JHEP

Published for SISSA by Springer

Received: October 15, 2015
Accepted: December 28, 2015
Published: February 3, 2016

Causality constraints on corrections to the graviton three-point coupling

Xián O. Camanho,[a] **José D. Edelstein,**[b,c] **Juan Maldacena**[d] **and Alexander Zhiboedov**[e]

[a] *Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Am Mühlenberg 1, D-14476 Golm, Germany*

[b] *Department of Particle Physics and IGFAE, Universidade de Santiago de Compostela, E-15782, Santiago de Compostela, Spain*

[c] *Centro de Estudios Científicos CECs, Casilla 1469, Valdivia, Chile*

[d] *School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, U.S.A.*

[e] *Department of Physics, Princeton University, Princeton, NJ 08544, U.S.A.*

E-mail: xian.camanho@aei.mpg.de, jose.edelstein@usc.es, malda@ias.edu, azhiboed@princeton.edu

Abstract: We consider higher derivative corrections to the graviton three-point coupling within a weakly coupled theory of gravity. Lorentz invariance allows further structures beyond the one present in the Einstein theory. We argue that these are constrained by causality. We devise a thought experiment involving a high energy scattering process which leads to causality violation if the graviton three-point vertex contains the additional structures. This violation cannot be fixed by adding conventional particles with spins $J \leq 2$. But, it can be fixed by adding an infinite tower of extra massive particles with higher spins, $J > 2$. In AdS theories this implies a constraint on the conformal anomaly coefficients $\left|\frac{a-c}{c}\right| \lesssim \frac{1}{\Delta_{\rm gap}^2}$ in terms of $\Delta_{\rm gap}$, the dimension of the lightest single trace operator with spin $J > 2$. For inflation, or de Sitter-like solutions, it indicates the existence of massive higher spin particles if the gravity wave non-gaussianity deviates significantly from the one computed in the Einstein theory.

Keywords: AdS-CFT Correspondence, Models of Quantum Gravity, Classical Theories of Gravity

ArXiv ePrint: 1407.5597

Open Access, © The Authors.
Article funded by SCOAP[3].

doi:10.1007/JHEP02(2016)020

JHEP02(2016)020

Contents

1 Introduction/motivation **1**

2 Flat space causality and shock waves **4**
- 2.1 Statement of flat space causality 4
- 2.2 Scattering through a plane wave in general relativity 5
- 2.3 Connection with the scattering amplitude computation 7
- 2.4 The effect of higher derivative interactions on particles with spin 8

3 General constrains on the on-shell three-point functions **10**
- 3.1 The phase shift in impact parameter representation from three-point functions 10
- 3.2 The possible forms of three-point functions in various theories 14
- 3.3 Problems with higher derivative corrections to the three-point functions 15
- 3.4 Scattering of gravitons in $D > 4$ dimensions 18
- 3.5 Scattering of gravitons in four dimensions 20

4 Fixing the causality problem by adding massive particles **21**
- 4.1 Massive spin two particles do not fix the problem in $D = 4$ 21
- 4.2 Massive spin two particles do not fix the problem in $D > 4$ 23
- 4.3 Exciting the graviton into new particles 24
- 4.4 Massive higher spin particles can solve the problem 26
- 4.5 Compositeness and the extra structures for graviton scattering 27

5 Anti-de Sitter discussion **28**
- 5.1 Motivation: the emergence of bulk locality should happen in the classical theory 28
- 5.2 Statement of AdS causality 28
- 5.3 The effect of higher derivative interactions on particles with spin 30
- 5.4 Implications for a, c in theories with large operator dimensions 32
- 5.5 Implications for dimensions of double trace operators 32

6 Wormholes and time advances **35**

7 Cosmological applications **36**

8 Conclusions **37**
- 8.1 Open problems 38

A Shapiro time delay **40**

B Three-point amplitudes and their sums **41**
- B.1 Scattering of a scalar and a graviton 43

JHEP02(2016)020

C Higher derivative terms from integrating out particles 44
C.1 QED case 44
C.2 Gravity case 45

D Causality and unitarity for a signal model 45

E Scattering in string theory 48

F Properties of the AdS shock wave 51

G Time advances and time machines 52

H Representations that couple to two gravitons 53

I The Weinberg-Witten theorem in D dimensions 55

1 Introduction/motivation

In this paper we consider weakly coupled gravity theories in the tree-level approximation. It is well-known that at long distances such theories should reduce to the Einstein gravity theory. However, at intermediate energies we can have higher derivative corrections. By intermediate energies we mean those that are low enough that the theory is still weakly coupled but high enough that we are sensitive to possible higher derivative corrections. An example of such a theory is weakly coupled string theory where the corrections appear at a length scale $\sqrt{\alpha'}$, which is much larger than the Planck length, l_p. The theory at energies comparable to $1/\sqrt{\alpha'}$ is still weakly coupled. In this case, the higher derivative corrections are accompanied by extra massive higher spin particles which appear at the same scale. For higher energies the description is via a string theory which departs significantly from ordinary local quantum field theory. It is reasonable to expect that this is a generic feature. Namely, that higher derivative corrections only arise due to the presence of extra states with masses comparable to the scales where the higher derivative corrections become important.

The objective of this paper is to sharpen this link for the simplest possible correction, that of the graviton three-point coupling. Due to the fact that the graviton has spin, the flat space on-shell three-point function is not uniquely specified. In general, it has three different possible structures. The most familiar is the one we get in the Einstein theory. The others can be viewed as arising from higher derivative terms in the gravitational action. The first new structure has two more derivatives. Relative to the size of the Einstein-Hilbert term, it scales like αp^2, where α is a new quantity with dimensions of length squared which characterizes the relative importance of the new term. Here p is the typical scale of the momenta, it is not a Mandelstam invariant, since they all vanish for on-shell three-point functions. We work in a regime where both of the three-point vertices are small, so that gravity is weakly coupled.

JHEP02(2016)020

We find that new three-point vertices lead to a potential causality violation unless we get contributions from extra particles. This causality violation is occurring when the theory is still weakly coupled. It occurs in a high-energy, fixed impact parameter, scattering process at a center of mass (energy)^2, s, which is large compared to $1/\alpha$ but still small enough for the coupling to be weak. In General Relativity this scattering process leads to the well-known Shapiro time delay [1], which is one of the classical tests of Einstein's theory [2]. See also [3]. When the graviton three-point vertex is corrected, the new terms can lead to a time advance, depending on the spin of the scattered graviton. At short enough impact parameter this time advance can overwhelm Shapiro's time delay and lead to a causality problem. This troublesome feature arises at an impact parameter of order $b^2 \sim \alpha$. At tree-level, this problem can only be fixed by introducing an infinite number of new massive particles with spin[1] $J > 2$ and m^2 comparable to α^{-1}. In other words, it cannot be fixed by adding particles with spins $J \leq 2$, or by considering the existence of extra dimensions.

These causality constraints are similar in spirit to those considered in [4, 5] but they differ in two ways. First of all, here the problem will be shown to arise for small t/s, but large s. Second, the fact that the graviton has spin is crucial. On the other hand, in both cases we have locally Lorentz-invariant Lagrangians that nevertheless can lead to causality violations in non-trivial backgrounds.

An example of a theory that is constrained by these considerations is given by the action

$$S = l_p^{2-D} \int d^D x \sqrt{g} \left[R + \alpha \left(R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} - 4 R_{\mu\nu} R^{\mu\nu} + R^2 \right) \right] , \tag{1.1}$$

where the second term is the Lanczos-Gauss-Bonnet term.[2] The constant α has dimensions of length squared. For $\alpha \gg l_p^2$, we will show that the theory is not causal. Furthermore, there is no way to make it causal by adding local higher curvature terms. In fact, our discussion refers to on-shell data, namely the three-point function, which reflects the real physical information and does not depend on the particular way that we write the Lagrangian. In other words, the discussion is invariant under field redefinitions.

Note that if we view gravity as a low-energy effective theory with a UV cutoff of order the Planck mass, M_p, and we add higher derivative terms with dimensionless coefficients which are of order one, then we have nothing to say. The remarks in this paper *only apply* to theories where the coefficients of the higher derivative terms are much larger. The "natural" value for the coefficient α if we view (1.1) as an effective gravity theory is $\alpha \sim l_p^2$. We will only constrain larger values of α. Of course, we are discussing this problem because it *is indeed possible* to have theories with $\alpha \gg l_p^2$, for example a weakly coupled string theory.

Another example where this discussion is relevant is the following. Imagine that we consider a large N gauge theory. Such a gauge theory is expected to have a weakly coupled string dual. This theory will have a weakly coupled graviton corresponding to the stress

[1] In $D > 4$ dimensions by spin $J > 2$ particles we mean particles in representations of the little group $SO(D-1)$ with both of the following two properties (see also appendix H): a) their maximal spin projection $J_{+-} \geq 2$; b) their representations are labeled by Young tableaux with three or more boxes.

[2] It is the dimensional continuation of the four-dimensional Euler density. In four dimensions it is a topological term, while in higher dimensions it is not topological, in fact it contributes to the three-point coupling of the graviton.

JHEP02(2016)020

tensor operator [6–8]. However, we are not guaranteed that the dual will be an ordinary Einstein gravity theory. It might even be a Vasiliev-like gravity theory [9, 10]. The field theory three-point functions of the stress tensor [11, 12] determine the three-point functions of the gravity theory [13, 14]. One can imagine a theory where the only light single trace operator is the stress tensor. It is natural to expect that this theory will have an Einstein gravity dual. It would be nice to prove that. For contact scalar interactions, it has been argued that there is a one-to-one correspondence between the solutions of crossing and bulk lagrangians [15], see also [16]. However, one can worry that the size of the vertices with higher derivatives might be comparable to the one in the Einstein theory. As a case in point, we can think about the graviton three-point coupling. This is constrained to be a linear combination of three structures, only one of which is the Einstein one. These extra structures necessarily lead to a causality problem unless we introduce new higher spin particles at the scale that appears in these new three-point functions. Thus we link the three-point function of the stress tensor to the operator spectrum of the theory. These three-point functions were constrained by causality in [14]. Here we get stronger constraints because we are making further assumptions about the operator spectrum of the theory.

As another application we consider the possibility that gravity waves during inflation were generated by a theory that indeed had these higher derivative corrections with a size comparable to the Hubble scale. This is a possibility which is allowed by conformal invariance and would be realized if the dual description to inflation (in the spirit of dS/CFT [17–19]) was a weakly coupled theory or if inflation occurs in a string theory where the string scale is close to the Hubble scale. The theory is still weakly coupled, so that scalar and tensor fluctuations are small. In this case the gravity wave non-gaussianities would be different from the ones in the Einstein theory [20, 21]. The observation of such gravity wave signals, combined with the arguments in this paper, would imply the existence of extra particles with spin $J > 2$ during inflation.

Finally, as a further motivation we should mention the grand dream of deriving the most general weakly coupled consistent theory of gravity. It is quite likely that the only such theory is a string-like theory, broadly defined. We are certainly very far away from this dream, but hopefully our simple observation about three-point functions could be useful. In particular, this observation highlights the importance of spin. Spin is likely to grow in importance as we consider constraints on the four-point function, given what we got for the much simpler three-point function. In [22–24] various interesting constraints were derived by using crossing symmetry and the correct factorization on the pole singularities. The constraints discussed in this paper are additional constraints, not covered by their analysis.

This paper is organized as follows. In section 2 we discuss the notion of asymptotic causality, both for asymptotically flat and asymptotically AdS spacetimes. We also discuss the propagation of particles and fields through a shock wave. The purpose of this discussion is to set the stage for the more general argument in the following section so that it becomes intuitively clearer. We present in section 3 the main thought experiment, which involves only on-shell amplitudes and does not refer explicitly to shock waves. In section 4 we discuss the effects of adding extra massive particles. We show that particles with spins two or less cannot solve the problem. We also discuss the appearance of these

JHEP02(2016)020

massive particles among the final states of the scattering process. In order to argue that the problem persists we offer an alternative presentation of the problem where we consider the analyticity properties of the S matrix in impact parameter representation. We discuss in section 5 various aspects of the AdS version of the thought experiment and its implications for properties of the dual conformal field theory. In section 6 we briefly mention the implications of a possible time advance in the context of wormholes. In section 7 we discuss a cosmological application: we link the possible existence of new structures for the gravity wave non-gaussianity to the presence of higher spin massive particles during inflation. We present our conclusions in section 8. Finally, we include a number of appendices to share some important but rather technical details.

2 Flat space causality and shock waves

In this section we consider the problem in asymptotically flat space, and start by discussing the causal structure. As a motivation for our later discussion we analyze the scattering of a probe graviton from the shock wave. We then present the argument for causality violation in purely on-shell terms.

2.1 Statement of flat space causality

When we consider a gravity theory in asymptotically flat space we expect to be able to define scattering amplitudes. In particular, this presupposes that one can fix the asymptotic structure of the spacetime so that we can compare times between the past and the future asymptotic regions. In dimensions $D > 4$ we expect to be able to do this. Let us give a simple argument. Imagine we have a series of observers that sit at a large distance L from each other, and also from the center. For simplicity, consider them at the vertices of a large spatial hypercube and moving along the time direction. We would like to argue that they can synchronize their clocks. If they were in flat space, they can just send signals to each other. On the other hand if there is an object of mass m in the interior, then the metric components decay as $1/r^{D-3}$. This gives rise to a redshift of order Gm/L^{D-3} at the position of the detectors. More importantly, as a signal travels from one detector to the other, staying at a distance of order L from the center, it will get delayed by an amount $\delta t = \delta L \sim Gm/L^{D-4}$ (see appendix A). In $D \geq 5$, we can make this as small as we want by moving out to large enough L. For $D = 4$ this delay does not go to zero and we have a problem. This has been discussed in [25], and it seems closely related to the soft graviton issues that arise in perturbative attempts to define the gravitational S-matrix. This issue is not present in AdS_4. In order to deal with the $D = 4$ case, Gao and Wald [26] have introduced another notion of causality saying that *we cannot send signals faster than what is allowed by the asymptotic causal structure of the spacetime.* In General Relativity, with the null energy condition, they argued that this notion of causality is respected; see [26] for a precise statement of the theorem. This holds in any number of dimensions. For $D > 4$ this becomes the more naive notion of causality introduced above. Note that this is an asymptotic notion and we are not relying on the locally defined lightcones as in [27].

JHEP02(2016)020

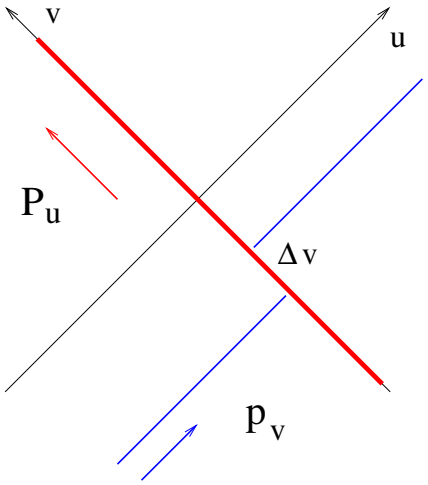


Figure 1. A particle creates a shock wave localized at $u = 0$. A second probe particle propagates on the geometry and experiences a time delay, Δv. The two particles are separated along the transverse directions, which are suppressed in this diagram.

We expect that this notion of causality is actually a *requirement* for any theory of quantum gravity. For asymptotically AdS theories we expect that we should not be able to send signals through the bulk faster than through the boundary. For theories of gravity that are dual to a quantum field theory in the boundary, this is implied by causality in the boundary theory. Also, since we expect that quantum gravity in asymptotically flat space is Lorentz invariant, then these time delays can lead to acausality or closed time-like curves (see appendix G).

Another reason to require this notion is to ensure that Lorentzian wormholes, such as the one obtained form the eternal Schwarzschild black hole and discussed in [28], do not lead to causality violations in the ambient spacetime. This is required by the so-called ER=EPR interpretation of such geometries [29–31]. In the rest of the paper we will *assume* this notion of causality and derive constraints on some higher derivative corrections.

2.2 Scattering through a plane wave in general relativity

A well-known general relativistic effect is that light going near a massive body (e.g. the Sun) would suffer a time delay relative to the same propagation in flat space. This is known as the Shapiro time delay [1] and constitutes one of the classical tests of General Relativity (see appendix A).

We will recall here the derivation of this time delay in the shock wave approximation, which will be all we need for our purposes. First we review the basic properties of shock wave solutions in gravitational theories [32]. This solution describes the gravitational field of an ultrarelativistic particle in a generic theory of gravity [33, 34] and is directly relevant for the high-energy scattering.

JHEP02(2016)020

A generic shock wave solution in flat space can be written in the following form

$$ds^2 = -dudv + h(u, x_i)\, du^2 + \sum_{i=1}^{D-2} (dx_i)^2 \,. \tag{2.1}$$

This geometry admits a covariantly constant null Killing vector $l^\mu \partial_\mu = \partial_v$. See figure 1.

We will be interested in the case when this geometry is sourced by a particle that moves very fast in the v direction. Classically, we can model such particle via the stress tensor [33]

$$T_{uu} = -P_u \delta(u) \delta^{D-2}(\vec{x}) \,, \tag{2.2}$$

where $r = \sqrt{\sum_{i=1}^{D-2} x_i^2}$ and $P_u < 0$ is the momentum of a particle. The Einstein equations then take the form

$$\partial_\perp^2 h(u, x_i) = -16\pi G |P_u| \delta(u) \delta^{D-2}(\vec{x}) \,. \tag{2.3}$$

The solution is

$$h(u, x_i) = \frac{4\Gamma(\frac{D-4}{2})}{\pi^{\frac{D-4}{2}}} \delta(u) \frac{G|P_u|}{r^{D-4}} \,. \tag{2.4}$$

Now we consider a probe particle that moves in the other light cone direction with momentum p_v and is such that it crosses the shock with impact parameter b. In other words, the displacement in the transverse dimensions is $r = b$. The metric (2.1) is a bit peculiar because of the delta function $\delta(u)$ in $h(u, x_i)$. We can remove this delta function at the transverse position b by defining a new coordinate

$$v = v_{\text{new}} + \frac{4\Gamma(\frac{D-4}{2})}{\pi^{\frac{D-4}{2}}} \frac{G|P_u|}{b^{D-4}} \theta(u) \,. \tag{2.5}$$

This then cancels the $\delta(u)$ term in the metric at $r = b$. Thus, the geodesic that goes through this point is continuous in the v_{new} coordinates. This means that in the original coordinates it suffers a shift

$$\Delta v = v_{\text{After crossing}} - v_{\text{Before crossing}} = \frac{4\Gamma(\frac{D-4}{2})}{\pi^{\frac{D-4}{2}}} \frac{G|P_u|}{b^{D-4}} > 0 \,. \tag{2.6}$$

This represents the Shapiro time delay, see figure 1.

In addition to this time delay, the trajectory is also subject to a deflection angle. We might worry that the deflection angle would hamper our ability to see a possible time delay or time advance from far away. In fact, we could consider two shocks in succession, but separated in the transverse direction. Considering the probe particle coming at $r = 0$, we set the shocks at $r = b$ opposite to each other as shown in figure 2. In this case, the probe particle does not get a net deflection, but the time delays add.

It is instructive to reproduce this formula for the time delay when we treat the probe as a quantum mechanical particle. The wave equation for a scalar field takes the form

$$\nabla^2 \phi = 0 \quad \longrightarrow \quad \partial_u \partial_v \phi + h \partial_v^2 \phi - \frac{1}{4} \partial_i^2 \phi = 0 \,, \tag{2.7}$$

JHEP02(2016)020

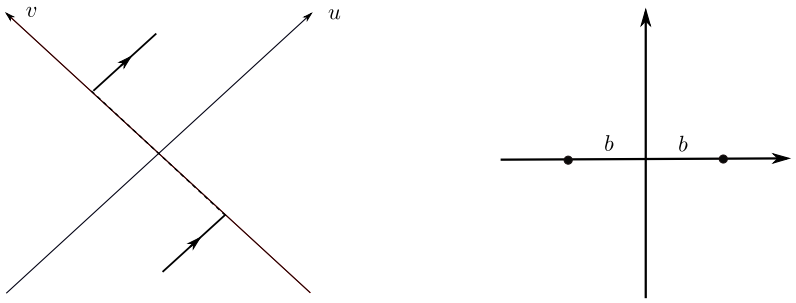


Figure 2. Left: we consider a particle propagating through the superposition of two left moving shock waves localized at $u = 0$. The particle trajectory is given by the arrows. Right: in the transverse plane we separate shock waves by distance $2b$ and send the particle between them so that the net deflection angle is zero. The time delay is the sum of the time delays due to each shock wave.

Let us now consider the change in the value of ϕ from $u = 0^-$ to $u = 0^+$. Since the variation of the h term is much faster than the variation in the other variables, we neglect the ∂_i derivatives and write

$$\begin{aligned}\phi(u = 0^+, v, x^i) &= e^{-\int_{0^-}^{0^+} du h \partial_v} \phi(u = 0^-, v, x^i) \\ &= e^{-\Delta v\, \partial_v} \phi(u = 0^-, v, x^i) = e^{-i \Delta v\, p_v} \phi(u = 0^-, v, x^i)\,, \end{aligned} \tag{2.8}$$

where Δv is given in (2.6). Thus we see that we reproduce the answer we got through the geodesic analysis. Note that $p_v = -i\partial_v$ is the generator of translations in v.

In quantum field theory we would end the discussion here. Any time advance in v, relative to the background Minkowski metric would be a problem. In gravity, the situation is more subtle, in principle, we need to make observations from asymptotically far away. However, in that case, the p_u energy also has a p_v dependence[3] and it contributes positively to the time delay. Since $p_u = \vec{q}^2/4p_v$, this is a small effect for large p_v. However, if we multiply by the total $u-$time elapsed, it can add to a very big time delay. Here we will assume that we can sit far enough from the shock to be able to neglect the dynamics of spacetime, but close enough that we can neglect the $v-$time delay produced by the p_u energy. This seems possible for small G.

2.3 Connection with the scattering amplitude computation

Let us reproduce the computation above using scattering amplitudes [3]. It is well-known that the shock wave computation can be reproduced using the so-called eikonal approximation [35]. Consider the scattering amplitude for gravitating scalar particles. It is given by

$$\mathcal{A}_{\text{tree}}(s, t) = -8\pi G \frac{s^2}{t}\,. \tag{2.9}$$

The eikonal approximation resums a particular set of diagrams (horizontal ladders) in the deflectionless limit when $\frac{t}{s} \to 0$. Under favorable circumstances,[4] the amplitude exponen-

[3] See section 3 for the relevant kinematic configuration.

[4] This particular point will be discussed later in more detail.

