Y. Zhou, Positivity bounds on vector boson scattering at the LHC, *Phys. Rev. D* **100**, 095003 (2019).
- [17] S. Melville and J. Noller, Positivity in the sky: Constraining dark energy and modified gravity from the UV, *Phys. Rev. D* **101**, 021502 (2020).
- [18] L. Alberte, C. de Rham, A. Momeni, J. Rumbutis, and A. J. Tolley, Positivity constraints on interacting spin-2 fields, *J. High Energy Phys.* **03** (2020) 097.
- [19] L. Alberte, C. de Rham, A. Momeni, J. Rumbutis, and A. J. Tolley, Positivity constraints on interacting pseudo-linear spin-2 fields, *J. High Energy Phys.* **07** (2020) 121.
- [20] S. Kim, T. Noumi, K. Takeuchi, and S. Zhou, Heavy spinning particles from signs of primordial non-Gaussianities: Beyond the positivity bounds, *J. High Energy Phys.* **12** (2019) 107.
- [21] Y.-J. Wang, F.-K. Guo, C. Zhang, and S.-Y. Zhou, Generalized positivity bounds on chiral perturbation theory, *J. High Energy Phys.* **07** (2020) 214.
- [22] N. Afkhami-Jeddi, S. Kundu, and A. Tajdini, A conformal collider for holographic CFTs, *J. High Energy Phys.* **10** (2018) 156.
- [23] X. O. Camanho, J. D. Edelstein, J. Maldacena, and A. Zhiboedov, Causality constraints on corrections to the graviton three-point coupling, *J. High Energy Phys.* **02** (2016) 020.
- [24] K. Hinterbichler, A. Joyce, and R. A. Rosen, Massive spin-2 scattering and asymptotic superluminality, *J. High Energy Phys.* **03** (2018) 051.
- [25] J. Bonifacio, K. Hinterbichler, A. Joyce, and R. A. Rosen, Massive and massless spin-2 scattering and asymptotic superluminality, *J. High Energy Phys.* **06** (2018) 075.
- [26] K. Hinterbichler, A. Joyce, and R. A. Rosen, Eikonal scattering and asymptotic superluminality of massless higher spin fields, *Phys. Rev. D* **97**, 125019 (2018).
- [27] M. Accettulli Huber, A. Brandhuber, S. De Angelis, and G. Travaglini, Eikonal phase matrix, deflection angle and time delay in effective field theories of gravity, [arXiv: 2006.02375](https://arxiv.org/abs/2006.02375).
- [28] D. Baumann, D. Green, H. Lee, and R. A. Porto, Signs of Analyticity in Single-Field Inflation, *Phys. Rev. D* **93**, 023523 (2016).
- [29] D. Baumann, D. Green, and T. Hartman, Dynamical constraints on RG flows and cosmology, *J. High Energy Phys.* **12** (2019) 134.
- [30] H. J. Bremermann, R. Oehme, and J. G. Taylor, Proof of dispersion relations in quantized field theories, *Phys. Rev.* **109**, 2178 (1958).

- [31] N. N. Bogoliubov, D. V. Shirkov, and S. Chomet, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959), Vol. 59.
- [32] K. Hepp, On the analyticity properties of the scattering amplitude in relativistic quantum field theory, *Helv. Phys. Acta* (Switzerland) **37**, 639 (1964).
- [33] I. T. Drummond and S. J. Hathrell, QED vacuum polarization in a background gravitational field and its effect on the velocity of photons, *Phys. Rev. D* **22**, 343 (1980).
- [34] R. LaFrance and R. C. Myers, Gravity's rainbow, *Phys. Rev. D* **51**, 2584 (1995).
- [35] T. J. Hollowood and G. M. Shore, Causality and micro-causality in curved spacetime, *Phys. Lett. B* **655**, 67 (2007).
- [36] T. J. Hollowood and G. M. Shore, The refractive index of curved spacetime: The fate of causality in QED, *Nucl. Phys. B* **795**, 138 (2008).
- [37] T. J. Hollowood and G. M. Shore, The causal structure of QED in curved spacetime: Analyticity and the refractive index, *J. High Energy Phys.* **12** (2008) 091.
- [38] T. J. Hollowood, G. M. Shore, and R. J. Stanley, The refractive index of curved spacetime II: QED, Penrose limits and black holes, *J. High Energy Phys.* **08** (2009) 089.
- [39] T. J. Hollowood and G. M. Shore, The effect of gravitational tidal forces on vacuum polarization: How to undress a photon, *Phys. Lett. B* **691**, 279 (2010).