JHEP02(2016)020

tiates in the impact parameter space (see e.g. [36, 37])

$$i\mathcal{A}_{eik} = 2s \int d^{D-2}\vec{b}\; e^{-i\vec{q}.\vec{b}} \left[e^{i\delta(b,s)} - 1 \right] , \tag{2.10}$$

where the phase is given by

$$\delta(b,s) = \frac{1}{2s} \int \frac{d^{D-2}\vec{q}}{(2\pi)^{D-2}} e^{i\vec{q}.\vec{b}} \mathcal{A}_{\text{tree}}(s, -\vec{q}^{\,2}) = \frac{\Gamma(\frac{D-4}{2})}{\pi^{\frac{D-4}{2}}} \frac{Gs}{b^{D-4}} . \tag{2.11}$$

This result matches the shock wave computation

$$\delta(b,s) = -p_v \Delta v_{|r=b} , \tag{2.12}$$

where we used $s = 4P_u p_v$. As we will see below this picture can be naturally generalized to the scattering of particles with spin.

2.4 The effect of higher derivative interactions on particles with spin

If we had a photon propagating through the plane wave, it will have the same time delay we computed in (2.6). However, if the Lagrangian contains certain higher order interactions, such as

$$S = \int d^D x \sqrt{-g} \frac{1}{4} \left[F_{\mu\nu} F^{\mu\nu} + \hat{\alpha}_2 R^{\mu\nu}{}_{\sigma\delta} F_{\mu\nu} F^{\sigma\delta} \right] , \tag{2.13}$$

then the second term gives rise to a time delay that depends on the polarization of the electromagnetic wave. The equations of motion take the form

$$\nabla^\mu F_{\mu\nu} - \hat{\alpha}_2 \, R_\nu{}^{\mu\alpha\beta} \nabla_\mu F_{\alpha\beta} = 0 . \tag{2.14}$$

In the shock wave background we have $R_{uiju} = \frac{1}{2}\partial_i \partial_j h$. For large p_v (2.14) is reduced to

$$\partial_u F_{vi} + (\delta_{ij} h + \hat{\alpha}_2 \partial_i \partial_j h) \, \partial_v F_{vj} = 0 , \tag{2.15}$$

where we made use of the Bianchi identity in that limit, $\partial_u F_{vi} = \partial_v F_{vi}$; and the total time delay then has the form

$$\begin{aligned} \Delta v &= \left[1 + \hat{\alpha}_2 \frac{\epsilon^i \epsilon^j \partial_{b^i} \partial_{b^j}}{\epsilon^i \epsilon^i} \right] \frac{4\Gamma(\frac{D-4}{2})}{\pi^{\frac{D-4}{2}}} \frac{G|P_u|}{b^{D-4}} \\ &= \frac{4\Gamma(\frac{D-4}{2})}{\pi^{\frac{D-4}{2}}} \frac{G|P_u|}{b^{D-4}} \left[1 + \frac{\hat{\alpha}_2 (D-4)(D-2)}{b^2} \left(\frac{(\epsilon.n)^2}{\epsilon.\epsilon} - \frac{1}{D-2} \right) \right] , \end{aligned} \tag{2.16}$$

where ϵ is the (real) transverse linear polarization direction of the electromagnetic wave. We also introduced $\vec{n} \equiv \frac{\vec{b}}{b}$. The derivatives in (2.16) come from the derivatives present in the Riemann tensor. The second term in (2.13) can be viewed as a spin dependent gravitational force.

We see that as b becomes small, $b^2 < \hat{\alpha}_2$, the second term in (2.16) can overwhelm the first and, depending on the sign of $\hat{\alpha}_2$ and the polarization, may lead to time advance instead of time delay. If the polarization is along $\vec{b}$, it has one sign, while if the polarization vector is orthogonal to $\vec{b}$, $\vec{\epsilon}.\vec{n} = 0$, then it has the other sign. If the polarization is a linear

JHEP02(2016)020

combination of these two, we first decompose the wave in linear polarizations along and perpendicular to $\vec{b}$ and then exponentiate (2.16) for each of these two cases separately, and then add the two results. In other words, the expression for the time delay is now a matrix that can be diagonalized by choosing the polarizations to be along $\vec{b}$ or perpendicular to $\vec{b}$. In conclusion, for either choice of the sign of $\hat{\alpha}_2$, there is a choice of polarization that can lead to time advance.

Thus, if we require the theory to be causal, we see that $\hat{\alpha}_2$ should be set to zero. More precisely, it should be small enough so that the computation we did above breaks down for some reason. An example of a theory where the coupling in (2.13) arises at tree level is bosonic string theory [38, 39]. We will see later that in string theory the potential causality problem is fixed by the presence of extra massive states.

As another example, let us consider the Gauss-Bonnet theory. This consists of the usual gravity action plus a specific R^2 interaction of the form

$$S = \frac{1}{16\pi G}\int d^D x\sqrt{-g}\left(R + \lambda_{GB}\left[R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2\right]\right) . \tag{2.17}$$

The term in brackets is a topological invariant in $D = 4$, but it is not topological for $D > 4$. This theory has been extensively studied because it has the nice feature that the equations for small fluctuations around any background are second order [40].

As explained in [34] the shock wave solution (2.1) is also an exact solution in the Gauss-Bonnet theory as well. We can consider propagation of a gravitational perturbation through the shock wave background. Before and after the shock the graviton moves as in flat space. All we need to know is what happens when it crosses the shock.

We consider a high-energy graviton δh_{ij} that propagates in the v direction with momentum p_v and traceless polarization in the transverse plane. Near the shock we approximate the equations as

$$\partial_u\partial_v\delta h_{ij} + (\delta_{ik} + 4\lambda_{GB}\partial_i\partial_k h)\,\partial_v^2\delta h_{kj} = 0\,. \tag{2.18}$$

Using (2.18) we can find the time delay which takes the following form

$$\begin{aligned}\Delta v &= \left[1 + 4\lambda_{GB}\frac{\epsilon^{ik}\epsilon^{jk}\partial_{b^i}\partial_{b^j}}{\epsilon.\epsilon}\right]\frac{4\Gamma(\frac{D-4}{2})}{\pi^{\frac{D-4}{2}}}\frac{G|P_u|}{b^{D-4}} \\ &= \frac{4\Gamma(\frac{D-4}{2})}{\pi^{\frac{D-4}{2}}}\frac{G|P_u|}{b^{D-4}}\left[1 + \frac{4\lambda_{GB}(D-4)(D-2)}{b^2}\left(\frac{(\epsilon.n)^2}{\epsilon.\epsilon} - \frac{1}{D-2}\right)\right] .\end{aligned} \tag{2.19}$$

Again, by choosing different polarizations we can get time advance for $b^2 \sim |\lambda_{GB}|$ for any sign of λ_{GB}. Notice that the formula for the time delay is very similar to the ones discussed in the context of energy correlators in AdS/CFT with the parameters depending on the impact parameter of scattering b (see e.g. [41]). We will see below that in the case of AdS the causality constraint interpolates between the usual energy correlator bounds and the flat space bounds obtained above.

So we see that imposing positivity in all channels exclude λ_{GB} completely unless new physics kicks in. In other words, the purely Gauss-Bonnet theory (2.17) is acausal. As an

JHEP02(2016)020

aside, the Gauss-Bonnet theory was found in [42–44] to violate the second law of black hole thermodynamics in certain processes.[5]

We could also consider fermions coupled to gravity. At the level of the on-shell scattering amplitudes there is an additional spin-flipping structure that leads to time advance. This additional structure does not come from any known local Lagrangian and is known to be ruled out by considering consistency conditions imposed on the four-point amplitude [24], at least, in four dimensions. If the same is true in all dimensions then the effects that we are discussing could not be observed for fermions. We leave exploration of this point for the future. Note that having an explicit local Lagrangian which leads to second order equations of motions guarantees that all consistency conditions for the four-point amplitude imposed in [24] are obeyed. Thus, none of the problems discussed in [24] arise in the case of photons or gravitons with the modified three-point functions that we are discussing.

3 General constrains on the on-shell three-point functions

The examples we have discussed so far have shown that there are causality issues with specific theories. In this section we will isolate the crucial elements that produce the problem. It turns out that these causality problems are produced purely and exclusively by the form of the *on-shell* three-point functions of the theory. Therefore they are insensitive to higher order contact terms. First we will explain why the three-point functions determine the time delay. Then we will recall some of the properties of three-point functions. Finally, we will present a thought experiment that makes the causality violation more manifest.

3.1 The phase shift in impact parameter representation from three-point functions

Let us consider the tree-level four-point amplitude $\mathcal{A}_4$. It depends on the kinematic invariants that we can produce with the four on-shell conserved momenta and polarization tensors of the external particles. Since it is a tree-level amplitude its only singularities are poles in the s, t and u Mandelstam variables. We can now consider external momenta such that $s \gg t$, but s small enough that the theory is still weakly coupled. We can take the first incoming particle to have very large momentum along p_u and the second with large momentum p_v.

One can show that, in impact parameter space the amplitude is given by (for review see appendix B)

$$\delta(\vec{b}, s) = \frac{1}{2s} \int \frac{d^{D-2}\vec{q}}{(2\pi)^{D-2}} e^{i\vec{q}.\vec{b}} \mathcal{A}_4(\vec{q}) \,. \tag{3.1}$$

[5] More precisely, [44] considered the compactification of the Gauss-Bonnet theory to four dimensions on a circle. Then the Gauss-Bonnet term becomes topological. This leads to a constant contribution to the black hole entropy. The sign of this constant contribution depends on the coefficient λ. When this contribution is negative, a small enough black hole can have negative entropy. When the constant is positive, a merger of two black holes can violate the second law, see [43]. The constraints arise when the Schwarzschild radius $r_S^2 \sim \lambda_{GB}$. In this sense they are similar to the ones we have. In both cases one needs $\lambda_{GB} \ll l_{Planck}$ in order to have a meaningful statement. [43] has the nice feature of also constraining purely topological terms in four dimensions. On the other hand, they might be modified by higher derivative corrections to the action, which could also modify black hole thermodynamics.

JHEP02(2016)020

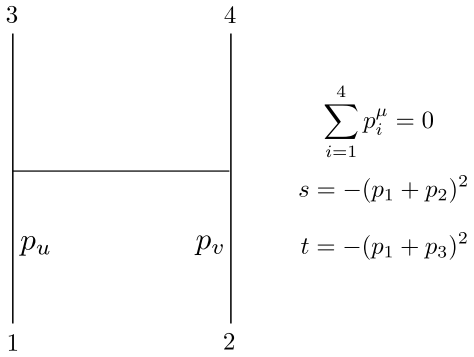


Figure 3. Kinematics of the two-to-two scattering that we are interested in. Particle 1 has a very large momentum p_u and particle two has a very large momentum p_v.

where $\mathcal{A}_4(\vec{q})$ is really a short-hand for the four-point amplitude evaluated in the following momentum configuration

$$\begin{aligned} p_{1\,\mu} &= \left(p_u, \frac{q^2}{16p_u}, \frac{\vec{q}}{2}\right) , \qquad p_{2\,\mu} = \left(\frac{q^2}{16p_v}, p_v, -\frac{\vec{q}}{2}\right) , \\ p_{3\,\mu} &= -\left(p_u, \frac{q^2}{16p_u}, -\frac{\vec{q}}{2}\right) , \quad p_{4\,\mu} = -\left(\frac{q^2}{16p_v}, p_v, \frac{\vec{q}}{2}\right) , \end{aligned} \tag{3.2}$$

with $s \simeq 4p_u p_v$ and $t \simeq -(\vec{q})^2$, where in all cases we indicated only the leading order term in the t/s expansion, assuming $t/s \ll 1$.

If we had scalar particles, the amplitude would depend only on the Mandelstam invariants and we can write $\mathcal{A}_4(\vec{q}) = \mathcal{A}_4(s, t = -\vec{q}^2)$. However, in the case of particles with spin, the amplitude depends also on the polarization vectors contracted with the various momenta.

Let us assume that $\vec{b}$ in (3.1) is chosen along $\vec{b} = (b, 0, 0, \cdots)$, $b = |\vec{b}|$. Then let us consider the integral over the first component of $\vec{q}$, call it q_1. Due to the exponential factor in (3.1) this integrand is suppressed if we give q_1 a positive imaginary part. Here we assumed that the amplitude does not grow exponentially. This is true if we consider particles with polynomial interactions.

Setting $q_1 = i\kappa$+real, we see that the exponential in (3.1) is suppressed in this region as $e^{-\kappa b}$, $\kappa > 0$. Thus we would get a vanishing result (in the large κ limit) unless we cross poles under this contour shift. In fact, we do cross poles. For example the pole at $t = m^2$ coming from the exchange of a particle of mass m in the t-channel leads to a pole at

$$\kappa_*^2 = m^2 + (\vec{q}_{\text{rest}})^2 , \tag{3.3}$$

where $\vec{q}_{\text{rest}}$ are the rest of the components of $\vec{q}$ except the first one. These are still real. The residue of the pole is given by a product of on-shell three-point functions. These three-point functions are non-vanishing because the intermediate momentum has one imaginary and one real component. Thus, we can think of the whole on-shell three-point function as being in two time directions. Note that this is a particular computation where the on-shell three-point function is relevant in ordinary signature. Notice that somewhat abstract notion of

JHEP02(2016)020

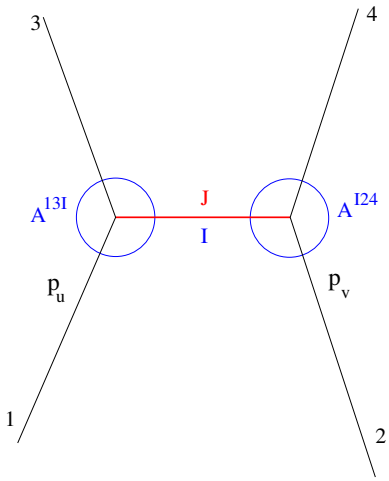


Figure 4. At fixed impact parameter we find a contribution to the amplitude given by the on-shell propagator in the intermediate channel, labeled by the letter I. The contribution involves a product of two on-shell intermediate amplitudes, $\mathcal{A}^{13I}$ and $\mathcal{A}^{I24}$, which are circled in the figure. Particles 1 and 3 carry a large momentum p_u and 2 and 4 have a large p_v. We get a factor of s^J from contracting p_u on the left side with p_v on the right side through the sum over polarizations of the intermediate particle.

complex factorization in mixed signature [23, 45] spacetimes has a direct physical meaning in the context of computing scattering amplitudes in the impact parameter representation.

The pole (3.3) gives a contribution to the amplitude of the form $e^{-\kappa_* b}$. For a massive particle this gives something going like e^{-mb} for large mb, a Yukawa-like potential.[6] For a massless particle, the integral over the rest of the components of $\vec{q}$ produces an inverse power of b, $1/b^{D-4}$. In addition, factors of $\vec{q}$ which are contracted with the polarization tensors give derivatives with respect to $\vec{b}$. In a theory which only contains a massless graviton we just get the massless graviton pole.

It should be noted that in impact parameter representation, for non-zero b, we only get a contribution from the diagram that contains an on-shell particle in the t-channel. In particular, s-channel exchanges do not contribute. The reason is that the two incoming particles have to actually overlap in order to give rise to the intermediate particle in the s-channel. Similarly, a four point contact interaction does not contribute for the same reasons. In both of these remarks we used that we are looking at the tree level diagrams to leading order in the weak coupling expansion. At higher orders there can be other contributions. However, since we are at weak coupling, we can ignore them. In these paragraph, we have used that the interactions are local. Any non-local effect has to come from the propagation of a physical particle in the t-channel.[7]

[6] More precisely, after we integrate over the rest of the components of $\vec{q}$ we get the following expression $(2\pi)^{\frac{2-D}{2}} \left(\frac{m}{b}\right)^{\frac{D-4}{2}} K_{\frac{D-4}{2}}(mb)$.

[7] If our theory contains extended objects, such as strings, we can actually create particles in the s-channel by going to small enough (but non-zero) impact parameter. We discuss this below.

JHEP02(2016)020

We are interested in the part of the amplitude that increases fastest with the Mandelstam variable s. Since there are no kinematic invariants in the on-shell three-point amplitudes, a factor of s can only come from factors of p_u or p_v that are in each of the three-point function contracted, via the sum of the polarizations of the intermediate particle, see figure figure 4 . With a particle of spin J, the maximum power we can get in the amplitude is $\mathcal{A}_4 \sim s^J$. This leads to a factor of s^{J-1} in the phase shift (3.1). In fact, this sum over polarizations over the intermediate state receives a large contribution only from one specific polarization, when the intermediate state has all u indices. In other words, the polarization tensor in the intermediate state appearing in the left three-point function has the form $\epsilon^{I\,u\cdots u}$ and the one in the right three-point function is $\epsilon^{I\,v\cdots v}$. These are contracted with the corresponding factors of the large momentum to give $(p_u p_v)^J \sim s^J$.

The polarization tensors of the external particles can be written as products of vectors $\epsilon_{\mu\nu} = \epsilon_\mu \epsilon_\nu$ where for particles one and three we can write

$$\epsilon^{1\,\mu} = \left(-\frac{\vec{q}.\vec{e}_1}{2p_u}, 0, \vec{e}_1 \right), \qquad \epsilon^{3\,\mu} = \left(\frac{\vec{q}.\vec{e}_3}{2p_u}, 0, \vec{e}_3 \right). \tag{3.4}$$

where $\vec{e}_{1,3}$ are vectors in the purely transverse directions. The on-shell three-point functions will contain factors of the form

$$\epsilon^1.\epsilon^3 = \vec{e}_1.\vec{e}_3\,, \qquad \epsilon^1.p^3 = \vec{q}.\vec{e}_1\,, \qquad \epsilon^3.p^1 = \vec{q}.\vec{e}_3\,. \tag{3.5}$$

The conclusion is that we can think of the polarizations of the external particles as contained effectively in the transverse space. In addition, when we contract them with the external states we get factors of $\vec{q}$. These translate into derivatives with respect to $\partial_{\vec{b}}$.

The final result from a massless pole is

$$\begin{aligned} \delta(\vec{b}, s) &= \frac{\mathcal{A}_3^{13I}(-i\partial_{\vec{b}})\mathcal{A}_3^{I24}(-i\partial_{\vec{b}})}{2s} \int \frac{d^{D-2}\vec{q}}{(2\pi)^{D-2}} \frac{e^{i\vec{q}.\vec{b}}}{\vec{q}^2} \\ &= \frac{\Gamma(\frac{D-4}{2})}{4\pi^{\frac{D-2}{2}}} \frac{\mathcal{A}_3^{13I}(-i\partial_{\vec{b}})\mathcal{A}_3^{I24}(-i\partial_{\vec{b}})}{2s} \frac{1}{|\vec{b}|^{D-4}}\,. \end{aligned} \tag{3.6}$$

where the three-point functions are evaluated on three momenta (3.2). As explained above, in each three-point function there is only one relevant polarization for the intermediate state that produces the factor of s^J. The polarizations of the external particles can be viewed as living purely in the transverse space, after we use (3.4) and (3.5).

In this way we can compute $\delta(\vec{b}, s)$ using on-shell methods. For more details see appendix B. This gives the phase shift to leading order in perturbation theory. The answers agree with with the infinitesimal form of the phase shift computed using the shock wave computation. In fact, the shock wave itself can be viewed as the on-shell graviton intermediate state we discussed above. The fact that the metric contains only the h_{uu} component is related to the fact that only one polarization of the intermediate state produces the most factors of s.

JHEP02(2016)020

3.2 The possible forms of three-point functions in various theories

Given that the answer for the time delay depends on the on-shell three-point functions, it is useful to recall their possible structures. In four dimensions we can use the usual helicity basis. The Einstein-Hilbert gravity action gives rise to the $++-$ and $--+$ three-point functions. We can also have $+++$ and $---$ structures which could come from $(\text{Riemann})^3$ terms. There are two combinations, one of them being parity violating.

In higher dimensions, $D > 4$, we have three possible structures for the graviton three-point functions, all parity preserving. They have the schematic structure, writing $\epsilon_{\mu\nu} = \epsilon_\mu \epsilon_\nu$,

$$\begin{aligned} \mathcal{A}_R &= (\epsilon_1.\epsilon_2\epsilon_3.p_1 + \epsilon_1.\epsilon_3\epsilon_2.p_3 + \epsilon_2.\epsilon_3\epsilon_1.p_2)^2 \,, \\ \mathcal{A}_{R^2} &= (\epsilon_1.\epsilon_2\epsilon_3.p_1 + \epsilon_1.\epsilon_3\epsilon_2.p_3 + \epsilon_2.\epsilon_3\epsilon_1.p_2)\epsilon_1.p_2\epsilon_2.p_3\epsilon_3.p_1 \,, \\ \mathcal{A}_{R^3} &= (\epsilon_1.p_2\epsilon_2.p_3\epsilon_3.p_1)^2 \,. \end{aligned} \tag{3.7}$$

The first is the usual one coming from Einstein gravity, while the second can arise from the Lanczos-Gauss-Bonnet term (2.17). The third one can arise from a $(\text{Riemann})^3$ term.[8] They can be viewed as products of the two possible structures that we can have for on-shell spin one particles. In a given theory, the total three-point function is given by a linear combination of these three answers

$$\mathcal{A}_{ggg} = \sqrt{32\pi G}\,[\mathcal{A}_R + \alpha_2\,\mathcal{A}_{R^2} + \alpha_4 \mathcal{A}_{R^3}] \,, \tag{3.8}$$

where α_2 and α_4 are two parameters with dimension of $(\text{length})^2$ and $(\text{length})^4$ respectively. Notice that we have an overall coupling, given by G, which we take to be parametrically smaller than the other two parameters.

In the high energy limit the three-point functions appearing in (3.6) simplify further and become

$$\begin{aligned} \mathcal{A}_R^{13I} &= 2p_u^2(\vec{e}_1.\vec{e}_3)^2 \,, \\ \mathcal{A}_{R^2}^{13I} &= 2p_u^2(\vec{e}_1.\vec{e}_3)(\vec{q}.\vec{e}_1)(\vec{q}.\vec{e}_3) \,, \\ \mathcal{A}_{R^3}^{13I} &= 2p_u^2(\vec{q}.\vec{e}_1)^2(\vec{q}.\vec{e}_3)^2 \,, \end{aligned} \tag{3.9}$$

for the three terms in (3.7). We used (3.4). The amplitude $\mathcal{A}^{I24}$ has a similar expression with $p_u \to p_v$ and $1,3 \to 2,4$. The reader can express these in terms of transverse traceless two index tensors by using the replacement rule $e_i e_j \to e_{ij}$ in (3.9) , where i, j are indices in the $D-2$ transverse directions.

Note that the third structure in (3.7) or (3.9) is not allowed in a supersymmetric theory. This can be seen as follows. In $D = 4$ this structure gives rise to $+++$ and $---$ amplitudes. However, by the methods described in [46] it is possible to show that

[8] $(\text{Riemann})^3 = R_{\mu\nu\sigma\delta}R^{\sigma\delta\rho\gamma}R_{\rho\gamma}{}^{\mu\nu}$. In four dimensions we can replace one of the curvatures by $R_{\mu\nu\sigma\delta} \to \widetilde{R}_{\mu\nu\sigma\delta} = \epsilon_{\mu\nu\rho\gamma}R^{\rho\gamma}{}_{\sigma\delta}$ to obtain the parity violating term. This parity violating term gives rise to the three-point vertex $\epsilon_{\mu\nu\delta\sigma}p_1^\delta p_2^\sigma \epsilon_{1,\mu\mu'}\epsilon_{2,\nu\nu'}p_2^{\mu'}p_1^{\nu'}\epsilon_{3,\gamma\rho}p_1^\rho p_2^\gamma$, which, despite appearances, is properly symmetric under exchange of any of the three particles. When this term is present the coefficient of the $+++$ amplitude is complex and the coefficient of the $---$ amplitude is the complex conjugate.

JHEP02(2016)020

supersymmetry implies that this structure should be set to zero. This is actually true in all dimensions. The reason is that we can start with a theory in D dimensions and consider external graviton three-point functions with four-dimensional kinematics. If we had a non-zero contribution from the third structure in D dimensions, then it would lead to a contribution to the $+++$ or $---$ amplitudes in four-dimensional kinematics. Since the arguments in [46] are purely kinematical, based on the symmetries of the theory, then they also force the amplitudes to vanish. In conclusion, supersymmetry implies that $\alpha_4 = 0$. In the heterotic string we have a non-zero α_2. And this is also true for compactifications of the string to D dimensions. Thus α_2 is compatible with half maximal supersymmetry. With maximal supersymmetry, e.g. $\mathcal{N} = 8$ in $D = 4$, we should also have $\alpha_2 = 0$. This can be understood from the fact that the four point amplitude for the full supergravity multiplet is determined up to a unique function. Thus, there is no freedom to introduce the polarization dependent terms that would arise if we had the freedom to switch on α_2. Of course, in standard maximal supergravity α_2 is not present, therefore the only value consistent with maximal supersymmetry is $\alpha_2 = 0$. As an aside, notice that this is also related to the fact that in four-dimensional $\mathcal{N} = 4$ superconformal theories, the three-point functions of the stress tensor are completely fixed by supersymmetry in terms of the two-point functions.

Between a graviton and two photons the number of possible three-point functions is two

$$\begin{aligned} \mathcal{A}^{13I}_{F^2} &= \epsilon^I_{\mu\nu}\left[p_1^\mu p_3^\nu(\epsilon_1.\epsilon_3) - \epsilon_1^\mu p_3^\nu(\epsilon_3.p_1) - \epsilon_3^\mu p_1^\nu(\epsilon_1.p_3)\right] , \\ \mathcal{A}^{13I}_{RFF} &= \epsilon^I_{\mu\nu} p_1^\mu p_3^\nu(\epsilon_1.p_3)(\epsilon_3.p_1) , \end{aligned} \tag{3.10}$$

so that $\mathcal{A}_{g\gamma\gamma} = \sqrt{32\pi G}\,[\mathcal{A}_{F^2} + \hat{\alpha}_2\,\mathcal{A}_{RFF}]$. One arising from the usual electrodynamics and the other from the second term in (2.13). Again this second structure is forbidden in a supersymmetric theory. As we did in (3.9) , in the high energy limit we can write them as

$$\begin{aligned} \mathcal{A}^{13I}_{F^2} &= 2p_u^2(\vec{e}_1.\vec{e}_3) , \\ \mathcal{A}^{13I}_{RFF} &= 2p_u^2(\vec{q}.\vec{e}_1)(\vec{q}.\vec{e}_3) . \end{aligned} \tag{3.11}$$

We emphasize that we work here with the on-shell three-point functions, independently of the precise way we write the Lagrangian. This discussion depends only on on-shell three-point functions and not on other contact terms. Any contact four point interaction does not give rise (at tree level) to the long range force at a non-zero value of the impact parameter.

3.3 Problems with higher derivative corrections to the three-point functions

We discussed above how the three-point functions give rise to the leading order expression for the phase shift $\delta(\vec{b}, s) = sF(\vec{b})$. If this result were exponentiated, as $e^{i\delta}$, then we could get a time advance problem similar to what we found for the shock waves. Here we would like to explain how to get a time advance problem without using the particular non-linear structure of shock waves. The goal is to present the problem in a way that depends only on very general principles.

First note that in order for time delay to be a problem we would like to find that the time delay $\Delta v = \partial_{p_{2,v}}\delta$ is larger than the quantum mechanical uncertainty that is implicit

JHEP02(2016)020

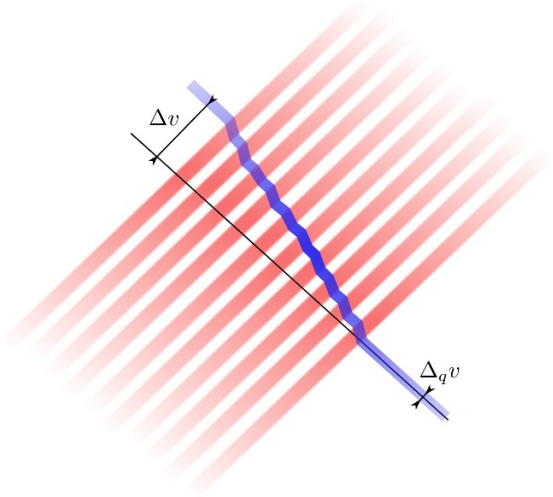


Figure 5. We imagine a particle going through a set of successive scattering events. The intrinsic quantum uncertainty in v is $\Delta_q v$. We have drawn a situation where there is a final time advance after going through all the shocks that is larger than the quantum uncertainty. In this figure we have neglected the delay of the u-localized particles.

in the definition of the wave packet for a particle of momentum $p_{2,v}$. This uncertainty is of the order of $\Delta_q v \sim 1/p_{2,v}$. Thus the figure of merit is

$$\frac{\Delta v}{\Delta_q v} = p_{2,v}\partial_{p_{2,v}}\delta = \delta\,, \tag{3.12}$$

here we used that δ is linear in s, and therefore linear in $p_{2,v}$ ($s \propto p_{1,u}p_{2,v}$). Thus in order to see a problem, we expect that δ should be greater than one. On the other hand, the validity of perturbation theory suggests that δ should be much less than one.

In order to amplify the effect, we imagine that particle number two undergoes N successive instances of particle number one, see figure 5. Through each instance it gets a small phase shift which leads to a factor in the out state of the form $(1 + i\delta)$ with small δ. If we repeat this N times we get

$$(1 + i\delta)^N \sim e^{iN\delta}\,, \qquad \delta \ll 1\,, \quad N \gg 1\,. \tag{3.13}$$

This is the total phase shift, and as explained in (3.12), we want $N\delta$ to be of order one. This can be achieved by taking $N \sim 1/\delta$. In addition, we would like to make sure that the approximations that we used remain valid. In particular, we have said that particle 2 remains localized at some distance b through the whole process. We can choose light-cone coordinates for the evolution of particle two so that u is time, then we have a non-relativistic problem with mass $m \sim p_v$. The spreading of the wavefunction during the time U that the whole process takes is $\Delta b \sim \sqrt{U/p_v}$. The time U that the process takes is determined as follows. We want to separate the N instances of the scattering process from each other so that we can view them as independent and we can use (3.13). The best we can localize each of the N particles of type 1 is by an amount $\Delta_q u \sim 1/p_u$. Thus $U = N/p_u$ and then

JHEP02(2016)020

$\Delta b \sim \sqrt{\frac{N}{s}}$. We want $\Delta b \ll b$. This translates into the condition

$$N \ll s b^2 \quad \longrightarrow \quad \frac{1}{s b^2} \ll \delta \,, \tag{3.14}$$

where we used that $N \sim 1/\delta$ in order to have a problem. We see that the simultaneous validity of (3.14) and (3.13) can be achieved if $sb^2 \gg 1$ and if s can also be increased so as to achieve

$$\frac{1}{s b^2} \ll \delta \ll 1 \,, \tag{3.15}$$

which can be done if δ grows with s. An additional issue is that we wanted to neglect the deflection angle. Then we can replace each of the 1 particles by a pair of particles localized in the transverse dimension at a distance b with respect to particle number 2 on opposite sides of its trajectory, as shown in figure 2.

For spin zero particles, notice that if we had a $g\phi^3$ vertex, then the scalar exchange leads to $\delta \sim \frac{1}{s} \frac{g^2}{b^{D-4}}$. In this case we cannot obey the conditions (3.15) to obtain a possible causality problem. In fact, (3.14) implies that $g^2/b^{D-6} > 1$, which means that we are at strong coupling. Thus we see that a crucial feature that we used is that the phase shift increases as a function of s. In the case that we exchange a spin one field, δ is independent of s and there is also no causality issue. The reason is simple, the spin two field is effectively changing the metric and causal structure while the spin zero or one fields are not.

The conclusion is that we have justified the exponentiation (3.13) and thus the derivation of the time delay problem in a way that does not depend on the details of the shock wave solution. Let us emphasize that the preceding shock wave discussion was motivational, but the thought experiment that we have set up in this section does not require us to rely on the non-linear structure of the shock wave. It was all derived from the on-shell three-point functions plus certain assumptions about the locality of the theory that allowed us to view each shock as an independent event.

In fact, there are some cases where the shock wave computation does not give the right answer. For example, consider the pure gravity case, where $\delta = Gs/b^{D-4}$. If the energy is large enough to form a black hole, then the time delay computed from the shock wave is not the correct description for the physics. We expect to form a black hole when the center-of-mass energy is such that the associated Schwarzschild radius $r_S^{D-3} = \sqrt{s}G$ is larger than b [47–49]. It is easy to check that this never happens if (3.15) is obeyed. It is not obeyed even if we consider the energy of the N particles that we used for the argument (with $N \sim 1/\delta$).

There is still one more complication that we need to deal with in order to make the argument clearer. In the shock wave discussion of section two, the spin of the particle creating the shock did not matter. However, with the modified three-point functions, the spins of the scattered particles can change. The full interaction is a spin dependent force which acts of both spins. It acts on both the spin of the left and the right moving particles (particle one and two, see figure 3). I will be necessary for us to be able to fix the polarizations of particles one and three, see figure 3. This can be achieved by replacing particle one, by a coherent state of particles. In this case, due to the usual Bose enhancement factor, particle three will have a larger probability of remaining with the same

JHEP02(2016)020

polarization. In this set up we can set the spins, or polarization vectors, of particles one and three to be the same. Or, more precisely, $\epsilon_3 = \epsilon_1^*$. Since we are at weak coupling we can consider a coherent state with a mean occupation number which is large enough for us to be able to neglect the spin flips but small enough that the total scattering amplitude is still small. In other words, the use of coherent states allows us to effectively select the final state for particle 3. More explicitly, say that we form a coherent state for the oscillator mode created by $a^\dagger$, $e^{\lambda a^\dagger}|0\rangle$. We could then have terms in the interaction Hamiltonian that leave this oscillator the same $H_{1,int} = h_1 a^\dagger a$ or that mix it with a second oscillator $H_{2,int} = h_2 b^\dagger a + h.c.$. Here h_1 and h_2 can act on other degrees of freedom. Then, for a coherent state, the matrix elements where there is no change are enhanced, due to the usual Bose enhancement factors, relative to the ones where there is a change

$$\begin{aligned} \langle 0|e^{\bar{\lambda} a}\, H_{1,int}\, e^{\lambda a^\dagger}|0\rangle &\propto |\lambda|^2 \langle h_1\rangle\,, \\ \langle 0|e^{\bar{\lambda} a} b\, H_{2,int}\, e^{\lambda a^\dagger}|0\rangle &\propto \lambda \langle h_2\rangle\,. \end{aligned} \tag{3.16}$$

We see that for large λ the first term is enhanced. We have not indicated explicitly the initial and final states on which h_1 and h_2 act, since they involve other oscillators. Since we are at weak coupling, we can choose λ large enough so that the first term dominates relative to the second term, while still the whole process is in the weakly coupled approximation, or the total effect of the interaction Hamiltonian is small. Of course an alternative way to say this is that we are creating a classical background with the particles of type 1 in figure 3 the terms that have a non-zero expectation value in this classical background dominate over the others. We want a classical background with small enough amplitude that we can still trust the leading order perturbative expansion of the interaction Hamiltonian.

The use of coherent states also allows us to select final states for particle 3 in figure 3 with a a small momentum p_v. Namely, we form a coherent state out of a superposition of particles with large momenta $p_{1,\mu}$. We need a superposition since we need to localize this particles within the transverse plane to a location smaller than b. Thus we have some dispersion in the transverse momenta $\vec{q}$. With a large p_u component of the momentum, we then get a small momentum along the $p_v = \frac{\vec{q}^2}{4p_u}$ direction. Since we have a coherent state, the particle 3 is also taken out of this superposition and has the same range of values for the momentum. Therefore the total momentum transfer in the t-channel along the $v$direction is very small. Then the kinematics chosen in (3.2) is representative for the process in question. This still allows a possibly large amount of momentum transfer along the u direction. We will discuss this in subsection 4.3 .

Another minor point, is that the phase shift δ represents a time delay for both particles, it affects both particle 1 and particle 2. So far we have been focusing only on the effects on particle 2. These coherent states also allow us to effectively select a final state for particles 3, so that we can focus more clearly on the time delay with which particle four emerges, see figure 3.

3.4 Scattering of gravitons in $D > 4$ dimensions

It is also easy to argue that α_2 and α_4 structures lead to causality problem in $D > 4$. To show this we consider the probe graviton that scatters off the coherent state. The phase

JHEP02(2016)020

shift takes in this case the following form (see appendix B for details)

$$\delta \sim Gs(e_1^{ij}e_3^{ij} + \alpha_2 e_1^{ij}e_3^{ik}\partial_{b^j}\partial_{b^k} + \alpha_4 e_1^{ij}e_3^{kl}\partial_{b^i}\partial_{b^j}\partial_{b^k}\partial_{b^l}) \times \\ \times(e_2^{ij}e_4^{ij} + \alpha_2 e_2^{ij}e_4^{ik}\partial_{b^j}\partial_{b^k} + \alpha_4 e_2^{ij}e_4^{kl}\partial_{b^i}\partial_{b^j}\partial_{b^k}\partial_{b^l})\frac{1}{|\vec{b}|^{D-4}}\,. \tag{3.17}$$

In order to find problems we will be choosing various polarizations for the particles. For example, let us firt consider particle 1 with polarization $e^1_{xx} = -e^1_{yy}$ and the other components equal to zero. Here x, y represent to two directions in the transverse plane. We call this the $\oplus$ polarization. We also choose $e^3 = e^1$. We enforce this by sending a coherent state with this polarization. Now for particle 2 we can choose the same polarization or the crossed polarization, called $\otimes$, given by $e^2_{xy} = e^2_{yx} = 1/\sqrt{2}$ and all other components equal to zero. Then we find the following. If $\vec{b}$ is along the $\hat{x}$ direction, then these two different polarizations for particle two do not mix as they go through the shock. They diagonalize the phase shift matrix. For small enough b the α_4^2 terms dominate since they are the most singular in the small b expansion. These terms have the form

$$\begin{aligned}
\delta_{\oplus\oplus} &\sim Gs\alpha_4^2 O_\oplus O_\oplus \frac{1}{b^{D-4}} = \frac{Gs\alpha_4^2}{b^{D+4}}(\text{positive})\,, \\
\delta_{\oplus\otimes} &\sim Gs\alpha_4^2 O_\oplus O_\otimes \frac{1}{b^{D-4}} = \frac{Gs\alpha_4^2}{b^{D+4}}(\text{negative})\,, \\
O_\oplus &= (\partial^2_{b_x} - \partial^2_{b_y})^2\,, \qquad O_\otimes = 4\partial^2_{b_x}\partial^2_{b_y}\,,
\end{aligned} \tag{3.18}$$

where we take the derivatives first and then set $\vec{b} = (b_x, 0, \cdots, 0)$. Where the terms in parenthesis are polynomials in D which are positive or negative definite.[9] Notice that the positivity of the first case is due to the following argument. If the polarizations of particle 2 and 4 are the same as those of 1 and 3, then the configuration is constrained by unitarity along the t-channel.[10] Therefore in this case we should get a strictly positive answer for the time delay for the contribution of any particle with a non-zero coupling. Since we obtained a negative time delay for the second case in (3.18) , we conclude that α_4 should be set to zero unless new particles are present.

Once α_4 has been set to zero, we can discuss α_2. In that case we can still choose the $\otimes_{xy}$ polarization for particles 1 and 3 and the $\otimes_{yz}$ polarizations for particles 2 and four. We then focus on the terms proportional to α_2^2 since they are the dominant terms at small b (once we have set $\alpha_4 = 0$). We then get

$$\begin{aligned}
\delta_{\otimes_{xy},\otimes_{yz}} &\sim Gs\alpha_2^2 \hat{O}_{xy}\hat{O}_{yz}\frac{1}{b^{D-4}} = -\frac{Gs\alpha_2^2}{b^D}2(D-4)(D-3)(D-2)\,, \\
\hat{O}_{xy} &= -\partial^2_{b_x} - \partial^2_{b_y}\,, \qquad \hat{O}_{yz} = -\partial^2_{b_y} - \partial^2_{b_z}\,.
\end{aligned} \tag{3.19}$$

Then we conclude that α_2 should also be set to zero unless new massive particles appear.

[9] For $D > 4$, (positive)$= (D-4)(D^2-4)D(336 + 128D + 20D^2 + 4D^3 + D^4)$ and (negative)$= -4(D-4)(D^2-4)D(D+3)(20 + 6D + D^2)$. For $D = 4$ the derivatives act on $-\log b$ and we get (positive) = -(negative) =80640.

[10] t-channel unitarity is the following statement. When the polarizations of 2 and 4 are related by conjugation and reflection along the $\vec{b}$ axis to the polarizations of 1 and 3 respectively, then the residue of the t-channel pole is positive.

JHEP02(2016)020

3.5 Scattering of gravitons in four dimensions

Let us now discuss in more detail the four-dimensional case, $D = 4$, which is a bit special. First, need to take into account that the Einstein term produces a $\log(L/b)$ time delay, where L is an IR cutoff. Second we need to take into account the parity violating structure. The logarithm can be taken into account by modifying the causality criterion in the form suggested by Gao and Wald, who define it by comparing to the behavior of the same metric far away. In this way the $\log L$ term is eliminated and it is easy for a power law behavior produced by α_4, which goes as $1/b^4$ to overwhelm the logarithm. Also we will later repeat the computations for AdS_4 space and we will see that $L \to R_{AdS_4}$. In conclusion, this is not a real issue.

Note that in four dimensions the α_2 structure is identically zero. This is related to the fact that the Gauss-Bonnet term becomes topological in four dimensions. Thus, we have only the Einstein term structure and the α_4 one. With four-dimensional kinematics, we can consider the situation with coherent states of particles of type 1. Let us choose the spin of these particles to be plus. Due to the coherent state considerations, the spin of particle 3 also needs to be plus (in the outgoing notation, or minus in the incoming notation). In other words the amplitude does not have a spin flip. In fact, without a spin flip the α_4 structure does not contribute in four dimensions. Thus in the vertex involving particles one, three, and the intermediate one, the only structure that contributes is the Einstein one. This contribution is effectively the same as the spin zero one. Then we can run an argument similar to the one above.

Let us now discuss the parity violating structure, together with the parity preserving one. By considering the coherent state for particle one (see figure 3), with definite helicity (positive or negative) we get that only the Einstein structure contributes to the $\mathcal{A}^{13I}$ three-point amplitude (see figure 4). The we get the following matrix form for the phase shift for particle two

$$\begin{aligned}\delta &= 2Gs\left(\mathbf{1} \;+\; \gamma\partial_\beta^4|-\rangle\langle+| \;+\; \gamma^*\partial_{\beta^*}^4|+\rangle\langle-|\right)(-\log|\beta|) \\ &= 2Gs\left(-\mathbf{1}\log|\beta| \;+\; \frac{3\gamma}{\beta^4}|-\rangle\langle+| \;+\; \frac{3\gamma^*}{\beta^{*4}}|+\rangle\langle-|\right),\end{aligned} \tag{3.20}$$

where we introduced a complex variable $\beta = b_1 + ib_2$ in the two dimensional transverse space. We also used $\partial_\beta = \frac{1}{2}\left(\partial_{b^1} - i\partial_{b^2}\right)$ and $e_\pm \propto e_x \mp ie_y$. Here γ is the coefficient of the $+++$ amplitude and γ^* the coefficient of the $---$ one. Notice that these physically imply that the particle two undergoes a spin flip. We see that δ is a two by two matrix in the space of helicities. In (3.20) , $\mathbf{1}$ represents the identity matrix in this two dimensional space. The matrix in (3.20) can be diagonalized by choosing the polarization directions

$$p_{1,2} \propto |+\rangle \pm \sqrt{\frac{\gamma\,\beta^{*4}}{\gamma^*\,\beta^4}}|-\rangle\,. \tag{3.21}$$

Then we find that

$$\delta_{1,2} = 2Gs\left(-\log|\beta| \pm 3\frac{|\gamma|}{|\beta|^4}\right). \tag{3.22}$$

Thus we see that we have a causality problem for small enough $b = |\beta|$.

JHEP02(2016)020

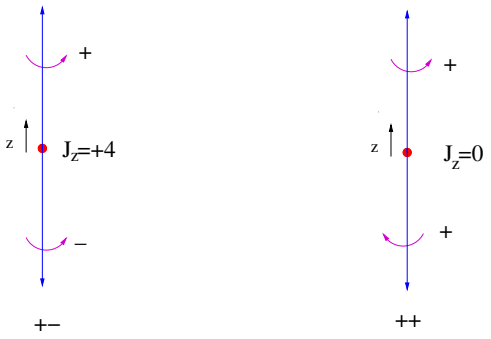


Figure 6. Consider the coupling of a massive spin two particle to two massless gravitons. Let us choose the kinematic configuration so that the massive particle decays into two massless gravitons along the $\hat{z}$ axis. The $+-$ helicity configuration is impossible since the angular momentum along the z axis would be $+4$. The $++$ configuration is allowed.

4 Fixing the causality problem by adding massive particles

Let us discuss how to evade the causality problem that we found above. This problem can be evaded by adding new particles at the scale α_2 or α_4. We will discuss the case of a weakly coupled theory where the problem should be fixed at tree level. This is indeed what happens in string theory, see appendix E. For cases involving loops see appendix C. Let us first consider the corrections to the time delay due to new particles being exchanged in the t-channel. The new particles have to lead to a phase shift growing like s, or a higher power of s. *Thus, we should add massive particles with spin* $J \geq 2$.

We will now argue that massive spin two particles do not help and that we need particles of higher spin. In particular, this will then rule out a solution involving Kaluza-Klein gravity, which would be a special example of the addition of massive spin two particles. For this reason we will analyze it in detail.

4.1 Massive spin two particles do not fix the problem in $D = 4$

Let us first discuss the four-dimensional case. Since the external states are massless spin two particles, the on-shell three-point vertices involve two massless particles and a massive spin two particle.

In four dimensions, we can label the massless particles by their helicities. An important result is that, in all incoming notation, the only non-zero amplitudes involve $++$ or $--$ helicities for the massless particles. In particular the $+-$ combinations are zero. The argument is essentially the same as in the Weinberg Witten theorem [50], or the statement that gravity does not have a local stress tensor operator.[11] Imagine that we have the massive spin two particle in its rest frame. We let it decay into two massless spin two particles. Let us suppose that the two decay products move in opposite directions along the $\hat{z}$ axis, see figure 6. In the $+-$ configuration the total sum of the spins of the decay

[11] The matrix elements of the stress tensor operator between two on-shell graviton states is like the coupling to a massive spin two particle, where the square of the momentum of the stress tensor, q^2 corresponds to the mass of the massive spin two particle.