- [40] T. J. Hollowood and G. M. Shore, "Superluminal" photon propagation in QED in curved spacetime is dispersive and causal, [arXiv:1006.1238](https://arxiv.org/abs/1006.1238).
- [41] T. J. Hollowood and G. M. Shore, The effect of gravitational tidal forces on renormalized quantum fields, *J. High Energy Phys.* **02** (2012) 120.
- [42] T. J. Hollowood and G. M. Shore, The unbearable beingness of light, dressing and undressing photons in black hole spacetimes, *Int. J. Mod. Phys. D* **21**, 1241003 (2012).
- [43] T. J. Hollowood and G. M. Shore, Causality violation, gravitational shockwaves and UV completion, *J. High Energy Phys.* **03** (2016) 129.
- [44] G. Goon and K. Hinterbichler, Superluminality, black holes and EFT, *J. High Energy Phys.* **02** (2017) 134.
- [45] C. de Rham and A. J. Tolley, Speed of gravity, *Phys. Rev. D* **101**, 063518 (2020).
- [46] C. de Rham, J. Francfort, and J. Zhang, Black hole gravitational waves in the effective field theory of gravity, *Phys. Rev. D* **102**, 024079 (2020).
- [47] L. Eisenbud, The formal properties of nuclear collisions, Ph. D. thesis, Princeton University, 1948.
- [48] E. P. Wigner, Lower limit for the energy derivative of the scattering phase shift, *Phys. Rev.* **98**, 145 (1955).
- [49] F. T. Smith, Lifetime matrix in collision theory, *Phys. Rev.* **118**, 349 (1960).
- [50] P. A. Martin, On the time delay of simple scattering systems, *Commun. Math. Phys.* **47**, 221 (1976).
- [51] I. I. Shapiro, Fourth Test of General Relativity, *Phys. Rev. Lett.* **13**, 789 (1964).
- [52] C. de Carvalho and H. Nussenzveig, Time delay, *Phys. Rep.* **364**, 83 (2002).
- [53] X. O. Camanho, G. Lucena Gómez, and R. Rahman, Causality constraints on massive gravity, *Phys. Rev. D* **96**, 084007 (2017).
- [54] G. D'Appollonio, P. Di Vecchia, R. Russo, and G. Veneziano, Regge behavior saves string theory from causality violations, *J. High Energy Phys.* **05** (2015) 144.
- [55] S. Gao and R. M. Wald, Theorems on gravitational time delay and related issues, *Classical Quantum Gravity* **17**, 4999 (2000).
- [56] E. Cáceres, A. S. Misobuchi, and J. F. Pedraza, Constraining higher order gravities with subregion duality, *J. High Energy Phys.* **11** (2019) 175.
- [57] P. Milonni, in *Fast Light, Slow Light and Left-Handed Light*, Series in Optics and Optoelectronics (Taylor & Francis, London, 2004), p. 262.
- [58] L. Brillouin, in *Wave Propagation and Group Velocity*, Series in Pure & Applied Physics (Academic Press, New York, 1960), p. 154.
- [59] N. Brunner, V. Scarani, M. Wegmüller, M. Legré, and N. Gisin, Direct Measurement of Superluminal Group Velocity and Signal Velocity in an Optical Fiber, *Phys. Rev. Lett.* **93**, 203902 (2004).
- [60] C. de Rham, Massive gravity, *Living Rev. Relativity* **17**, 7 (2014).
- [61] J. S. Toll, Causality and the dispersion relation: Logical foundations, *Phys. Rev.* **104**, 1760 (1956).
- [62] B. Bellazzini, M. Lewandowski, and J. Serra, Positivity of Amplitudes, Weak Gravity Conjecture, and Modified Gravity, *Phys. Rev. Lett.* **123**, 251103 (2019).
- [63] L. Alberte, C. de Rham, S. Jaitly, and A. J. Tolley, Positivity bounds and the massless spin-2 pole, [arXiv:2007.12667](https://arxiv.org/abs/2007.12667).
- [64] Y. Aharonov, A. Komar, and L. Susskind, Superluminal behavior, causality, and instability, *Phys. Rev.* **182**, 1400 (1969).
- [65] C. de Rham and S. Melville, Gravitational Rainbows: LIGO and Dark Energy at Its Cutoff, *Phys. Rev. Lett.* **121**, 221101 (2018).
- [66] C. De Rham, L. Heisenberg, and A. J. Tolley, Spin-2 fields and the weak gravity conjecture, *Phys. Rev. D* **100**, 104033 (2019).
- [67] R. R. Caldwell, Green's functions for gravitational waves in FRW space-times, *Phys. Rev. D* **48**, 4688 (1993).