JHEP02(2016)020

products along the $\hat{z}$ axis is $+4$ or -4. However, the initial massive particle had spin at most ± 2. Therefore a $+-$ configuration is impossible. With a $++$ or $--$ configuration there is not problem because the sum of the spins is zero. One can also write down explicitly the corresponding three-point amplitude

$$\begin{aligned}&\widetilde{\alpha}_4 \epsilon^I_{\mu\nu} p_1^\nu p_3^\mu \left[(\epsilon_1.p_3)(\epsilon_3.p_1) - (\epsilon_1.\epsilon_3)(p_1.p_3)\right]^2 \\ &\qquad \to 2\widetilde{\alpha}_4 p_u^2 \left[(\vec{e}_1.\vec{q})(\vec{e}_3.\vec{q}) + \frac{m_I^2}{2}(\vec{e}_1.\vec{e}_3)\right]^2 , \end{aligned} \tag{4.1}$$

where $\epsilon^1_{\mu\nu} = \epsilon^1_\mu \epsilon^1_\nu$, and we used that the component of $\epsilon^{I\mu\nu}$ that contributes the largest factor of s in the sum over intermediate states is $\epsilon^{Iuu} = 2$. We have denoted the coupling by $\widetilde{\alpha}_4$ since it reduces to the α_4 structure in the massless limit. Here 1 and 3 are the massless particles. Of course, $p_1.p_3$ is given by the mass of the massive particle. We see that this result is invariant under $\epsilon^1 \to \epsilon^1 + p^1$, and so on. In the second line of (4.1) we have written the three-point amplitude including the leading terms in the high energy limit. This is written in terms of the purely transverse polarization vectors (or tensors) introduced in (3.4). In $D = 4$ there is also a parity violating structure which we will not need to write explicitly.

If we now consider particles 1 and 3 in figure 3 to be associated to a coherent state with definite spin, then we have no spin flip allowed and this coherent state does not couple to the massive spin two particles. Therefore in four dimensions the massive spin two particles cannot solve the problem, they simply do not couple to the type of source that we are considering. Note that it is important that the massless intermediate gravitons are still coupling to the 1-3 coherent state through the Einstein three-point function, and as discussed in section 3.5, it leads to a causality problem for particle two in figure 3.

We can further show that the massive spin two particle with a coupling (4.1) by itself also leads to a causality problem and should therefore not be present. In fact, it will be useful for our later argument to understand this in more detail. For simplicity let us set to zero the parity violating massive structure. For the coherent state that involves particles one and three in figure 3 we choose the $\oplus$ polarization with $e^1_{xx} = -e^1_{yy}$ and the other components equal to zero, and the same for e^3. Here x, y represent the two directions in the transverse plane. Now for particle 2 we can choose the same polarization or the crossed polarization, called $\otimes$, given by $e^2_{xy} = e^2_{yx} = 1/\sqrt{2}$ and all other components equal to zero. Then we find the following. If $\vec{b}$ is along the $\hat{x}$ direction, then these two different polarizations for particle two do not mix as they go through the shock. They diagonalize the phase shift matrix. If the polarizations of particle 2 and 4 are the conjugate to those of 1 and 3, and reflected along $\vec{b}$, then the configuration is constrained by unitarity along the t-channel to give a strictly positive answer for the contribution to the time delay of any particle with a non-zero coupling. On the other hand, if we average over all polarizations for particle 2, it is possible to see that the terms involving α_4 or $\widetilde{\alpha}_4$ (the massive particle contributions) all vanish. Thus, the contribution from the crossed polarization has to have the opposite sign. In other words, unitarity fixes a plus sign for the time delay for one polarization and this implies a negative sign for the other. Indeed, it is possible to see this

JHEP02(2016)020

explicitly by computing the massive particle contribution to both answers, which are

$$\begin{aligned}\delta_{\oplus,\oplus} &= 4Gs\Big(\sum_m \widetilde{\alpha}_4^2 O_m O_m\Big) K_0(mb)\,,\\ \delta_{\oplus,\otimes} &= -4Gs\Big(\sum_m \widetilde{\alpha}_4^2 O_m O_m\Big) K_0(mb)\,,\\ O_m &\equiv \partial_{b_x}^2\partial_{b_y}^2 - \frac{m^4}{8}\,,\end{aligned} \tag{4.2}$$

where the first subindex of δ is the polarization of particles 1 and 3 and the second that of particles 2 and 4 in figure 3. By acting with this operator explicitly one can see that it gives a negative answer in the second case. This is independent of the sign of $\widetilde{\alpha}_4$. In fact, it is also negative for the contribution of the massless case when we have the α_4 structure on both sides. The full phase shift also has the General Relativity contribution. Once we have a single massive particle, it is possible to go to a small enough b so that we overwhelm the positive contributions from the General Relativity vertices.

This shows that in a theory with up to spin two particles we cannot solve the causality problem that arises when α_4 is nonzero. In addition, we see any massive spin two particles, even if present, they should have $\widetilde{\alpha}_4 = 0$ in order not to cause further causality problems.

4.2 Massive spin two particles do not fix the problem in $D > 4$

Now we now move on to a higher dimensional gravity theory, $D > 4$. The three-point amplitudes for two gravitons and a massive spin two particle now have two possible structures, first the one in (4.1), which can be multiplied by a coefficient that we will still call $\widetilde{\alpha}_4$. And a second one of the form

$$\begin{aligned}&\widetilde{\alpha}_2\epsilon^I_{\mu\nu}\left[\epsilon_1^\mu p_3^\nu(\epsilon_3.p_1) + \epsilon_3^\mu p_1^\nu(\epsilon_1.p_3) - p_1^\mu p_3^\nu(\epsilon_1.\epsilon_3) - \epsilon_1^\mu\epsilon_3^\nu(p_1.p_3)\right]\times\\ &\qquad\times\left[(\epsilon_1.p_3)(\epsilon_3.p_1) - (\epsilon_1.\epsilon_3)(p_1.p_3)\right]\\ &\to 2\widetilde{\alpha}_2 p_u^2\left[e^1_{ki}q^i e^3_{kj}q^j + \frac{m^2}{2}e^1_{ij}e^3_{ij}\right].\end{aligned} \tag{4.3}$$

where again $\epsilon^I_{\mu\nu}$ is the intermediate state polarization vector and we used that we only care about its $\epsilon^{Iuu} = 2$ component. We have introduced a new coefficient $\widetilde{\alpha}_2$. In the second line we have indicated the form that it takes in the high energy limit. In the last line the polarization tensors are purely in the transverse directions and q is the momentum transfer.

We can first consider a setup with four-dimensional kinematics. Namely, we can consider particles 1 and 3 to be associated to a coherent state which is uniformly distributed along $D-4$ of the original dimensions. In this case the problem is essentially four-dimensional and the three-point amplitudes involving α_2 and $\widetilde{\alpha}_2$ (both massless and massive) do not contribute. If we want to avoid causality problems, and without spin > 2 particles, we conclude that both α_4 and $\widetilde{\alpha}_4$ should be zero. The argument is the same as the one we presented in the four-dimensional discussion. Note that since we are getting to four dimensions by effectively dimensionally reducing the higher dimensional theory, then the parity violating four-dimensional structure does not arise.

JHEP02(2016)020

We would now like to rule out the possibility of having contributions with non-zero α_2. For this we will assume that α_4 and $\widetilde{\alpha}_4$ are zero, as shown by the previous argument.

Let us first consider the case $D = 5$. Now we have three transverse directions, let us call them $x,\ y,\ z$. We choose the polarizations of 1 and 3 to be of the $\oplus_{xy}$ type. Inserting this into (4.3) we see that this produces a factor of $O_{xy} = m^2 - \partial_{b_x}^2 - \partial_{b_y}^2$ acting on the massive propagator, which is simply $\frac{1}{b}e^{-mb}$. For the particles 2 and 4 we choose the polarization $\otimes_{yz}$, which also produces a similar operator O_{yz}. The final result has the form

$$\begin{aligned}\delta_{\oplus_{xy},\otimes_{yz}} &\sim Gs\left(\sum_m(\widetilde{\alpha}_2)^2 O_{xy}O_{yz}\frac{1}{b}e^{-mb}\right)\\ &= -Gs\sum_m(\widetilde{\alpha}_2)^2\frac{(b^3m^3+5b^2m^2+12bm+12)}{b^5}e^{-mb}\,,\end{aligned} \tag{4.4}$$

where we have set $\vec{b} = (b, 0, 0)$ after taking the derivatives. We see that for any sign of $\widetilde{\alpha}_2$ this produces a negative result. Furthermore, in the massless case, $m = 0$, this also gives the part of the graviton contribution proportional to α_2^2. The graviton also contains other contributions involving the ordinary Einstein piece on both three-point functions, as well as a mixed term. These contributions behave like $1/b$ and α_2/b^3 respectively. Thus, for small b, the graviton contribution involving α_2^2 dominates, since it goes as $\alpha_2^2 b^{-5}$. Therefore, we conclude that if α_2 is non-zero, then by going to small enough b we get a causality problem. Furthermore, this problem cannot be fixed by adding massive spin two particles.

For $D > 5$ one can run a similar argument. But of course, we could also set up the problem with five-dimensional kinematics. In other words, we choose a coherent state spread over $D - 5$ of the dimensions and we get the same as what we discussed above.

The final conclusion is that if we have extra structures in the graviton three-point function (if α_2 or α_4 are nonzero), they lead to a causality problem *which cannot be fixed by adding massive particles with spins* $J \leq 2$.

With a risk of being repetitive, let us summarize the argument that rules out massive spin two particles. First we go to four-dimensional kinematics where the massless or massive couplings proportional to α_2 or $\widetilde{\alpha}_2$ do not contribute and we rule out both α_4 and the similar coupling $\widetilde{\alpha}_4$ to massive intermediate gravitons. Then we go to five (or higher) dimensional kinematics and we rule out both α_2 and $\widetilde{\alpha}_2$. In particular, notice that, even in the case of ordinary Einstein gravity, with $\alpha_2 = \alpha_4 = 0$, we have ruled out tree level couplings to massive gravitons (or massive spin two particles).

4.3 Exciting the graviton into new particles

In the above discussion we ignored the possibility of exciting a graviton when it passes through the shock and transforming it into a new state.[12] In this section we discuss this possibility and conclude that it cannot fix the problem.

Since the available energy is large, compared to $1/b$, it is possible to turn the incoming graviton into an outgoing massive particle, let us call it X. If we use coherent states for

[12] In string theory these are called tidal excitations of the string.

JHEP02(2016)020

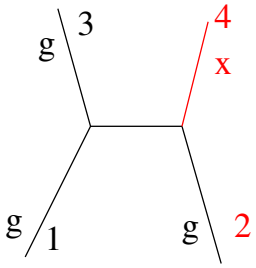


Figure 7. When the particles scatter, the graviton can become another massive particle, here labeled by X.

particles 1 and 3, then we suppress the processes where particle 3 becomes a new massive particle and we only have to worry about the possibility of particle 2 turning into this new massive particle; see figure 7. This can happen even if the mass of the new particle is much larger than b^{-1}, but smaller than $\sqrt{s}$. The reason is that there can be some p_u energy transfer from particles 1 and 3 to particle 4 in figure 3.

Let us view the process of the graviton 2 passing through the shock as a signal transmission problem. Focusing on the v dependence of the signal we can say that the out-signal, given an in-signal must be causal. Namely, if the in-signal vanishes for $v < 0$, then the out-signal must vanish for $v < 0$. In Fourier space these signals are related by $f_{\rm out}(\omega) = S(\omega) f_{\rm in}(\omega)$, with $\omega = -p_v$. Here we are using that particles 1 and 3 carry a negligible amount of p_v, since we are using $v-$translation invariance. Of course, if we have physical particles, we cannot localize them sharply in v because they only have positive frequencies. In order to obtain a sharp causality bound we need to invoke the vanishing of the field commutators, $[\phi_{\rm out}(v), \partial_v \phi_{\rm in}(v')] = 0$ for $v < v'$ (we put the derivative to remove possible zero mode issues).[13] As reviewed in appendix D , causality implies that $S(\omega)$ is analytic in the upper half plane. In addition, the fact that we can produce other particles can only make the strength of the graviton signal in the future smaller. This, in turn, implies that the S matrix element for graviton going into graviton, call it $S_{gg}(\omega)$ should be smaller than one in the upper half plane,

$$|S_{gg}(\omega)| \leq 1\,, \qquad \text{for} \qquad \text{Im}(\omega) > 0\,. \tag{4.5}$$

See appendix D for further discussion.

In situations where we have some time advance for the graviton, we are getting an infinitesimal matrix element of the form

$$S_{gg} = 1 - i\Delta v p_v = 1 + i\Delta v \omega\,, \tag{4.6}$$

with $\Delta v < 0$. Then if we set $\omega \to i\gamma$, with $\gamma > 0$ we get $S_{gg} = 1 - \gamma \Delta v$ which is bigger than one in the upper half plane. Note that we do not need to go to very large values of γ to

[13] In a theory of gravity we do not have local field operators. However, we can imagine defining such operators in the asymptotic past and future. More precisely, in order to run into the causality problems we need to put them far enough to the past and future that we can neglect the change in the spacetime metric but close enough so that the quantum mechanically dictated momentum p_u of the signal particles does not wash out the possible time advance. Here we will assume that it is possible to do this.

JHEP02(2016)020

obtain a violation, we only need p_v or γ, to be large enough so that this impact parameter description is good enough.

In this presentation of the argument, it is clear that adding extra particles as possible extra final states does not help. We need to modify S_{gg}. In other words, the transformation of the graviton to X is irrelevant for this argument because we are considering the graviton-graviton S matrix element. The transformation to X and then back to the graviton can contribute to this matrix element at higher orders in G. But, this cannot fix the problem we run into (4.6), which is of first order in G.

4.4 Massive higher spin particles can solve the problem

Now we consider the exchange of a massive spin $J > 2$ particle in the t-channel. Its contribution to the phase shift will rise with energy like Gs^{J-1} and at high energies it will dominate over the graviton contribution. This can happen even in the regime that the theory is weakly coupled. If we have a single contribution of this type, we also run into a problem. The problem is the following. We can think of the propagation of the particle number 2 as signal transmission problem where time is v. In other words, we start with a signal $f_{\rm in}(v)$ which vanishes for $v < 0$, then the out-signal $f_{\rm out}(v)$ should be zero for $v < 0$. In addition we want that the total norm of the out wave should not grow. From these two conditions we can deduce that the $S(\omega)$ matrix as a function of the "energy", $\omega = -p_v$, should be analytic in the upper half complex ω-plane and, in addition, it should be bounded $|S(\omega)| \leq 1$ in the upper half plane. See appendix D for a review of these properties. However, a particle of spin $J > 2$ leads to a contribution $S \propto 1 + iGs^{J-1} + \cdots$ which becomes bigger than one in some regions of the upper half complex s plane. Notice that the problem arises already at weak coupling, for a small value for Gs^{J-1}.

Thus, a *finite* number of higher spin particles does not fix the problem. In fact, it generates problems of its own. On the other hand, an *infinite* number of particles with higher spin can solve the problem. An example is string theory. This problem has been discussed extensively in the classic papers by Amati, Ciafaloni and Veneziano [51–55] (see also [56]). In fact, the amplitude Reggeizes

$$\mathcal{A} \sim \frac{Gs^2}{t} s^{\frac{\alpha' }{2} t} e^{-i\pi \frac{\alpha'}{4} t} \,, \tag{4.7}$$

for large s and small t ($s\alpha' \gg 1$, $t\alpha' \ll 1$). This expression has a cut in the s-channel, due to the creation of physical states along the s-channel. These are simply the massive closed string states that are present along the s-channel. For spacelike t, $t < 0$, we see that this effectively leads to a phase shift that decreases faster than s at large s. Taking (4.7) and transforming to the impact parameter representation we find that for $b^2 \ll \alpha' \log s$ we get a behavior [53]

$$\delta \sim \text{Pol} \frac{Gs}{(\log(\alpha' s/4))^{\frac{D-4}{2}}} \left(\frac{1}{D-4} - \frac{b^2}{2(D-2)\alpha' \log(\alpha' s/4)} + \cdots + i \frac{\pi}{4 \log(\alpha' s/4)} + \cdots \right) , \tag{4.8}$$

JHEP02(2016)020

where Pol $= 1 + \alpha_2 \epsilon.\partial_b \epsilon.\partial_b + \cdots$ is the part coming from the polarization tensors, which includes the new structures in the three-point functions.[14] This is indeed compatible with causality. We also get a large imaginary part that is reflecting the fact that we are creating strings along the s-channel. Notice that we had argued before that in a local theory we expect that by going to impact parameter space we can suppress tree level s-channel processes. This is not true in string theory, which contains extended objects. Furthermore, since their size increases with mass logarithmically, we see that at high energies their effects appear at $b^2 \sim \alpha' \log(s\alpha')$ rather than the more naive expectation of $b^2 \sim \alpha'$. This justifies the small t expansion used in (4.7). Further aspects of the string case are discussed in appendix E.

The conclusion is that an infinite number of higher spin particles can solve the problem. We need a tower of particles with increasing spins and intricate relations between them so that the expansion can be resummed into an amplitude that does not have a problem. Besides string theory, we do not know if there are other ways of doing this.

4.5 Compositeness and the extra structures for graviton scattering

In this subsection, with the string theory case as an inspiration, we make some remarks about the extra structures for the graviton scattering.

Imagine that the graviton has a composite structure.[15] Let us imagine that the graviton is given by a pair of particles which are in a bound state given by a wavefunction $\psi(r)$ where r is the relative distance. We further assume that these particles scatter through the shock via the usual General Relativity three-point functions. However, since the two particles feel slightly different forces we find that the total scattering amplitude will have the form

$$\delta_0(b_0) + \partial_i \partial_j \delta_0(b) \langle \psi_\epsilon | r_i r_j | \psi_\epsilon \rangle + \cdots \tag{4.9}$$

where δ_0 is the general relativistic expression (2.11) , and the subindex ϵ indicates the spin of the graviton. Notice that since the Laplacian $\nabla_b^2 \delta_0 = 0$, the only terms that can contribute to the second part are those which are not rotationally invariant in the relative coordinates. These are possible because the graviton spin or polarization ϵ. Notice that this is a simple argument for the presence of extra structures in the graviton three-point function. Even though we motivated this with a graviton composed with two particles, the same final formula works if the graviton is made out of many more elementary constituents as it happens in string theory, when it is a string. In any case, the size of the new structure, α_2, due to compositeness, is of order $\alpha_2 \sim r_s^2$ where r_s is the typical size of the graviton. Given that the bound state has this typical size, then we also expect that it can be excited to other states with masses $m^2 \sim 1/r_s^2$. Indeed, by imposing the causality constraint, we found that there should be new particles with masses of this order of magnitude.

[14] For type II superstrings Pol= 1.

[15] Note that we cannot make the graviton as a zero energy bound state in a local relativistic quantum field theory that contains a stress tensor operator [50].

JHEP02(2016)020

5 Anti-de Sitter discussion

The case of asymptotically AdS space is very similar to the asymptotically flat space one. The causal structure is defined by the causal structure of the boundary. We then require that signals that go through the bulk cannot go faster than signals that remain on the boundary. As argued in [26, 57], General Relativity with the null energy condition implies that this is obeyed.

In terms of the dual CFT, this is just the statement that CFT observers cannot exchange information faster than light. Equivalently boundary CFT operators commute outside the boundary light-cone.

5.1 Motivation: the emergence of bulk locality should happen in the classical theory

In this subsection we discuss in more detail the AdS considerations that motivated the present paper.

We expect that the dual of a large N gauge theory should be a weakly coupled string theory with coupling $g_s \sim 1/N$. This should be true both at weak and strong 't Hooft coupling. As we increase the 't Hooft coupling we are supposed to interpolate between a Vasiliev-like theory and an ordinary Einstein-like theory at strong 't Hooft coupling. This whole interpolation happens within the classical string theory in the bulk. Of course, in the ordinary Einstein description we see a local theory in the bulk. Thus the emergence of bulk locality is something that should be contained within *classical* string theory. It is for this reason that it is interesting to understand the constraints of tree level interactions of gravitons and the link between the masses of the higher spin particles and the size of the corrections to Einstein gravity. Here we attempted to link them via the causality considerations for the simplest gravitational interactions. Given the interest of the AdS case, we will discuss in more detail some of its features.

5.2 Statement of AdS causality

In asymptotically AdS gravitational theories the causal structure given by the Minkowski light-cone on the boundary of AdS. This allows us to formulate the causality criterion in a very simple manner. None of the subtleties that existed in flat space regarding the definition of causality appear for gravitational theories in AdS. The basic condition we would like to impose is that

$$\langle \Psi | [T_{\mu\nu}(y), T_{\rho\sigma}(0)] | \Psi \rangle = 0, \quad y^2 > 0\,, \tag{5.1}$$

where Ψ is some nontrivial state in the theory. We will be interested in the commutator computed on the shock wave background [58–62]. The Aichelburg-Sexl shockwave in AdS can be created by inserting a pair of operators creating a coherent state bulk wavefunctions that localize the bulk stress tensor on the light ray [58, 63, 64].

As before, instead of computing the commutator, we study propagation of an energetic graviton through the bulk and impose positivity of the time delay. The whole discussion is similar to the flat space one. The only difference is that the t-channel propagator is now

JHEP02(2016)020

in AdS. Instead to computing it directly, we consider plane wave solutions which encode intermediate state gravitons with the properties we need. Plane wave solutions in AdS_D have the form

$$ds^2 = \frac{-dudv + h(u, y_i, z)\, du^2 + \sum_{i=1}^{D-3} dy_i^2 + dz^2}{z^2}\,. \tag{5.2}$$

Here the function h is only constrained by the Laplace equation in the transverse hyperbolic space

$$z^{D-2}\partial_z(z^{-(D-2)}\partial_z h) + \partial_{y_i}^2 h = 0\,, \tag{5.3}$$

which can be equivalently written as

$$\nabla_{D-2}^2 f - (D-2)f = 0\,, \tag{5.4}$$

where $f = \frac{h}{z}$. In appendix F we give an argument that this is a solution, to all orders in the derivative expansion.