- [68] J. Martin and R. H. Brandenberger, Trans-Planckian problem of inflationary cosmology, *Phys. Rev. D* **63**, 123501 (2001).
- [69] T. Jacobson, Trans Planckian redshifts and the substance of the space-time river, *Prog. Theor. Phys. Suppl.* **136**, 1 (1999).
- [70] D. J. Gross and E. Witten, Superstring modifications of Einstein's equations, *Nucl. Phys.* **B277**, 1 (1986).
- [71] R. R. Metsaev and A. A. Tseytlin, Curvature cubed terms in string theory effective actions, *Phys. Lett. B* **185**, 52 (1987).
- [72] T. Bunch and L. Parker, Feynman propagator in curved space-time: A momentum space representation, *Phys. Rev. D* **20**, 2499 (1979).
- [73] G. M. Shore, "Faster than light" photons in gravitational fields: Causality, anomalies and horizons, *Nucl. Phys.* **B460**, 379 (1996).
- [74] G. M. Shore, Accelerating photons with gravitational radiation, *Nucl. Phys.* **B605**, 455 (2001).
- [75] E. P. Wigner and L. Eisenbud, Higher angular momenta and long range interaction in resonance reactions, *Phys. Rev.* **72**, 29 (1947).



- [76] L. Alberte, C. de Rham, S. Jaitly, and A. J. Tolley, Positivity bounds and the massless spin-2 pole, [arXiv:2007.12667](#).
- [77] J. Schwinger, Gauge invariance and mass, *Phys. Rev.* **125**, 397 (1962).
- [78] G. Dvali, G. Gabadadze, and M. Porrati, 4-D gravity on a brane in 5-D Minkowski space, *Phys. Lett. B* **485**, 208 (2000).
- [79] A. O. Barvinsky and G. A. Vilkovisky, The generalized Schwinger-Dewitt technique in gauge theories and quantum gravity, *Phys. Rep.* **119**, 1 (1985).
- [80] A. O. Barvinsky, Yu. V. Gusev, V. V. Zhytnikov, and G. A. Vilkovisky, Asymptotic behaviors of one loop vertices in the gravitational effective action, *Classical Quantum Gravity* **12**, 2157 (1995).
- [81] A. O. Barvinsky, Yu. V. Gusev, G. A. Vilkovisky, and V. V. Zhytnikov, Asymptotic behaviors of the heat kernel in covariant perturbation theory, *J. Math. Phys. (N.Y.)* **35**, 3543 (1994).
- [82] I. G. Avramidi, The covariant technique for calculation of one loop effective action, *Nucl. Phys.* **B355**, 712 (1991).
- [83] I. G. Avramidi, The nonlocal structure of the one loop effective action via partial summation of the asymptotic expansion, *Phys. Lett. B* **236**, 443 (1990).
- [84] A. O. Barvinsky, Yu. V. Gusev, V. V. Zhytnikov, and G. A. Vilkovisky, Covariant perturbation theory. 4. Third order in the curvature, [arXiv:0911.1168](#).
- [85] A. O. Barvinsky, Yu. V. Gusev, G. A. Vilkovisky, and V. V. Zhytnikov, The basis of nonlocal curvature invariants in quantum gravity theory (third order), *J. Math. Phys. (N.Y.)* **35**, 3525 (1994).
- [86] G. A. Vilkovisky, Expectation values and vacuum currents of quantum fields, *Lect. Notes Phys.* **737**, 729 (2008).
- [87] A. Codello and O. Zanusso, On the non-local heat kernel expansion, *J. Math. Phys. (N.Y.)* **54**, 013513 (2013).
- [88] J. F. Donoghue and B. K. El-Menoufi, Nonlocal quantum effects in cosmology: Quantum memory, nonlocal FLRW equations, and singularity avoidance, *Phys. Rev. D* **89**, 104062 (2014).
- [89] R. E. Langer, On the connection formulas and the solutions of the wave equation, *Phys. Rev.* **51**, 669 (1937).
- [90] S. Iyer and C. M. Will, Black-hole normal modes: A WKB approach. I. Foundations and application of a higher-order WKB analysis of potential-barrier scattering, *Phys. Rev. D* **35**, 3621 (1987).
- [91] N. G. Sanchez, Scattering of scalar waves from a Schwarzschild black hole, *J. Math. Phys. (N.Y.)* **17**, 688 (1976).
- [92] G. T. Horowitz and N. Itzhaki, Black holes, shock waves, and causality in the AdS/CFT correspondence, *J. High Energy Phys.* **02** (1999) 010.
- [93] P. Cameron and M. Dunajski, On Schwarzschild causality in higher dimensions, [arXiv:2004.00086](#).