We can also have a plane wave with a delta function source so that instead of (5.3) we write

$$z^{D-2}\partial_z(z^{-(D-2)}\partial_z h) + \partial_{y_i}^2 h = -16\pi G|P_u|\delta(u) z_0^{D-2}\delta^{D-3}(\vec{y} - \vec{y}_0)\delta(z - z_0)\,. \tag{5.5}$$

where the r.h.s. corresponds to an insertion of a delta-function source in the hyperbolic space in (5.4). The Green's function in the hyperbolic space is well-known, so that we get

$$\begin{aligned} h &= \frac{z\varpi(\rho)}{1-\rho^2}\delta(u) \\ &= 16\pi G|P_u|\delta(u)\frac{z(4\pi)^{\frac{2-D}{2}}\Gamma(\frac{D}{2})}{(D-1)(D-2)}\left(\frac{\rho^2}{1-\rho^2}\right)^{2-D} {}_2F_1\left(D-2, \frac{D}{2}, D; -\frac{1-\rho^2}{\rho^2}\right)\,, \end{aligned} \tag{5.6}$$

where $\vec{y}_0$ and z_0 are the coordinates of the source in the bulk, and

$$\rho = \sqrt{\frac{(z-z_0)^2 + |\vec{y} - \vec{y}_0|^2}{(z+z_0)^2 + |\vec{y} - \vec{y}_0|^2}}\,. \tag{5.7}$$

It can be checked that P_u is also the total momentum from the boundary point of view by integrating the boundary stress tensor (read off from the small z expansion of this metric) on the boundary. For example, in AdS_4 and AdS_5 (5.6) takes the following form

$$\begin{aligned} \varpi_4(\rho)/(G|P_u|) &= -8\left(1 - \rho^2 + (1+\rho^2)\log\rho\right)\,, \\ \varpi_5(\rho)/(G|P_u|) &= 2\frac{(1-\rho)^4}{\rho}\,. \end{aligned} \tag{5.8}$$

More generally, we have

$$\begin{aligned} \lim_{\rho\to 0}\varpi(\rho) &\sim \frac{1}{\rho^{D-4}}\,, \\ \lim_{\rho\to 1}\varpi(\rho) &\sim (1-\rho)^{D-1}\,. \end{aligned} \tag{5.9}$$

Let us understand better the symmetries of the problem. The shock wave is localized around $u = 0$ and is probed by a particle which is localized in v. The role of the transverse

JHEP02(2016)020

plane in flat space is played here by the transverse H_{D-2}. It is convenient to think of the probe crossing this hyperbolic space at the center.[16]

The shock centered at $(\vec{y}_0, z_0)$ has a number of Killing vectors that depend on $f(u)$. For arbitrary $f(u)$ the background has an obvious $\mathrm{SO}(D-3)$ symmetry that rotates $\vec{y}$ and a translation in v. For $f(u) = \delta(u)$ the geometry has extra Killing vectors which enhance the rotational symmetry to $\mathrm{SO}(D-2)$, as we had in flat space. In the original coordinates (5.2) , some of the extra symmetries involve special conformal generators. Since we are working in the high energy limit, we effectively have a delta function in u, so that we also expect to have this extra $\mathrm{SO}(D-2)$ symmetry. See [61] for a coordinate system that makes this manifest.

General properties of the AdS shock wave and its different limits are considered in appendix F. In particular, when the center of the shock goes to the boundary $z_0 \to 0$ the problem becomes very similar to the one arising in the computation of energy correlators [14], whereas in the limit $z_0 \to \infty$ it reduces to the setup used in [62] to study causality. Our formulas will reduce to the ones considered before in those limits.

5.3 The effect of higher derivative interactions on particles with spin

Now we would like to consider different type of probes and compute the time delay for them. We start with a simple example of a scalar probe and then move to the case of particles of spin one and two.

If we consider a minimally coupled scalar in the shock wave background its equation of motion takes the form

$$\nabla^2 \phi = 0\,. \tag{5.10}$$

In our setup we are interested in corrections to this equation which are second order in derivatives.[17] Considering terms yielding two derivative equations of motion for ϕ we have to consider terms like

$$\mathcal{H}^{\mu\nu} \nabla_\mu \nabla_\nu \phi\,, \tag{5.11}$$

where the tensor $\mathcal{H}$ is made from the background metric, Riemann and covariant derivatives. However one can check that there is no two index symmetric tensor that is not vanishing on-shell [61]. Of course, this statement is equivalent to the uniqueness of the scalar-scalar-graviton three-point vertex. In the high energy limit we get similar to flat space

$$\partial_u \partial_v \phi + f(u) h(z, \vec{y}) \partial_v^2 \phi = 0\,, \tag{5.12}$$

which produces the time delay

$$\Delta v = \frac{\varpi(\rho)}{1-\rho^2}\,, \tag{5.13}$$

reproducing the flat space computation for small ρ. We assumed that the perturbation crosses the shock at $z = 1$ and $\vec{y} = 0$.

[16] From the CFT point of view we can create such a probe by acting with the operator of given energy and zero momentum in the AdS Poincare coordinates for which $u = 0$ is the future null infinity as explained in [14]. In [14] terms we are working here in the y-coordinates, while the operator with given momentum is inserted in the x-coordinates. In pure AdS case isometries of H_{D-2} at fixed $u = 0$ correspond to the usual Lorentz symmetry group in the x-coordinates of [14].

[17] Since we need derivatives to bring down large factors of momentum.

JHEP02(2016)020

For the gauge boson we imagine at the level of two derivative the following equation

$$\nabla^\mu F_{\mu\nu} + \mathcal{H}_\nu{}^{\mu\alpha\beta}\nabla_\mu F_{\alpha\beta} = 0\,, \tag{5.14}$$

where $\mathcal{H}$ is built out of the Riemann tensor and its covariant derivatives. Using the properties of the background discussed above (we defer the details to appendix F) one can check that the only term that we can have is a correction analogous to that appearing in the case of flat space

$$\nabla^\mu F_{\mu\nu} - \hat{\alpha}_2\, \check{R}_\nu{}^{\mu\alpha\beta}\nabla_\mu F_{\alpha\beta} = 0\,, \tag{5.15}$$

where $\check{R}$ reduces to the Weyl tensor on-shell (see appendix F for details).

If we compute the time delay using the same action (5.15), considering that each mode corresponds to a different constant polarization, ϵ_i, we get

$$\Delta v = \frac{\varpi(\rho)}{1-\rho^2}\left(1 - \hat{\alpha}_2\left(1-\rho^2\right)^2 \frac{\varpi'(\rho) - \rho\varpi''(\rho)}{4\rho\varpi(\rho)}\left(\frac{\epsilon.n^2}{\epsilon.\epsilon} - \frac{1}{D-2}\right)\right)\,. \tag{5.16}$$

The final result is very similar to the one obtained in flat space. The only different is that the polarization dependent delay is slightly more complicated. The flat space result (2.16) is reproduced by considering $\rho \to 0$ limit, whereas the energy correlator constrained is recovered in the limit $\rho \to 1$.

Similarly, in the case of gravity we are interested in the most general form of the second order equations. We choose to parameterize the equations of motions for perturbations as follows

$$\delta R_{\mu\nu} + \alpha_2\, \check{R}_{(\mu}{}^{\rho\alpha\beta}\delta R_{\nu)\rho\alpha\beta} + \frac{\alpha_4}{2}\,[\nabla_{(\mu}\nabla_{\nu)}\check{R}^{\alpha\beta\rho\sigma}]\delta R_{\alpha\beta\rho\sigma} = 0\,, \tag{5.17}$$

where the parameters α_i are in units of the AdS radius R_{AdS} that we set to one. In this case, even though there are several possible ways to contract the indices in each of the above terms, we may concentrate on the contributions to the transverse equations of motion that are the ones yielding the time delay. For that purpose (5.17) is the most general parametrization.[18]

The time delay for these equations of motion is then given by

$$\begin{aligned}
\Delta v &= \frac{\varpi(\rho)}{1-\rho^2}\left[1 + t_2(\rho)\left(\frac{(\epsilon.n)^2}{\epsilon.\epsilon} - \frac{1}{D-2}\right) + t_4(\rho)\left(\frac{(\epsilon.n)^4}{(\epsilon.\epsilon)^2} - \frac{2}{D(D-2)}\right)\right]\,,\\
t_2(\rho) &= \left(1-\rho^2\right)^2 \frac{\varpi' - \rho\varpi''}{4\rho\,\varpi(\rho)}\left(-\alpha_2 + \alpha_4 \frac{D(1+\rho^2)^2 - 2\rho^2}{\rho^2}\right)\,,\\
t_4(\rho) &= -D\alpha_4\left(1-\rho^2\right)^2 \frac{\varpi' - \rho\varpi''}{4\rho\,\varpi(\rho)} \frac{D(1+\rho^2)^2 + 2(1+\rho^4)}{4\rho^2}\,,
\end{aligned} \tag{5.18}$$

where $\vec{n}$ is a vector pointing from the center of the shock to the probe particle, and ϵ is the polarization of the probe particle. These time delays can become negative for small enough ρ if α_2 or α_4 are non-zero.

[18] There is also the possibility of considering more than two derivatives acting on the perturbation as $\nabla^n \delta R$. These contributions in general change the number of degrees of freedom of the theory and will not be considered here.

JHEP02(2016)020

These results can be also reproduced using slightly different method of evaluating the on-shell action in an explicit gravitational theories in the shock wave background like Lovelock or quasi-topological theories (see e.g. [65, 66]).

In the limit $\rho \to 0$ these constraints reproduce the flat space analysis of section B.1 with $\rho = \frac{b}{2}$. In the $\rho \to 1$ limit the above result reproduces constraints discussed in the past. If we take the limit $\rho \to 1$ by taking the shock center to the boundary $z_0 \to 0$ we recover energy correlator computation [14]. If we on the other hand consider $\rho \sim 1$ by taking the shock center to the horizon $z_0 \to \infty$ we recover the shock wave discussed by Hofman [62].

5.4 Implications for a, c in theories with large operator dimensions

Imagine an abstract CFT_4 with large $N \gg 1$ and large gap $\Delta_{\text{gap}} \gg 1$, where Δ_{gap} is the dimension of the lightest higher spin single trace operator. This theory is described by a gravitational theory in the bulk with potentially some higher derivative corrections. String theory inspired intuition suggests that higher derivative corrections should be suppressed by the $\frac{1}{\Delta_{\text{gap}}}$ factor. Our argument shows that this is indeed the case for the simplest corrections, which are the ones affecting the three-point function of stress tensor. Indeed, as we showed in (5.18) , we run into a potential problems with causality at impact parameters $\rho_c \sim (\alpha_2)^{1/2}, (\alpha_4)^{1/4}$. Since this can only be fixed by higher spin particles, we conclude that Δ_{gap} has to be small enough so that it can start correcting the amplitude before we run into this problem.

For $\Delta_{\text{gap}} \gg 1$ then the relevant impact parameters are such that we can approximate the formulas by the flat space limit. Thus, we get the bound of the type

$$(\alpha_2)^{1/2} \lesssim \frac{1}{\Delta_{\text{gap}}}\,, \qquad (\alpha_4)^{1/4} \lesssim \frac{1}{\Delta_{\text{gap}}}\,, \tag{5.19}$$

where $\lesssim$ stands for some numerical coefficient that we cannot fix using our simple analysis.

In the case of $\mathcal{N} = 1$ superconformal theories, $\alpha_4 = 0$ by supersymmetry and $\alpha_2 \propto \frac{a-c}{c}$. Then we get

$$\left|\frac{a-c}{c}\right| \lesssim \frac{1}{\Delta_{\text{gap}}^2}. \tag{5.20}$$

In the case of $\mathcal{N} = 4$ SYM (5.19) is satisfied trivially since $\Delta_{\text{gap}} \sim \lambda^{1/4}$ and $a = c$.

It will be very interesting to find an independent field theoretic argument that leads to the bounds of the type (5.19) or (5.20). It would also be nice to find the precise numerical coefficients in (5.19) and (5.20). In the supersymmetric case, $a - c$ was argued to control asymptotic density of BPS operators in [67]. It would be very interesting to understand if there is any relationship between their work and our analysis.

5.5 Implications for dimensions of double trace operators

This computation of the time delay can be viewed as a special four-point correlation function with particular wave functions for external operators. We can use the OPE to expand the four-point function computation in terms of the time evolution eigenstates (or, equivalently, local operators). One can ask then the following question: what is the relation between the time delay and the OPE data?

JHEP02(2016)020

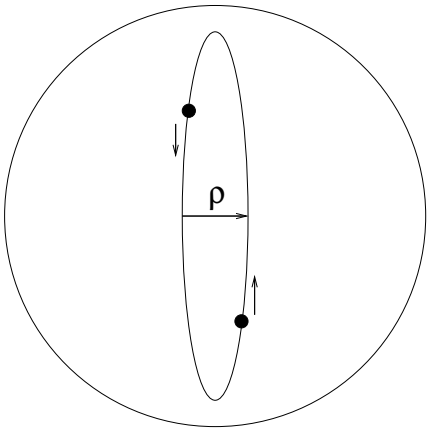


Figure 8. We consider two energetic particles in AdS that oscillate back and forth with energy E and angular momentum J. This models the high twist, high spin double trace operator in the dual CFT.

This question was addressed in a nice series of papers [58, 63, 64] where the time delay was shown to be equal to anomalous dimension of double trace operators of the type $\mathcal{O}(\partial^2)^n\partial_{\mu_1}\dots\partial_{\mu_j}\mathcal{O}$ for large n and large j

$$\begin{aligned}\delta(s,\rho) &= -\pi\gamma(n,j)\,,\\ \rho &= \frac{j}{j+2n}\,,\\ s &= 4n(j+n)\,.\end{aligned} \tag{5.21}$$

where the relations between parameters on both sides of the equation are reviewed below. Intuitively, (5.21) follows from the fact that the phase shift of the correlator $e^{i\delta}$ is given by $e^{-i(\Delta_*-2\Delta_\mathcal{O})\Delta\tau}$ where Δ_* is the dimension of the operator that dominates the OPE, and $\Delta\tau$ is the global AdS time that passed from the beginning to the end of the process. The time delay that we computed corresponds to particles starting at the boundary, getting close to each other at the center and then reaching the boundary again, see figure 8. It takes $\Delta\tau=\pi$ for this process to occur. From this fact (5.21) follows. Even though (5.21) was derived in General Relativity in the limit when $\delta\ll 1$ it follows simply from the AdS graviton diagram exchanged. In the impact parameter representation, only the on-shell t-channel exchange diagram constributes. This diagram is fixed in terms of three-point function $\langle\mathcal{O}\mathcal{O}T_{\mu\nu}\rangle$ in a generic gravitational theory with generic three-point couplings similar to our flat space analysis.

Let us briefly review the results in [58, 63, 64]. The basic idea is the following: the state created by $\mathcal{O}(\partial^2)^n\partial_{\mu_1}\dots\partial_{\mu_j}\mathcal{O}$ for large n and large j can be thought of as two highly energetic particles that follow null geodesics in AdS, see figure 8.

For two geodesics that are characterized by total energy E and spin J the minimal separation is achieved in global coordinates at

$$\rho=\frac{J}{E}\simeq\frac{j}{j+2n}\,. \tag{5.22}$$

JHEP02(2016)020

where we matched energy and spin of the pair of particles to the ones of the double trace operator $J = j$, $E = 2\Delta + 2n + j$ and used that $n, s \gg 1$, with Δ also of order one.

Thus, we see that probing distances much smaller than AdS radius ($\rho \ll 1$) corresponds to considering operators with $n \gg j$. On the other hand $n \ll j$ corresponds to scattering at very large impact parameters. The Mandelstam variable s is given by

$$s = E^2 - J^2 \simeq 4n(j+n)\,, \tag{5.23}$$

and the relation to the anomalous dimensions is that

$$\pi\gamma(n,j) = p_v \Delta v = -Gs\frac{\varpi(\rho)}{1-\rho^2}\,. \tag{5.24}$$

Let us consider different limits of this formula. First, consider very large impact parameters $\rho \sim 1$ or $j \gg n$. In this limit we get

$$\gamma(n,j) \sim -G\frac{n^{D-1}}{j^{D-3}}\,, \tag{5.25}$$

which is in agreement with the general results derived using the crossing equation [68–70]. In the opposite limitof small impact parameters scattering, $n \gg j$ and $\frac{j}{n} > \frac{1}{\Delta_{\rm gap}}$, we have

$$\gamma(n,j) \sim -Gn^2\left(\frac{n}{j}\right)^{D-4}\,. \tag{5.26}$$

We have several comments to add to this story. First, these results should be universal and applicable to generic CFTs with large N and large gap. To write the answer in an abstract form we have to use the relation of G to the two-point function of stress tensors which is well-known and is roughly $G \sim \frac{1}{c_T}$. It means that it should be possible to derive them using crossing equation which still be dominated by the stress tensor exchange. Probably the relevant limit is $z \to 0$, $\bar{z} \to 1$ with $\frac{z}{1-\bar{z}}$ being fixed. It would be nice to reproduce the formulas above using crossing equations.

Second, we see that causality, or positivity of the time delay, implies the constraint $\gamma(n,s) < 0$, which generalizes the ones that were previously known [69–71] for asymptotically large $s \gg n$. Again it would be very interesting to understand how to prove these constraints purely from the field theory point of view.

Third, we see that considering double trace operators of the type $T_{\mu\nu}(\partial^2)^n \partial_{\mu_1} \dots \partial_{\mu_j} T_{\rho\sigma}$ we get new structures due to the dependence on polarization which potentially lead to causality violations and bounds (5.19), (5.20). Of course, the same is true about the double trace operators that involve the conserved current J_μ. It will be very interesting to understand them from the purely CFT viewpoint. Note that in the scattering process the polarizations of particles 3 and 4 in figure 3 can change relative to those of particles 2 and 4, so that the phase shift is an operator that acts on this space of polarization tensors. Both t-channel unitarity, as well as our considerations, constrain only some matrix elements of this general matrix. While we leave the general case for the future, we note that, in some cases, we can ensure that the polarizations of 3 and 4 do not change by using conservation

JHEP02(2016)020

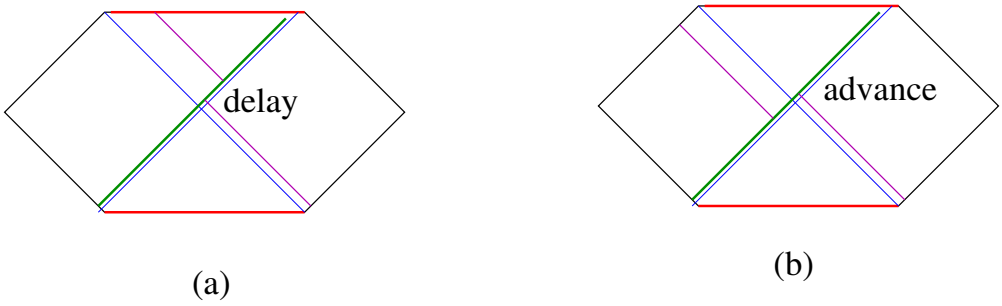


Figure 9. We consider a Lorentzian wormhole configuration that is described, near each wormhole, by the maximally extended Schwarzschild solution. We send a (green) particle from the left very close to the past horizon. We then send a (purple) particle from the right. (a) If this particle gets a time delay, it will fall into the singularity. (b) If the particle gets a time advance, then it can make it out of the other black hole and we would have a way of sending signals through the wormhole. Here the blue lines represent the average position of the horizon of the black hole, defined by null lines that are very far away from the two particles we send in. We assume that the impact parameter is much smaller than the Schwarzschild radius.

of angular momentum along the directions orthogonal to the impact parameter direction. In these cases our considerations apply and we can say that the anomalous dimensions of the corresponding double trace operators should be negative. The positivity statement applies to the part of the phase shift that grows with the Mandelstam invariant s, which translate into the growth with n via (5.23).

However, in our case, this positivity requirement is not obvious from the CFT point of view. It would be nice to see whether this is a general requirement or is one that is present only in theories with a local bulk dual.

6 Wormholes and time advances

General relativity has Lorentzian wormhole solutions that join far away points by short Einstein-Rosen bridges. The simplest configuration is the maximally extended Schwarzschild solution interpreted as an approximation to the metric of two distant black holes which share a single interior. As discussed in [28], these solutions do not lead to a violation of causality in the ambient space because it is not possible to send signals through the wormhole [72].

The inability to send signals through the wormhole depends crucially on the fact that we have a Shapiro time *delay* as opposed to a time advance. For example, if one sends a fast moving particle from the left side, then a particle send from the right will suffer a time delay that will make it go into the singularity, see e.g. [33, 73]. However, if that particle were to suffer a time advance, as opposed to a time delay, then it would be able to go through the wormhole and we would have a violation of causality, see figure 9. Note that we can make the wormhole big enough that we can neglect the higher derivative corrections in the description of the background metric. It can also be big enough that we can neglect the backreaction of the two particles. So we are considering a situation where the impact

JHEP02(2016)020

parameter $b \ll$(Schwarzschild radius). In the $D = 4$ case, the Schwarzschild radius acts at the IR cutoff of the logarithm.

This impossibility of sending signals is crucial for interpreting the wormhole as an EPR state of two disconnected systems [29–31].

7 Cosmological applications

The gravity wave non-gaussianities produced by inflation are a direct measure of the graviton three-point vertex during inflation [17, 20]. To leading order in the slow roll approximation we can do the computation in de Sitter space. The symmetries of de Sitter imply that only two different parity preserving structures are possible. These correspond to the two parity preserving structures that we have in four-dimensional flat space. One is the one produced by the Einstein action and the other can be produced by a term in the action of the form $M_p^2 \alpha_4 R^3$, where R is a Riemann tensor (not the Ricci tensor). The relative size between the two types of gravity wave non-gaussianity is proportional to

$$\frac{\langle hhh \rangle_{R^3}}{\langle hhh \rangle_{\text{Einstein}}} \propto \alpha_4 H^4 \,, \tag{7.1}$$

where H is the Hubble scale during inflation. Of course, both are small compared to the two point function, $\frac{\langle 3\text{point} \rangle}{\langle 2\text{point} \rangle^{3/2}} \sim \frac{H}{M_p}$. See [20] for the explicit expressions.

Thus, if the gravity wave three-point function was measured *and it was found that this exotic new structure is present at a level comparable to the Einstein one*, then one concludes that α_4 is of the order of the Hubble scale. The considerations in this paper imply that there should also be new particles with spins $J > 2$ with masses comparable to the Hubble scale during inflation. Thus, this would be an indirect evidence for string theory during inflation.

Note that $\alpha_4 H^4 \sim 1$ implies that supersymmetry had to be broken at the Planck scale and not at a lower scale, since the $+++$ and $---$ structures are forbidden by supersymmetry.[19] Let us be a bit more explicit about this point. If the short distance theory is supersymmetric, then the field theory Lagrangian does not contain the couplings giving rise to α_4. Now, since supersymmetry is broken, this three-point function could arise from integrating out massive particles. These are expected to contribute to α_4 as

$$\alpha_4 \sim \frac{1}{M_p^2} \left(\frac{1}{m_B^2} - \frac{1}{m_F^2} \right) , \quad \longrightarrow \quad \alpha_4 H^4 \sim \frac{H^2}{M_p^2} \ll 1 \,, \tag{7.2}$$

which is very small. Where, to maximize the effect, we assumed that the masses of the bosons and fermions, as well as their differences, are of order H. We then see from (7.1) and (7.2) that in this supersymmetric scenario the contributions are very small. Thus, in order for the right hand side of (7.1) to be of order unity (or say a few percent) the supersymmetry should be broken at the Planck scale (during inflation) so that the three-point vertex is present in the original classical theory. Notice that most of the string inflation models *do not* predict a large α_4 since they are based on compactifications of the ten dimensional superstring. It should be a model where the string length is comparable

[19] We thank Nima Arkani-Hamed for emphasizing this to us.

JHEP02(2016)020

to the Hubble radius and with a very weak coupling to account for the small experimental upper bound for H/M_p [74].

Note that if one imagines that inflation is given a dual description in the spirit of dS/CFT and the dual field theory is weakly coupled, then one expects that $\alpha_4 H^4 \sim 1$. This is what happens in the Vasiliev theory [75]. Of course, this theory also contains massless higher spin particles. It is also not suitable for building an inflationary model because the scalar does not appear to obey the slow roll conditions.

If the gravity waves produced by inflation are as large as to explain the signal seen by BICEP2 [76], then probing the gravity wave three-point functions might be possible (with a lot of optimism!).[20]

8 Conclusions

In this paper, we studied causality constraints on higher derivative corrections to the graviton three-point function. We considered a weakly coupled theory and studied higher derivative corrections which are important before the theory becomes strongly coupled. These are higher derivative corrections that arise in the classical regime of the theory.

The constraints arise from a thought experiment where we scatter two gravitons at relatively high energy and fixed impact parameter. The energy is high compared to the inverse of the impact parameter but low compared to the scale where the theory becomes strongly coupled. More explicitly, we have the very small overall coupling G and we consider corrections to the three-point functions which scale as powers of αp^2 relative to the Einstein one. The three-point amplitudes are very small because G is very small. But α is a fixed quantity and we look at impact parameters of the order of $b^2 \sim \alpha$. We found that when the impact parameter $b^2 \sim \alpha$, then we see a causality violation. In this impact parameter representation, and in the field theory regime (without higher spin particles), the time delay comes from the singularities in the t-channel, which is simply a pole at $t = 0$ for the massless theory. More precisely it comes from the part that is quadratic in s of its residue. In other words, terms going like s^2/t at $t = 0$. It is important that while $t = 0$, the momentum transfer itself is non-zero.[21] The overall sign of the time delay, then depends on the contractions of the polarization tensors of the external particles with the momentum transfer in the t-channel.

We have argued that this type of tree level causality violation can only be fixed, at tree level, by higher spin particles at a mass scale $m^2 \sim 1/\alpha$. In string theory, this issue is fixed because the amplitude Reggeizes. Namely, it has a behavior $s^{2+\frac{\alpha' t}{2}}$ for large s and fixed t. This is due to extended strings being exchanged in the s-channel. If the amplitude Reggeizes, then corrections appear at a scale $b^2 \sim \alpha' \log(s\alpha')$. Due to the presence of the logarithm we did not find a sharp bound between the corrections to the graviton three point amplitude, α, and the Regge slope α'.

[20] At this time, there are alternative explanations for this signal [77, 78], so we might have to wait till the dust settles.

[21] It is a null, non-zero momentum.

JHEP02(2016)020

We should stress that in this discussion we have assumed that we have a weakly coupled gravitational theory. We have also assumed the notion of asymptotic causality which says that the causal structured determined by the far away regions of spacetime cannot be violated by its interior regions.

The analysis in this paper was also motivated by trying to understand better the AdS/CFT correspondence. In particular, if we consider large N gauge theories we know that we have a weakly coupled theory in the bulk. However, we do not know under what conditions that weakly coupled theory is a local in the bulk. It is clear that the absence of light massive higher spin states is a requirement. Here we have tried to address the question of whether it is sufficient. We have only studied the simplest possible correction to gravity, its three-point function. We have argued that, as long as higher spin particles are very massive, there cannot be higher derivative corrections to the three-point functions. Previous discussions argued against such corrections by saying that they would make the theory strongly coupled at energies that are lower than the Planck scale [15, 16], but still parametrically larger than the scale of the corrections.[22] Here we have strengthened the bound by linking the scale of corrections to the appearance of new particles at the same scale.

As a more concrete statement, we are linking the values of the constants appearing in the stress tensor three-point functions to the dimensions of the lightest particles with higher spins, $J > 2$. In other words, $\frac{a-c}{c} \lesssim \frac{1}{\Delta_{\rm gap}^2}$. Unfortunately, we could not determine the precise numerical constant in this inequality.

Using the results in [58, 63, 64], we can link the time delay for a high energy scattering process in the bulk to the anomalous dimensions of certain double trace operators. These double trace operators have the rough form $T\partial_+^j(\partial^2)^n T$. They have both relative spin and relative radial excitations. The anomalous dimensions are $\gamma(n,j) \sim -\delta(s,b)/\pi$, where the values of b and s are given in terms of j, n in (5.22) (5.23). The requirement that the time delay is positive leads to the statement that the anomalous dimensions for some of these operators should be negative.

For the de Sitter case, this analysis has *potentially* interesting phenomenological applications. *If* the gravity wave three-point functions were measured, we expect to see the structure predicted by Einstein theory. However a new structure is also possible. This new structure is the only one allowed by the approximate scale and conformal invariance of the inflationary phase. *If* such new structure was found with a strength comparable to the Einstein one, then it would be a direct signal of dramatically new physics at the Hubble scale: a tower of higher spin particles. Which is a rather drastic departure from ordinary field theory at the inflationary scales. It is not clear how likely this inflationary scenario is in the space of possible inflationary theories.

8.1 Open problems

It would be nice to derive these constraints in a more direct way. If one understood directly the constraints of unitarity and causality at the level of the four point function,

[22] In [15, 16], or in talks referring to those papers, they impose the bound $\alpha \lesssim \frac{1}{G(\Delta_{\rm gap})^4}$ (with $G \sim 1/N^2$). This bound comes from demanding that the theory remains perturbatively unitary at the scale $\Delta_{\rm gap}$ where new particles appear. Here we argued for the stronger bound $\alpha \lesssim \frac{1}{\Delta_{\rm gap}^2}$.

JHEP02(2016)020

then one would not need to resort to the exponentiation argument we discussed in section 3.3. Furthermore, it might lead to sharper bounds that include numerical factors.

Our discussion of massive intermediate particles in mixed representations was not complete.[23] These are representations that have maximal spin two in the uv-plane, but have additional indices in the other directions (see appendix H). We suspect that a finite number of these cannot solve the causality problem, but we did not prove it.

In the AdS case, it would be nice to derive the constraints from the conformal bootstrap point of view. This is in the spirit of [15], but it would involve the stress tensor as an external state. One of the main messages from this paper is the importance of spin in order to derive constraints. The structure constants (or three-point functions) of operators with spin are very numerous but, as we have shown, there can be powerful constraints on them. These constraints are not so easily seen when we scatter external operators with no spin. Notice that, even the simplest bounds for a and c discussed in [14], which should be valid for arbitrary CFTs, have not been derived from the conformal bootstrap approach.

It would be nice to further constrain the interactions of all the higher spin particles and derive the general structure of the tree level theory. This is the program that was pursued in the sixties and that led to string theory. However, we would like to know how unique string theory is. Methods developed to tackle this problem might also be useful for analyzing large N gauge theories such as large N QCD.

It would also be nice to see whether in the de Sitter context there is a sharp bound for the gravity wave three-point correlators analogous to the one in [14] for AdS correlators. Our de Sitter discussion assumed the local gravity description and a locally flat space discussion in the bulk, so it applies most clearly when $\alpha_4 H^4$ is somewhat smaller than one, but still of order one compared with H/M_p.

Acknowledgments

We would like to thank N. Arkani-Hamed, S. Caron-Huot, A. Dymarsky, S. Giddings, S. He, D. Hofman, Y. Huang, Z. Komargodski, D. McGady, R. Myers, J. Penedones, L. Rodina, D. Simmons-Duffin, E. Skvortsov, A. Strominger and R. Wald for discussions. X.C. and J.E. thank the Institute for Advanced Study for hospitality at the initial stages of this work. J.E. is supported in part by MINECO and FEDER (grant FPA2011-22594 and FPA2014-52218), by Xunta de Galicia (GRC2013-024), and by the Spanish Consolider-Ingenio 2010 Programme CPAN (CSD2007-00042). J.M. is supported in part by U.S. Department of Energy grant DE-SC0009988. The Centro de Estudios Científicos (CECs) is funded by the Chilean Government through the Centers of Excellence Base Financing Program of Conicyt.

[23] These are not present in $D = 4$, so that the existence of an infinite tower of higher spin states is clear in $D = 4$, or if $\alpha_4 \neq 0$ in higher dimensions. In higher dimensional theories with only $\alpha_4 = 0$ but non-zero α_2, in order to establish that the tower is really infinite we need to rule out the possibility that the causality problem is fixed with a finite number of mixed representations. We leave this to the future.

JHEP02(2016)020

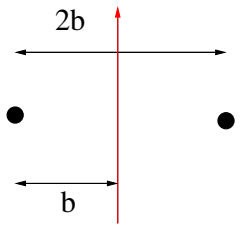


Figure 10. We consider the metric produced by two massive sources at distance $2b$. We then send a particle between them and measure the time delay. There is no deflection angle, but there is a non-zero time delay. We do the computation to leading order in the r_S/b expansion, by simply consider a linear superposition of the two metrics.

A Shapiro time delay

The physical effect discussed in the paper is known in General Relativity as the Shapiro time delay [1]. The usual setup to discuss the Shapiro time delay is to consider propagation of light near a massive body (a star or a planet). Here we would like to consider a slight variation of it by considering propagation of light between two massive bodies. We consider the masses to be equal and consider the geodesic that is equally separated from each of them, such that the deflection is absent (each of the masses bends the trajectory in the opposite direction such that the net deflection is zero). The time delay on the other hand accumulates. Recall that the Schwarzschild metric takes the form

$$ds^2 = -\left(1-\frac{r_S^{D-3}}{r^{D-3}}\right)dt^2 + \frac{dr^2}{1-\frac{r_S^{D-3}}{r^{D-3}}} + r^2 d\Omega^2 \,. \tag{A.1}$$

We are interested in the metric produced by the superposition of two equal masses. In the Schwarzschild coordinates to first order in the mass, or r_S, we get

$$ds^2 = ds^2_{\text{Mink}} + \sum_i \frac{r_S^{D-3}}{r_i^{D-3}}(dt^2 + dr_i^2) \,, \tag{A.2}$$

with $r_i = |\vec{x} - \vec{x}_i|$. Let us now consider two masses separated along the direction x^1, one at $x^1 = -b$ and one at $x^1 = b$. Then we consider a probe particle moving along the direction $x^{D-1} = z$. By symmetry, it will stay at $x^1 = \ldots = x^{D-2} = 0$ if it starts there with velocity along the z direction. We have that $r = \sqrt{b^2+z^2}$ for both masses. We also find that $dr = dz\, z/r$. We get

$$dt^2\left(1-2\frac{r_S^{D-3}}{r^{D-3}}\right) = dz^2\left(1+2\frac{r_S^{D-3}}{r^{D-3}}z^2/r^2\right) . \tag{A.3}$$

Then the time delay is

$$\Delta t = \int_{-\infty}^{\infty} dz\, r_S^{D-3}\frac{1}{(b^2+z^2)^{\frac{D-3}{2}}}\left(1+\frac{z^2}{z^2+b^2}\right) = \frac{r_S^{D-3}}{b^{D-4}}\frac{(D-2)\sqrt{\pi}\Gamma(\frac{d}{2}-2)}{2\Gamma(\frac{D-1}{2})} \,. \tag{A.4}$$

JHEP02(2016)020

We now do the same for a particle moving at a velocity v. We find

$$\begin{aligned} dt^2\left(1-2\frac{r_S^{D-3}}{r^{D-3}}\right) &= dz^2\left(1+2\frac{r_S^{D-3}}{r^{D-3}}\frac{z^2}{z^2+b^2}\right)+d\tau^2\,,\\ \frac{dt}{d\tau}\left(1-2\frac{r_S^{D-3}}{r^{D-3}}\right) &= \gamma\,, \qquad \gamma^{-2}=1-v^2\,,\\ dt &= \frac{dz}{v}\left(1+\left(2+2\frac{z^2}{z^2+b^2}-\frac{1}{v^2}\right)\frac{r_S^{D-3}}{r^{D-3}}\right)\,. \end{aligned} \tag{A.5}$$

The result of doing the integral over z is the same as in (A.4) multiplied by a factor of

$$\Delta t = \frac{1}{v}\frac{(2+\frac{1}{D-3}-\frac{1}{v^2})}{(1+\frac{1}{D-3})}\Delta t_{\text{relativistic}}\,. \tag{A.6}$$

We see that for slow velocities, we indeed find a negative time delay (time advance) proportional to $1/v^3$. This is time advance relative to the particle moving with the same velocity v in flat space (not relative to a particle moving at the speed of light!).

We could have repeated the computation in different coordinate system. For example, we can consider the computation in the so-called isotropic coordinates given by

$$\begin{aligned} ds^2 &= -\left(\frac{1-\frac{r_S^{D-3}}{4r^{D-3}}}{1+\frac{r_S^{D-3}}{4r^{D-3}}}\right)^2 dt^2+\left(1+\frac{r_S^{D-3}}{4r^{D-3}}\right)^{\frac{4}{D-3}}\sum_{i=1}^{D-1}dx_i^2\,,\\ r^2 &= \sum_{i=1}^{D-1}x_i^2\,. \end{aligned} \tag{A.7}$$

One can easily check that the time delay for the geodesic in these coordinates coincides with the one computed in Schwarzschild coordinates to leading order in r_S.

B Three-point amplitudes and their sums

In this appendix we recollect different three-point amplitudes that involve a graviton and two other particles and present different polarization sums that appear in the time delay. In particular cases we reproduce the shock wave computations but the results obtained using on-shell amplitudes are much more general and are valid in any theory with given on-shell three-point amplitudes.

Let us be more explicit on the kinematics we are interested in. As defined in the bulk of the paper we consider the following kinematics

$$\begin{aligned} p_{1\,\mu} &= \left(p_u, \frac{q^2}{16p_u}, \frac{\vec{q}}{2}\right)\,, & p_{2\,\mu} &= \left(\frac{q^2}{16p_v}, p_v, -\frac{\vec{q}}{2}\right)\,,\\ p_{3\,\mu} &= -\left(p_u, \frac{q^2}{16p_u}, -\frac{\vec{q}}{2}\right)\,, & p_{4\,\mu} &= -\left(\frac{q^2}{16p_v}, p_v, \frac{\vec{q}}{2}\right)\,, \end{aligned} \tag{B.1}$$

JHEP02(2016)020

thereby $s \simeq 4p_u p_v$ and $t \simeq -(\vec{q})^2$. We define the polarization tensors as follows

$$\epsilon^{1\,\mu} = \left(-\frac{\vec{q}.\vec{e}_1}{2p_u}, 0, \vec{e}_1 \right), \qquad \epsilon^{3\,\mu} = \left(\frac{\vec{q}.\vec{e}_3}{2p_u}, 0, \vec{e}_3 \right),$$
$$\epsilon^{2\,\mu} = \left(0, \frac{\vec{q}.\vec{e}_2}{2p_v}, \vec{e}_2 \right), \qquad \epsilon^{4\,\mu} = \left(0, -\frac{\vec{q}.\vec{e}_4}{2p_v}, \vec{e}_4 \right). \tag{B.2}$$

An important point is that all $\epsilon_i.p_j$ are of order 1 and, thus, are sub-leading in the high energy limit compared to powers of s. Thus, in all our computations we can think of polarization tensors as being purely transverse, see (3.5).

We focus on the massless scalar, vector and graviton three-point amplitudes which are parameterized as follows

$$\begin{aligned}
\mathcal{A}_{hhg} &= \sqrt{32\pi G}\epsilon_{\mu\nu}p_1^\mu p_3^\nu\,, \\
\mathcal{A}_{\gamma\gamma g} &= \sqrt{32\pi G}\epsilon_{\mu\nu}\left([p_1^\mu p_3^\nu(\epsilon_1.\epsilon_3) - \epsilon_1^\mu p_3^\nu(\epsilon_3.p_1) - \epsilon_3^\mu p_1^\nu(\epsilon_1.p_3)] + \hat{\alpha}_2 p_1^\mu p_3^\nu(\epsilon_1.p_3)(\epsilon_3.p_1)\right)\,, \\
\mathcal{A}_{ggg} &= \sqrt{32\pi G}\left[(\epsilon_1.\epsilon_2\epsilon_3.p_1 + \epsilon_1.\epsilon_3\epsilon_2.p_3 + \epsilon_2.\epsilon_3\epsilon_1.p_2)^2\right. \\
&\qquad +\alpha_2(\epsilon_1.\epsilon_2\epsilon_3.p_1 + \epsilon_1.\epsilon_3\epsilon_2.p_3 + \epsilon_2.\epsilon_3\epsilon_1.p_2)\epsilon_1.p_2\epsilon_2.p_3\epsilon_3.p_1 \\
&\qquad \left.+\alpha_4(\epsilon_1.p_2\epsilon_2.p_3\epsilon_3.p_1)^2\right]\,.
\end{aligned} \tag{B.3}$$

where we used the $\epsilon^{\mu\nu} \to \epsilon^\mu \epsilon^\nu$ form of the polarization tensor for the graviton . Using these three-point amplitudes we can compute the time delay using (3.1) in the high energy limit. In the high energy limit the relevant part of the sum over graviton polarization tensors takes a very simple form

$$\sum_i \epsilon^i_{\mu\nu}(q)(\epsilon^i_{\rho\sigma}(q))^* \sim \frac{1}{2}\left(\eta_{\mu\rho}\eta_{\nu\sigma} + \eta_{\mu\sigma}\eta_{\nu\rho}\right)\,, \tag{B.4}$$

so that it leads to factors of s^2 when we contract with the p_1 or p_3 momenta from the left side and with p_2 or p_4 momenta on the right side, see figure 4. Note that the large components of the external momenta are transverse to the momentum $\vec{q}$ of the intermediate particle, $p_1 + p_3 = (0, 0, \vec{q})$, see (B.1).

The results are the following

$$\begin{aligned}
\sum \mathcal{A}_{hhg}(-i\partial_{\vec{b}})\mathcal{A}_{hhg}(-i\partial_{\vec{b}}) &= 8\pi G s^2\,, \\
\sum \mathcal{A}_{hhg}(-i\partial_{\vec{b}})\mathcal{A}_{\gamma\gamma g}(-i\partial_{\vec{b}}) &= 8\pi G s^2\left(e_2.e_4 + \hat{\alpha}_2 e_2.\partial_b e_4.\partial_b\right)\,, \\
\sum \mathcal{A}_{hhg}(-i\partial_{\vec{b}})\mathcal{A}_{ggg}(-i\partial_{\vec{b}}) &= 8\pi G s^2\left(e_2^{ij}e_4^{ij} + \alpha_2 e_2^{ij}e_4^{ik}\partial_{b^j}\partial_{b^k} + \alpha_4 e_2^{ij}e_4^{kl}\partial_{b^i}\partial_{b^j}\partial_{b^k}\partial_{b^l}\right)\,, \\
\sum \mathcal{A}_{\gamma\gamma g}(-i\partial_{\vec{b}})\mathcal{A}_{\gamma\gamma g}(-i\partial_{\vec{b}}) &= 8\pi G s^2\left(e_1.e_3 + \hat{\alpha}_2 e_1.\partial_b e_3.\partial_b\right)\left(\epsilon_2.e_4 + d_2 e_2.\partial_b \epsilon_4.\partial_b\right)\,, \\
\sum \mathcal{A}_{\gamma\gamma g}(-i\partial_{\vec{b}})\mathcal{A}_{ggg}(-i\partial_{\vec{b}}) &= 8\pi G s^2\left(e_1.e_3 + \hat{\alpha}_2 e_1.\partial_b e_3.\partial_b\right) \\
&\qquad \left(e_2^{ij}e_4^{ij} + \alpha_2 e_2^{ij}e_4^{ik}\partial_{b^j}\partial_{b^k} + \alpha_4 e_2^{ij}e_4^{kl}\partial_{b^i}\partial_{b^j}\partial_{b^k}\partial_{b^l}\right)\,, \\
\sum \mathcal{A}_{ggg}(-i\partial_{\vec{b}})\mathcal{A}_{ggg}(-i\partial_{\vec{b}}) &= 8\pi G s^2(e_1^{ij}e_3^{ij} + \alpha_2 e_1^{ij}e_3^{ik}\partial_{b^j}\partial_{b^k} + \alpha_4 e_1^{ij}e_3^{kl}\partial_{b^i}\partial_{b^j}\partial_{b^k}\partial_{b^l}) \\
&\qquad \left(e_2^{ij}e_4^{ij} + \alpha_2 e_2^{ij}e_4^{ik}\partial_{b^j}\partial_{b^k} + \alpha_4 e_2^{ij}e_4^{kl}\partial_{b^i}\partial_{b^j}\partial_{b^k}\partial_{b^l}\right)\,.
\end{aligned} \tag{B.5}$$

JHEP02(2016)020

where all of the operators are acting on the propagator $1/b^{D-4}$. To reproduce the usual shock wave computations only the first three formulas are relevant. In the case of electrodynamics matching with (2.16) is manifest. In the case of Gauss-Bonnet theory we have from (2.19) $\alpha_2 = \frac{\lambda_{GB}}{4}$, $\alpha_4 = 0$.

Consider now the coupling of the graviton to a massive spin two particle which is a relevant amplitude for the discussion in the bulk of the paper. We get

$$\begin{aligned}\mathcal{A}_{gg\widetilde{g}} = \widetilde{\alpha}_2\epsilon_{\mu\nu}\left[\epsilon_1^\mu p_3^\nu(\epsilon_3.p_1) + \epsilon_3^\mu p_1^\nu(\epsilon_1.p_3) - p_1^\mu p_3^\nu(\epsilon_1.\epsilon_3) - \epsilon_1^\mu\epsilon_3^\nu(p_1.p_3)\right] \\ \left[(\epsilon_1.\epsilon_3)(p_1.p_3) - (\epsilon_3.p_1)(\epsilon_1.p_3)\right]] \\ +\widetilde{\alpha}_4\epsilon_{\mu\nu}p_1^\mu p_3^\nu\left[(\epsilon_1.p_3)(\epsilon_3.p_1) - (\epsilon_1.\epsilon_3)(p_1.p_3)\right]^2 .\end{aligned} \tag{B.6}$$

Notice that instead of three structures that are present in the massless amplitude we have only two.

For the coupling to spin four particle or higher we can have three structures [79]. They take the same form as above with extra indices of particle of spin J being contracted with momenta. An additional structure takes the form

$$\begin{aligned}\widetilde{c}_{\text{extra}}\epsilon_{\mu\nu\rho\sigma}.p_1\ldots p_1\left[\epsilon_1^\mu p_3^\nu(\epsilon_3.p_1) + \epsilon_3^\mu p_1^\nu(\epsilon_1.p_3) - p_1^\mu p_3^\nu(\epsilon_1.\epsilon_3) - \epsilon_1^\mu\epsilon_3^\nu(p_1.p_3)\right] \\ \left[\epsilon_1^\rho p_3^\sigma(\epsilon_3.p_1) + \epsilon_3^\rho p_1^\sigma(\epsilon_1.p_3) - p_1^\rho p_3^\sigma(\epsilon_1.\epsilon_3) - \epsilon_1^\rho\epsilon_3^\sigma(p_1.p_3)\right] .\end{aligned} \tag{B.7}$$

We are interested in the contribution of new particles to the time delay. The computation is almost identical to the one we did for the graviton exchange. The first difference is that we have to compute the Fourier transform of the massive propagator in $D-2$ dimensions. The result is given by $(\frac{m}{b})^{\frac{D-4}{2}}K_{\frac{D-4}{2}}(mb)$ which decays exponentially fast for $mb \gg 1$. The second difference comes from slightly different structure of the three-point functions.

Similarly, we can write down the contribution to the time delay of some higher spin particle the difference being that the sum of three-point amplitudes give s^J and we have an extra structure in the three-point amplitude for external gravitons which is an analogous to the Einstein one.

The formulas above are valid in $D > 4$, whereas in $D = 4$ α_2 and $\widetilde{\alpha}_2$ structures above are absent[24] but instead we have a parity odd structure. In $D \geq 5$ parity odd structures are absent in the class of amplitudes considered above.

B.1 Scattering of a scalar and a graviton

As an example let us reproduce the computation for the graviton that we did using the shock wave. It corresponds to scattering from the energetic scalar particle (we could have considered graviton-graviton scattering as well if we average of all polarizations of gravitons 1 and 3),[25]

$$\sum_{\text{states}}\mathcal{A}^{hhg}(-i\partial_{\vec{b}})\mathcal{A}^{ggg}(-i\partial_{\vec{b}}) \propto Gs^2(e_2^{ij}e_4^{ij} + \alpha_2 e_2^{ij}e_4^{ik}\partial_{b^j}\partial_{b^k} + \alpha_4 e_2^{ij}e_4^{kl}\partial_{b^i}\partial_{b^j}\partial_{b^k}\partial_{b^l})\frac{1}{b^{D-4}} . \tag{B.8}$$

[24] They are identically zero.

[25] More precisely we set polarizations one and three to be equal and then we sum over all of them. Physically we are considering a sequence of coherent state with the various alternative polarizations. This makes the α_2 and α_4 contributions vanish in the $\mathcal{A}^{13I}$ part of the amplitude.

JHEP02(2016)020

Setting $\alpha_4 = 0$, $\alpha_2 = -\frac{\lambda_{GB}}{4}$ we reproduces the computation from the previous section (2.19).

Let us now study the constraints that follow from the positivity of the time delay. Introducing new variables

$$\begin{aligned} t_2 &= \frac{(D-2)(D-4)\alpha_2}{b^2} - \frac{4(D-4)(D-2)D\alpha_4}{b^4}\,, \\ t_4 &= \frac{(D-4)(D-2)D(D+2)\alpha_4}{b^4}\,, \end{aligned} \tag{B.9}$$

we can write the phase shift in the form familiar from the study of the energy correlators

$$\delta(s,\vec{b}) \sim 1 + t_2\left(\frac{(e.n)^2}{e.e} - \frac{1}{D-2}\right) + t_4\left(\frac{(e.n)^4}{(e.e)^2} - \frac{2}{D(D-2)}\right)\,, \tag{B.10}$$

which coincides with the formula (3.6) in [41, 80].[26] Here $\vec{n} = \frac{\vec{b}}{|\vec{b}|}$ and $\vec{e}$ is the graviton polarization of particles two and four. Thus, positivity constraints from causality are identical to the ones obtained in their analysis with identification of parameters as above (B.9), namely

$$\begin{aligned} 1 - \frac{t_2}{D-2} - \frac{2t_4}{D(D-2)} &\geq 0\,, \\ \left(1 - \frac{t_2}{D-2} - \frac{2t_4}{D(D-2)}\right) + \frac{t_2}{2} &\geq 0\,, \\ \left(1 - \frac{t_2}{D-2} - \frac{2t_4}{D(D-2)}\right) + \frac{D-3}{D-2}(t_2+t_4) &\geq 0\,. \end{aligned} \tag{B.11}$$

So we get the bounds on α_2 and α_4 depending for how small a b we can trust the computation. If the computation is trustworthy for arbitrarily small b we are forced to set α_2 and α_4 to zero.

C Higher derivative terms from integrating out particles

C.1 QED case

Note that the action (2.13) also arises in QED (quantum electrodynamics) after we integrate out the massive electron [81]. In that case $\hat{\alpha}_2 \propto \frac{e^2}{m^2}$. This is a one loop effect, suppressed by the coupling e^2. In this paper we have taken the coupling to be very small, so that we would have treated this $\hat{\alpha}$ as being essentially zero. The discussion of this paper concentrated on the case in which the higher curvature corrections were present at tree level, so that the causality problem had to be solved by tree level physics. In this appendix, we consider this loop generated term in QED and we will show that the potential causality problem is solved by one loop effects.

In QED, when we get to an impact parameter of order m^{-1} we cannot be satisfied with the low energy action (2.13). Fortunately, the necessary one loop diagrams were computed in [82]. Using their results it is possible to go to the impact parameter representation (doing

[26] In [41] $d = D - 1$.

JHEP02(2016)020

the Fourier transform) and check that for $b > 1/m$ we get a result that agrees with the simple Lagrangian (2.13), but for $b < 1/m$ we get a different result which displays no causality problem. In other words, the potential causality problems arising from (2.13) would appear at $b \sim \sqrt{|\hat{\alpha}_2|} \sim e/m$, but at a larger distance, $b \sim 1/m$, the computation should be already modified and we obtain results consistent with causality.[27] We can view this modification as arising from the propagation of electron positron pairs along the t-channel.

Notice, that, in addition, when the photon goes through the shock we can have electron positron pair creation. We can view this as another example of tidal excitations. Indeed in QED, the photon has a non-zero probability of becoming an electron positron pair.

It is also possible to consider an example where we have N species of electrons. In this case the loop effect has an additional factor of N so that we can have a large and "classical" contribution to $\hat{\alpha}_2 \sim \frac{e^2 N}{m^2}$. We call it classical because it can have a finite value in the $e^2 \to 0$ limit if Ne^2 remains fixed. As before, here we expect no problems since the original theory did not have any causality problem. We had suggested in the paper that in theories where $\hat{\alpha}_2$ is present in the classical theory, then higher spin fields should solve the problem. In this example, these higher spin fields are just electron-positron bilinears. Due to large N O(N) counting, these bilinear behave as classical fields. We can say that those fields are described by a Vasiliev-like theory.

C.2 Gravity case

In D dimensions, integrating out N particles of mass m is generically expected to produce corrections to the Einstein Lagrangian of the form

$$M_p^{D-2} R + N m^{D-4} R^2 + N m^{D-6} R^3 \,, \tag{C.1}$$

which lead to $\alpha_2 \sim \frac{N m^{D-4}}{M_p^{D-2}}$ and $\alpha_4 \sim \frac{N m^{D-6}}{M_p^{D-2}}$.

The amplitudes computed using this effective action should be modified when the impact parameter b is smaller than $b_m \sim 1/m$. We can easily check that $\alpha_2/b_m^2 \sim \alpha_4/b_m^4 \sim N \frac{m^{D-2}}{M_p^{D-2}}$. Since we expect that in such a theory $N m^{D-2} \lesssim M_p^{D-2}$, for the theory to be local at scale m.

Since N can be very large, we see that the corrections in (C.1) can be viewed as "classical", in the sense that they are much larger than the values expected from effective field theory applied to the Einstein action. In this case, we see that the potential causality problem raised by these extra terms is solved by adding the mass m particles. In this case we did not need to include any elementary higher spin fields. Nevertheless, there is a sense in which the weakly coupled higher spin fields are still present. In fact, the theory of N particles of mass m, for large N, could have a Vasiliev-like dual with weakly coupled higher spin fields.

D Causality and unitarity for a signal model

In this appendix we review the constraints from causality and unitarity in the context of a simple signal model. We imagine a signal propagating along one dimension. We have an

[27] See [83] and references therein for a related discussion of this problem.

JHEP02(2016)020

initial signal which is a function of time $f_{\rm in}(t)$ and an out-signal $f_{\rm out}(t)$ which, in Fourier space is given by $f_{\rm out}(\omega) = S(\omega) f_{\rm in}(\omega)$ or

$$f_{\rm out}(t) = \int dt' \int d\omega S(\omega) e^{-i\omega(t-t')} f_{\rm in}(t') \,. \tag{D.1}$$

Causality implies that if $f_{\rm in}(t') = 0$ for $t' < 0$, then $f_{\rm out}(t) = 0$ for $t < 0$. By unitarity we mean that the L_2 norm of the out-signal should be smaller than that of the in-signal $\int dt |f_{\rm out}(t)|^2 \leq \int dt |f_{\rm in}(t)|^2$. Now it is well known that the Fourier transform of a function which vanishes for $t < 0$ is analytic in the upper half ω plane. This follows directly from the explicit integral expression for the Fourier transform. Then if $f_{\rm in} = 0$ for $t < 0$ we find that both $f_{\rm in}(\omega)$ and $f_{\rm out}(\omega)$ are analytic in the upper half plane. This also implies that $S(\omega)$, which is given their ratio, is also analytic. One might worry that $S(\omega)$ could have poles at zeros of $f_{\rm in}(\omega)$. However, we can change the location of the zeros of $f_{\rm in}(\omega)$ by choosing different functions. Therefore $S(\omega)$ is analytic in the upper half plane.

We will now prove that unitarity implies that $|S(\omega)| \leq 1$ in the upper half plane. With some foresight, we pick a particular $f_{\rm in}(t)$ of the form

$$f_{\rm in}(t) = e^{-\gamma t} e^{-i\omega_0 t} \theta(t) \sqrt{2\gamma} \,, \tag{D.2}$$

with $\gamma > 0$ and ω_0 real. Note that $||f_{\rm in}||^2 = \int dt |f_{\rm in}(t)|^2 = 1$. For $Im(\omega) > 0$ we can now write

$$\begin{aligned} |f_{\rm out}(\omega)|^2 &\leq \left| \int_0^\infty dt e^{i\omega t} f_{\rm out}(t) \right| \leq \int_0^\infty dt |e^{i\omega t}|^2 \int_0^\infty dt |f_{\rm out}(t)|^2 = \frac{1}{2Im(\omega)} ||f_{\rm out}||^2 \\ |f_{\rm out}(\omega)|^2 &\leq \frac{1}{2Im(\omega)} \\ |S(\omega)|^2 &= \frac{|f_{\rm out}(\omega)|^2}{|f_{\rm in}(\omega)|^2} \leq \frac{1}{2Im(\omega)} \frac{1}{|f_{\rm in}(\omega)|^2} \,. \end{aligned} \tag{D.3}$$

Here we used the Cauchy-Schwartz inequality. Note that $|e^{i\omega t}|^2 = e^{-2Im(\omega)t}$. We also used that $||f_{\rm out}||^2 \leq ||f_{\rm in}||^2 = 1$. We can now set $\omega = \omega_0 + i\gamma$ and find that $f_{\rm in}(\omega_0 + i\gamma) = 1/\sqrt{2\gamma}$ for the specific function (D.2). Inserting this into (D.3) we then find that $|S(\omega_0 + i\gamma)| \leq 1$, which is what we wanted to prove, since ω_0 and γ are arbitrary.

In conclusion, we find that $S(\omega)$ should be analytic and bounded $|S(\omega)| \leq 1$ in the upper half plane. These are necessary and sufficient conditions.[28]

Let us now briefly mention how this is connected to the field theory situation. We consider light cone coordinates u and v. We consider a perturbation that is translation invariant in v but is localized in the u coordinate. We call this "the shock". We expand the fields in the v coordinate at some $u_{\rm in}$ and then we expand them again at some $u_{\rm out}$ after the shock. To make contact with the above discussion we call $t = v$ and $p_v = -\omega$. We can expand the field as

$$\phi(t) = \int_0^\infty \frac{d\omega}{\sqrt{\omega}} (a_\omega e^{-i\omega t} + a_\omega^\dagger e^{i\omega t}) \,. \tag{D.4}$$

[28] Note that some functions which are analytic in the upper half plane, such as $S(\omega) = e^{i\omega^3}$ are actually not causal.

JHEP02(2016)020

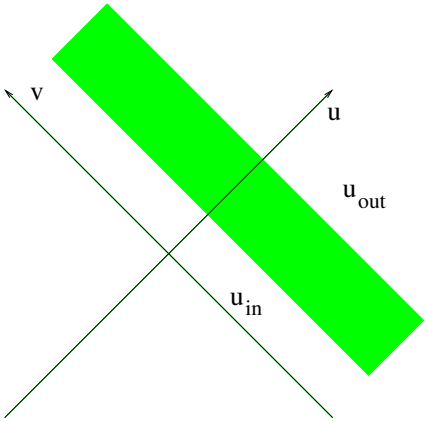


Figure 11. We consider a v independent perturbation that is localized in the u direction, given here by the shaded region. We then consider signal propagating along the u direction, which are v dependent and demand causality. We can consider an S matrix that connects the region before the perturbation to the region after the perturbation.

We can do this for $\phi_{\rm in}$ and $\phi_{\rm out}$ in terms of $a_{\rm in}$ and $a_{\rm out}$. These oscillators then are related by

$$a_{\omega,\rm out} = S(\omega) a_{\omega,\rm in}\,, \qquad a^\dagger_{\omega,\rm out} = S(\omega)^* a^\dagger_{\omega,\rm in}\,. \tag{D.5}$$

This defines $S(\omega)$ for positive ω. For negative ω we can define $S(-\omega) = S(\omega)^*$. Alternatively, we can define

$$S(\omega) = -\int_{-\infty}^{\infty} dt e^{i\omega t}[\phi_{\rm out}(t), i\partial_t \phi_{\rm in}(0)] = -\int_0^{\infty} dt e^{i\omega t}[\phi_{\rm out}(t), i\partial_t \phi_{\rm in}(0)]\,. \tag{D.6}$$

The commutation relations for the in and out oscillators require that $|S(\omega)|^2 = 1$.

In a case where there is particle mixing, but no particle creation, then the fields have indices $\phi^i_{\rm in}$ and $\phi^j_{\rm out}$. Now S is a matrix which obeys $S_{ij}(-\omega) = S_{ij}(\omega)^*$ and $S_{ij}(\omega) S_{kj}(\omega)^* = \delta_{ik}$ due to the commutation relations of the oscillators before and after the shock. In the preceding discussion we have neglected the transverse dimensions. We can now remedy that by including the momentum in the transverse dimensions as part of the indices we are discussing here.

If we consider a signal that is made out of physical particles, one might correctly worry that the fact that $\omega > 0$ will preclude us from localizing the signal in time. In order to avoid this issue we can consider a coherent state of the form

$$|\Psi\rangle = e^{i\int dt f_{\rm in}(t)\phi_{\rm in}(t)}|0\rangle\,, \tag{D.7}$$

with a real function $f_{\rm in}$. This is a state that could be produced by adding a hermitian term to the Hamiltonian at some early time $u_{\rm in}$. On this state we have the expectation values

$$\langle\Psi|\partial_t\phi_{\rm in}(t)|\Psi\rangle = f_{\rm in}(t)\,, \qquad \langle\Psi|\partial_t\phi_{\rm out}(t)|\Psi\rangle = f_{\rm out}(t)\,, \tag{D.8}$$

where the functions are related as in the signal model.[29] Here we assumed a linear relation between the in- and out-signals. Furthermore, we can also consider the expectation values

[29] $f_{\rm out}$ is real if $f_{\rm in}$ is real when $S(-\omega) = S(\omega)^*$.

JHEP02(2016)020

of the normal ordered product $T_{vv} = T_{tt} =: \partial_t\phi(t)\partial_t\phi(t) :$. When this is evaluated on the state (D.7) , and integrated over t we find that the answer is given by

$$-P_v^{\text{in,out}} = \int dv T_{vv} = \int dt (f_{\text{in,out}}(t))^2 = ||f_{\text{in,out}}||^2 \,. \tag{D.9}$$

Thus, the condition that the total light-cone momentum P_v should not increase implies the norm condition $||f_{\text{out}}||^2 \leq ||f_{\text{in}}||^2$.

More precisely, we can consider the signal f_{in} exciting a mode involving a graviton with a given polarization. The signal f_{out} is the same mode of the graviton. In addition, the initial graviton could go into other massive particles. Then the condition that the total P_v in the out-graviton mode should be no bigger than the initial P_v, which was all contained in the graviton mode, leads to the norm condition (or unitarity condition) for the signal model. In conclusion, the graviton-graviton matrix element $S_{gg}(\omega)$ obeys all the assumptions of the signal model. Therefore, it should be analytic and $|S_{gg}(\omega)| \leq 1$ in the upper half plane.

Note that we have assumed here a perfect v-translation symmetry for the perturbation that creates the shock. In our scattering problem, see figure 3, particles 1 and 3 have small p_v momentum. Thus, in this discussion, we have neglected this small momentum. This is reasonable for $sb^2 \gg 1$.

E Scattering in string theory

String theory in flat space is the simplest example of a theory that follows into the category of weakly coupled gravitational theories with higher derivative corrections that are subject of our analysis. As explained in the introduction, a motivation for this work was actually to argue that string theory is inevitable, at least, under certain assumptions.

It is well known that effective gravitational actions in string theory contain higher derivative corrections at the string scale [84, 85]. In particular graviton three-point amplitudes can contain the higher derivative terms that we constrained in this paper using causality. The potential problem is fixed by extra particles with string scale masses. Here we would like to understand how this is happening in detail.

Let us first recall the form of the three-point gravity amplitudes in bosonic, heterotic and type II string theories [86]

$$\begin{aligned} \mathcal{A}^{ggg} &= \sqrt{32\pi G}\epsilon_1^{\mu_1\mu_1'}\epsilon_2^{\mu_2\mu_2'}\epsilon_3^{\mu_3\mu_3'}N_{\mu_1\mu_2\mu_3}\bar{N}_{\mu_1'\mu_2'\mu_3'}\,, \\ N^{\mu_1\mu_2\mu_3} &= k_2^{\mu_1}\eta^{\mu_2\mu_3} + k_3^{\mu_2}\eta^{\mu_1\mu_3} + k_1^{\mu_3}\eta^{\mu_1\mu_2} + \frac{\alpha'}{2}\epsilon k_2^{\mu_1}k_3^{\mu_2}k_1^{\mu_3}\,, \end{aligned} \tag{E.1}$$

and we have $\epsilon_{bos} = \bar{\epsilon}_{bos} = \epsilon_{het} = 1$ and $\epsilon_{II} = \bar{\epsilon}_{II} = \bar{\epsilon}_{het} = 0$. Translating it to our notations we get

$$\begin{aligned} \alpha_2^{bos} &= 2\alpha_2^{het} = \alpha'\,, & \alpha_2^{II} &= 0\,, \\ \alpha_4^{bos} &= \frac{(\alpha')^2}{4}\,, & \alpha_4^{het} &= \alpha_4^{II} = 0\,. \end{aligned} \tag{E.2}$$

JHEP02(2016)020

The vanishing of some of these corrections can be understood from supersymmetry, as explained in section 3.2. For the purposes of this paper, the type II case is not interesting since there are not corrections at all for the graviton three-point function.

The high energy scattering problem in string theory was studied in a nice series of papers by Amati, Ciafaloni and Veneziano [51–55], see also e.g. [34, 87]. Let us first review their picture. The scattering can be described in terms of a phase shift defined as

$$\begin{aligned} \delta(\vec{b},s) &= [\mathrm{POL}]\delta^{ACV}(\vec{b},s)\,, \\ \delta^{ACV}(\vec{b},s) &= \int \frac{d^{D-2}\vec{q}}{(2\pi)^{D-2}} e^{i\vec{q}.\vec{b}} C(s,t,u)\,, \\ C(s,t,u) &= \frac{\Gamma(-\frac{\alpha' s}{4})\Gamma(-\frac{\alpha' t}{4})\Gamma(-\frac{\alpha' u}{4})}{\Gamma(1+\frac{\alpha' s}{4})\Gamma(1+\frac{\alpha' t}{4})\Gamma(1+\frac{\alpha' u}{4})}\,, \end{aligned} \tag{E.3}$$

where [POL] represents a factor that depends on the polarizations and is a polynomial in the momenta. We will only need its form in a specific limit. In the high energy limit $C(s,t,u)$ has the celebrated Regge behavior

$$C(s,t,u) \sim \frac{\Gamma(-\frac{\alpha' t}{4})}{\Gamma(1+\frac{\alpha' t}{4})} \left(-i\frac{s\alpha'}{4}\right)^{-2+\frac{t\alpha'}{2}}\,. \tag{E.4}$$

This Regge form is reflecting the creation of particles in the s-channel. The infinite sequence of s-channel poles is becoming a cut, the cut arising when $s \to s e^{2\pi i}$. The creation of the massive s-channel states is also related to the fact that we get an imaginary part from the $(-i)^{\frac{t\alpha'}{2}}$ factor in (E.4) is saying that the most likely process is to create a massive closed string, rather than scattering the gravitons.

For large s, only small q will contribute and we can approximate the prefactor in (E.4) by $1/t$. Then the integral becomes

$$\delta^{[ACV]} \propto \int \frac{d^{D-2}\vec{q}}{2(2\pi)^{D-2}} \frac{e^{i\vec{q}.\vec{b}} \left(-i\frac{s\alpha'}{4}\right)^{-\frac{\vec{q}^2\alpha'}{2}}}{\vec{q}^2} = \frac{1}{(\frac{\alpha' Y}{2})^{\frac{D-4}{2}}} \int_0^1 d\rho \rho^{\frac{D-6}{2}} e^{-\frac{b^2}{2\alpha' Y}\rho}\,, \tag{E.5}$$

where $Y = \log(-is\alpha'/4)$. We have two characteristic behaviors, depending on whether b^2 is larger or smaller than $\alpha' \log(s\alpha')$. For large b we get the usual $1/b^{D-4}$ behavior. For small b we get the result in brackets in (4.8). Note that since the transition region occurs for a b^2 which is larger than α' by a $\log s$ factor, we can, a posteriori, justify the fact that we have approximated the prefactor in (E.4) by $1/t$.

In our field theory discussion we had represented the phase shift as a sum over poles. We can wonder how this applies to the string theory discussion. Notice that from (E.4) we get a Gaussian integrand factor of the form

$$e^{i\vec{q}.\vec{b}} e^{-(\vec{q})^2 \frac{\alpha'}{2} \log(-is\alpha'/4)}\,. \tag{E.6}$$

Let us assume that $\vec{b} = (b, 0, \cdots, 0)$, so that it points along the first coordinates. When we do the integral over the first component of $\vec{q}$, call it q_1, we get a saddle point for (E.6) at

$$q_s = i\frac{b}{\alpha' \log(-is\alpha'/4)}\,. \tag{E.7}$$

JHEP02(2016)020

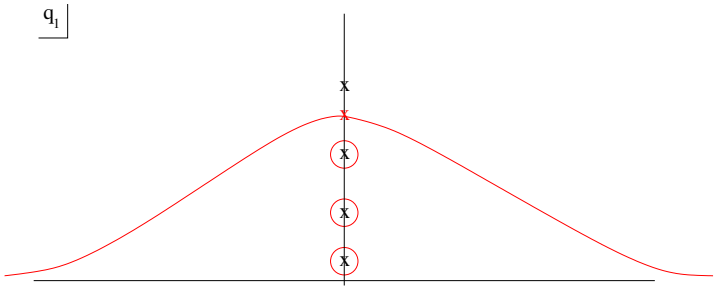


Figure 12. We display the complex q_1 plane. We have displayed the poles in the t-channel by black crosses. The saddle point (E.7) of the Gaussian integral has been denoted by a red cross. We have shifted the original integration contour for q_1, which was along the real axis, to the complex plane so that it passes through the saddle point (E.7). In the process we have picked up some of the poles in the t-channel.

It is thus convenient to shift the contour to this location, (E.7) , where this saddle point contribution gives something of the order of

$$e^{-\frac{b^2}{2\alpha' \log(-is\alpha'/4)}} . \tag{E.8}$$

This is not the whole answer, since by shifting the contour to this location, we can pick up some poles from the prefactor (E.4) , see figure 12. We always pick up the pole at $t = 0$,[30] which was the center of our gravity discussion, but we can also pick up the poles at $t = \frac{4}{\alpha'} n$ for $n < -q_s^2$, where q_s is given in (E.7). Notice that at these saddles the t dependent part of the exponent gives us factors of s^{2n} as we expect for the corresponding spins,[31] once we take into account that [POL] contains a factor of s^4. When b is large $b^2 \gg \alpha' \log(s\alpha')$, we pick up many poles, but when b is small we only pick up the $t = 0$ pole, but, even then, the integral is more accurately computed using (E.5). Note that the factor [POL] in string theory phase simplifies at $t = 0$ and becomes the product of three-point amplitudes we discussed in the body of the paper and that we added as Pol in (4.8). In other words [POL]$\to$Pol as $t \to 0$.

In particular, note that the residues of poles associated to the massive states go as $\frac{1}{(n!)^2} e^{-b\sqrt{\frac{4n}{\alpha'}}} s^{2n}$. As a function of n, these contributions decrease and then start increasing again with a transition at a value of n corresponding to the saddle point (E.7).[32]

We can ask: why don't we include all poles in the t-channel?. If we were to include all poles in the t-channel, we would obtain the wrong answer. The reason it is wrong is that in string theory the argument that we can shift the contour is not correct because of contributions for large values of q. Such large values of q were never meant to be included in the integral, since the kinematics of the process we consider restricts the real values of $\vec{q}^2$ to be much smaller than s. In deriving the physical picture, we certainly assumed that

[30] The location of this pole in the q_1 plane depends on $\vec{q}_{\rm rest}$ which we take to be real.

[31] For a closed string with $N_L = N_R = n$, the maximum spin is $J = 2 + 2n$.

[32] Here we are assuming that both b^2 and $\log s\alpha'$ are large with a ratio larger than one but fixed, so that we can neglect the $1/(n!)^2$ in this discussion. This last factor makes the sum converge for large n, but this convergent answer, as discussed below, is not the correct one.

JHEP02(2016)020

$\vec{q}^2 \ll s$.[33] Indeed, if we look at (E.4) we get a very small contribution from large real values of $\vec{q}^2$. On the other hand, if we were to keep s fixed and we formally look at large real values of $\vec{q}^2$ in the original expression, (E.3) , then we would encounter the u channel poles. The conclusion is that shifting the contour for the q_1 integral, while it can be done formally, it does not represent the real physical computation we want to do. Approximating the integrand using (E.4) , and then integrating gives the physically correct answer.

One can qualitatively say that, for large $b^2/(\alpha' \log s\alpha')$, we get a contribution of some of the t-channel poles, as in figure 12, and then the rest of the poles are completely resummed via the saddle point integral in figure 12. Their contribution should be better thought of as coming from the creation of extended objects in the s-channel.

Another remark we want to make is the following. It was shown in [34] that the plane wave solution is a solution to all orders in the α' expansion. This gravitational plane wave encodes the contribution from the $t = 0$ pole. However, we have seen that sometimes we get a subleading contribution from the other poles, due to massive states along the t-channel. These mean that the physical scattering process contains extra contributions not captured by the plane wave.[34]

In string theory we can also take into account the tidal excitations. In this case the phase shift can be viewed as an operator that maps the two initial gravitons to two final generic string states. Refs. [51–55] have shown that this operator has the remarkably simple expression $\hat{\delta} \propto \int d\sigma d\sigma' \delta_{\rm grav}(\hat{X}_L(\sigma) - \hat{X}_R(\sigma'))$ where X_L and X_R are the transverse space positions of the string on the worldsheet and $\delta_{\rm grav}(b)$ is the ordinary gravity phase shift. This is valid for distances $b^2 \gg \alpha' \log(s\alpha')$. The effects of these tidal excitations do not help in resolving the causality issues discussed here and are unrelated to the appearance of closed strings in the s-channel discussed above. See [51–55] for further discussion.

F Properties of the AdS shock wave

Let us first examine the problem of higher derivative corrections for the AdS shock wave. As in the case of flat space the shock wave at hand continue to be an exact solution when arbitrary higher derivative corrections are included.

The argument for this is identical to the one in section 5 of [61]. The vector $l_\mu = \partial_\mu u = \{1, 0, 0, 0 \ldots\}$ in coordinates u, v, y_i, z. We can now compute the vector V_μ that the argument talks about. We find $V_\mu = \{0, \cdots, 0, -2/z\}$. In other words $V_z = -2/z$ and the rest of the components are zero. This obeys that $V_\mu l^\mu = 0$ as required in their argument.

In addition, one can also compute the substracted Riemann tensor

$$\check{R}_{\mu\nu\alpha\beta} = R_{\mu\nu\alpha\beta} + 2g_{\mu[\alpha}g_{\beta]\nu} \,.$$

It is indeed of the form stated in [61] with the symmetric tensor K given by

$$K_{y^i,y^j} = \frac{1}{2}(\delta_{ij}\partial_z h/z^3 - \partial_i\partial_j h/z^2) \,,$$

[33] Momentum conservation and on-shell conditions impose $q^2 < \frac{s}{4}$.

[34] This is not in contradiction with [34], since these extra terms can be viewed as a non-perturbative contribution in α'.

JHEP02(2016)020

$$K_{z,y^i} = K_{y^i,z} = -\frac{\partial_{y^i}\partial_z h}{z^2}\,, \qquad (F.1)$$
$$K_{zz} = \frac{1}{2}(\partial_z h/z^3 - \partial_z^2 h/z^2)\,,$$

with the rest of the components equal to zero. This K obeys that $K_{\mu\nu}l^\nu = 0$ as required by [61]. Notice that this form of K also leads to the equation of the form (5.3), when the Riemann tensor is used in Einstein's equations, namely $K^\mu_\mu = 0$. Once the former equation is imposed $\check{R}$ reduces to the Weyl tensor.

Using this shock wave we can once again compute the time delay for different theories. It is convenient to evaluate Riemann tensor in the coordinates that make rotation symmetry manifest. The result is

$$\hat{R}_{uiju}|_{\vec{y}=0;z=1} = K_{ij} = -f(u)\left(1-\rho^2\right)\frac{\varpi'(\rho)-\rho\varpi''(\rho)}{8\rho}\left(n^i n^j - \frac{1}{D-2}\delta^{ij}\right)\,, \qquad (F.2)$$

where $i = 1, \ldots, D-2$ so that we span $\vec{y}$ and z components. In the bulk of the paper we consider propagation of different perturbations in the shock wave background described above. We first write the most general form of the second order equations of motion and then compute the time delay in the high energy limit.

Let us consider several limits of the shock wave to make contact with previous investigations of similar type. First, we expect to recover the flat space shock wave for probes that come close enough to the center of the shock or, equivalently, for $\rho \to 0$. Indeed, it is easy to check that this asymptotic is correctly recovered (5.9).

Let us consider a couple of other limits. We can fix $\vec{y}_0$ and send $z_0 \to 0$ in which case we have a source at the boundary and the shock wave takes the form

$$h \sim \frac{z^{D-3}}{(z^2+|\vec{y}-\vec{y}_0|^2)^{D-2}}\,, \qquad (F.3)$$

which is exactly the shock wave considered in the energy correlator problem [14]. In the opposite limit $z_0 \to \infty$ we get

$$h \sim z^{D-3}\,, \qquad (F.4)$$

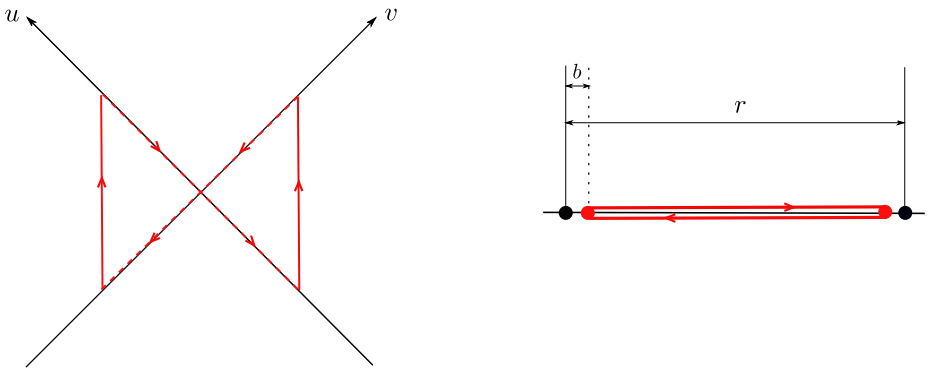
and the time delay in this background was computed, for example, in [62].

G Time advances and time machines

In this appendix we would like to argue that a negative time delay enables us to build a time machine which leads to closed time-like curves. This is a standard argument [88, 89] and the only thing we check is that the long range gravitational forces do not prevent us from setting it up. The setup we would like to consider is the following. We have two shock waves that correspond to energetic particles with momenta $q_{1,u} = \frac{\sqrt{s}}{2}$ and $q_{2,v} = \frac{\sqrt{s}}{2}$ separated by distance r in the transverse plane. The first shock is localized around $u = 0$ and the second one around $v = 0$. We would like the separation to be such that $r \gg r_S$, namely we are in the regime where the black hole formation does not occur

$$r^{D-3} \gg r_S^{D-3} = G\sqrt{s}\,. \qquad (G.1)$$

JHEP02(2016)020



a) b)

Figure 13. a) We imagine a background that consists of two shock waves located at $u = 0$ and $v = 0$ widely separated in the transverse directions which is not presented on the figure.The arrows show the motion of the probe massless particle projected on to the u, v plane. b) Same motion but projected on the transverse plane. The two background shocks are separated by r and the probe passes at a short distance b from each of them. The vertical region of the path in (a) corresponds to the horizontal motion in (b). We can build a closed time-like curve as depicted on the picture by crossing this pair of shocks if time advances are allowed. We need mirrors to reverse the motion in the transverse plane as we pass through the shocks.

Next we would like to consider a test particle that propagates through both shocks in such a way that it ends up at the same position where it started, thus, forming a closed time-like curve. The situation is depicted on figure 13. When the test particle crosses each of the shocks it gets shifted by $\Delta v = \Delta u \sim \frac{G\sqrt{s}}{b^{D-4}}$. Between the shock the particle travels the distance of order r. Thus, we want the time shift to be $\frac{G\sqrt{s}}{b^{D-4}} = \left(\frac{r_S}{b}\right)^{D-3} b \sim r \gg r_S$ which becomes

$$\left(\frac{r_S}{b}\right)^{D-4} \gg 1\,. \tag{G.2}$$

We also want $b \gg \widetilde{r}_S$ where $\widetilde{r}_S$ is the Schwarzschild radius for the shock wave-test particle pair, $\widetilde{r}_S^{D-3} \sim G\sqrt{p\sqrt{s}}$, where p is the energy of the probe particle. Together with (G.2) it implies $\sqrt{s} \gg p$ where p is the energy of the probe particle. We also need that $pb \gg 1$. The conclusion is that in $D > 4$ we can construct closed time-like curve using negative time delays by choosing b,r and s appropriately. For example, if we have a causality problem that appears at a scale $b^2 \sim \alpha_2$, then we take this value for b. Since we are at weak coupling, we know that the Planck length $l_p \ll b$. We can then pick $\sqrt{s} l_p \sim X^{1+a}\,, p\, l_p \sim X^{1-a'}$, with $X = (b/l_p)^{D-3}$ and we can choose $a > 0$, $a - a' < 0$, $1 - a' + 1/(D-3) > 0$ to ensure that $r_S \gg b$, $\widetilde{r}_S \ll b$ and $pb \gg 1$. We can achieve this with $a' = 1/2$ and $a = 1/4$, for example.

H Representations that couple to two gravitons

Here we would like to understand better what are the representations of little group SO($D-1$) of massive particles that can couple to two gravitons. We are interested only in bosonic fields since a single fermion does not couple to two gravitons.

JHEP02(2016)020

Let us consider the decay of a massive particle in its rest frame so that $p_2 = (M, \vec{0})$. Gravitons produced have $\vec{p}_1 = -\vec{p}_3 = \vec{p}$. We characterize the original particle by some polarization tensor $e_{i_1 \dots i_k}$ which has only spatial components and is traceless with respect to any pair of indices. We do not specify the symmetry properties of this tensor yet.

We characterize gravitons by polarization tensors e_1 and e_3 such that $\vec{e}_1.\vec{p} = \vec{e}_3.\vec{p} = 0$. We have three type of contractions (see also [79])

$$\begin{aligned} \mathcal{A}_1 &= e_{i_1 \dots i_k} p^{i_1} \dots p^{i_k} (e_1.e_3)^2 \,, \\ \mathcal{A}_2 &= e_{i_1 \dots i_k} e_1^{i_1} e_3^{i_2} p^{i_3} \dots p^{i_k} (e_1.e_3) \,, \\ \mathcal{A}_3 &= e_{i_1 \dots i_k} e_1^{i_1} e_1^{i_2} e_3^{i_3} e_3^{i_4} p^{i_5} \dots p^{i_k} \,. \end{aligned} \tag{H.1}$$

The first amplitude $\mathcal{A}_1$ exists only for particles in the symmetric traceless representations (Young tableau that consists of single horizontal row with k boxes). Actually all three amplitudes are allowed for symmetric representation and we discussed them in the bulk of the paper in detail.

For the second and third amplitude we can add more rows to the Young diagram. Properties of these, so-called, mixed-symmetry tensors are nicely reviewed, for example, in appendix E of [90]. By thinking about the representation in terms of tensors which are manifestly anti-symmetric with respect to indices in a given column we can read off possible representations. In particular the fact that we have only three different vectors means that we can have at most three rows.

Let us write all the amplitudes in a covariant manner. The general prescription is the following

$$e_1^i e_3^j \to E_{13}^{\mu\nu} \equiv \epsilon_1^\mu p_3^\nu (\epsilon_3.p_1) + \epsilon_3^\mu p_1^\nu (\epsilon_1.p_3) - p_1^\mu p_3^\nu (\epsilon_1.\epsilon_3) - \epsilon_1^\mu \epsilon_3^\nu (p_1.p_3) \,. \tag{H.2}$$

We then have for the amplitudes

$$\begin{aligned} \mathcal{A}_1 &= \epsilon_{\mu_1 \dots \mu_k} p_1^{\mu_1} \dots p_1^{\mu_k} \left[(\epsilon_1.\epsilon_3)(p_1.p_3) - (\epsilon_1.p_3)(\epsilon_3.p_1) \right]^2 \,, \\ \mathcal{A}_2 &= \epsilon_{\mu_1 \dots \mu_k} E_{13}^{\mu_1 \mu_2} p_1^{\mu_3} \dots p_1^{\mu_k} \left[(\epsilon_1.\epsilon_3)(p_1.p_3) - (\epsilon_1.p_3)(\epsilon_3.p_1) \right] \,, \\ \mathcal{A}_3 &= \epsilon_{\mu_1 \dots \mu_k} E_{13}^{\mu_1 \mu_3} E_{13}^{\mu_2 \mu_4} p_1^{\mu_5} \dots p_1^{\mu_k} \,. \end{aligned} \tag{H.3}$$

To compute the time delay we need to use the completeness relation

$$\sum_i \epsilon^*_{\nu_1 \dots \nu_k} \epsilon_{\mu_1 \dots \mu_k} = \Pi_{\nu_1 \dots \nu_k | \mu_1 \dots \mu_k} \,, \tag{H.4}$$

and contract both sides of the amplitude, where Π is a projector on to the space orthogonal to the intermediate momentum.

As the next step we would like to focus on those diagrams that produce s^a with $a \geq 2$ for the amplitude. Everything that is less is irrelevant for causality violation. In the language of Young tableaux it corresponds to having $a \geq 2$ boxes in the first row. It will be curious to understand if mixed symmetry fields with $a = 2$ can resolve the causality problem we observed in the bulk of the paper. As an example of such field would be $(2, 2)$ (a square with four boxes) or $(2, 2, 2)$ fields (a vertical rectangle of 2×3 boxes). There is a short list of representations of this kind.

JHEP02(2016)020

One could wonder whether these particles alone (without the infinite tower of higher spin particles) could solve the causality problem. We think that the answer is no. We leave a full exploration of this question for the future.

I The Weinberg-Witten theorem in D dimensions

The Weinberg-Witten theorem [50] constrains the matrix elements of a conserved stress tensor on single graviton states. They showed that the stress tensor cannot be Lorentz invariant, conserved, and such that it measures the energy of the gravitons. Their original argument is for $D = 4$. Here we extend it to arbitrary number of spacetime dimensions D. The matrix elements of a conserved stress tensor on single graviton states, $\langle \epsilon', p'|T_{\mu\nu}(q)|p, \epsilon\rangle$, has the Lorentz transformation properties of an on shell three point amplitude between two massless gravitons and a massive spin two particle. The "mass" is the same as the square of the four momentum of the stress tensor operator, q^2.[35]

The generic three point amplitude for a massive spin two and two massless gravitons has the form

$$\begin{aligned}\mathcal{A}_{gg\widetilde{g}} = \widetilde{c}_2\widetilde{\epsilon}_{\mu\nu}\,&[\epsilon_1^\mu p_3^\nu(\epsilon_3.p_1) + \epsilon_3^\mu p_1^\nu(\epsilon_1.p_3) - p_1^\mu p_3^\nu(\epsilon_1.\epsilon_3) - \epsilon_1^\mu\epsilon_3^\nu(p_1.p_3)]\\ &[(\epsilon_1.\epsilon_3)(p_1.p_3) - (\epsilon_3.p_1)(\epsilon_1.p_3)]]\\ +\widetilde{c}_4\widetilde{\epsilon}_{\mu\nu}p_1^\mu p_3^\nu\,&[(\epsilon_1.p_3)(\epsilon_3.p_1) - (\epsilon_1.\epsilon_3)(p_1.p_3)]^2\,,\end{aligned} \tag{I.1}$$

where the indices one and three denote the massless gravitons.

We are interested in the matrix element that corresponds to the measurement of particle's energy

$$\lim_{p'\to p}\langle \epsilon', p'|T_{\mu\nu}|p, \epsilon\rangle = p_\mu p_\nu \epsilon'.\epsilon\,. \tag{I.2}$$

We approach the point $p' = p$ from the space-like direction. In other words, the transferred momentum is space-like $q^2 = (p' - p)^2 > 0$.

Let us now consider $p_1 = p = (p_u, 0, \vec{0})$ and $q = (0, 0, \vec{q})$ so that $p_3 = p' = (p_u, \frac{\vec{q}^2}{4p_u}, \vec{q})$. We also consider the polarization tensor of the graviton to be purely transverse e_{ij}. Then for the outgoing graviton we have $\epsilon' = (0, \frac{\vec{q}.\vec{\epsilon}'}{2p_u}, \vec{\epsilon}')$.

To measure the energy in the limit of zero transfer we should have

$$\lim_{p'\to p}\langle e', p'|T_{\mu\nu}|p, e\rangle = p_u p_u e'_{ij}e_{ij}\,. \tag{I.3}$$

Computing the three-point amplitude to leading order in p_u we get

$$\begin{aligned}\mathcal{A}_{gg\widetilde{g}} &= -\widetilde{c}_2(q^2)p_u p_u(\epsilon_1.\epsilon_3)\,[(\epsilon_1.\epsilon_3)(p_1.p_3) - (\epsilon_3.p_1)(\epsilon_1.p_3)]]\\ &\quad +\widetilde{c}_4(q^2)p_u p_u\,[(\epsilon_1.p_3)(\epsilon_3.p_1) - (\epsilon_1.\epsilon_3)(p_1.p_3)]^2\\ &= \frac{1}{2}\widetilde{c}_2(\vec{q}^2)\vec{q}^2 p_u p_u\left(e'_{ij}e_{ij} - e'_{ij}e_{ik}\frac{q^j q^k}{\vec{q}^2}\right)\\ &\quad +\frac{1}{4}\widetilde{c}_4(\vec{q}^2)\vec{q}^4 p_u p_u\left(e'_{ij}e_{ij} - 2e'_{ij}e_{ik}\frac{q^j q^k}{\vec{q}^2} + e'_{ij}e_{kl}\frac{q^i q^j q^k q^l}{\vec{q}^4}\right)\,.\end{aligned} \tag{I.4}$$

[35] When $q^2 \neq 0$, the transversality condition on the polarization tensor, $q^\mu\widetilde{\epsilon}_{\mu\nu} = 0$ corresponds to the conservation condition, $q^\mu T_{\mu\nu}(q) = 0$.

JHEP02(2016)020

It is clear that for any $\widetilde{c}_2$ and $\widetilde{c}_4$ we cannot get (I.3) due to the non-trivial dependence on polarization of the graviton. This shows that it is impossible to have a theory with a stress tensor and massless gravitons at the same time. This is the famous result of Weinberg and Witten.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] I.I. Shapiro, *Fourth test of general relativity*, *Phys. Rev. Lett.* **13** (1964) 789 [INSPIRE].

[2] C.M. Will, *The confrontation between general relativity and experiment*, *Living Rev. Rel.* **17** (2014) 4 [arXiv:1403.7377] [INSPIRE].

[3] G. 't Hooft, *Graviton dominance in ultrahigh-energy scattering*, *Phys. Lett.* **B 198** (1987) 61 [INSPIRE].

[4] Y. Aharonov, A. Komar and L. Susskind, *Superluminal behavior, causality and instability*, *Phys. Rev.* **182** (1969) 1400 [INSPIRE].

[5] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis and R. Rattazzi, *Causality, analyticity and an IR obstruction to UV completion*, *JHEP* **10** (2006) 014 [hep-th/0602178] [INSPIRE].

[6] J.M. Maldacena, *The large-N limit of superconformal field theories and supergravity*, *Int. J. Theor. Phys.* **38** (1999) 1113 [hep-th/9711200] [INSPIRE].

[7] E. Witten, *Anti-de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150] [INSPIRE].

[8] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Gauge theory correlators from noncritical string theory*, *Phys. Lett.* **B 428** (1998) 105 [hep-th/9802109] [INSPIRE].

[9] M.A. Vasiliev, *Nonlinear equations for symmetric massless higher spin fields in (A)dS$_d$*, *Phys. Lett.* **B 567** (2003) 139 [hep-th/0304049] [INSPIRE].

[10] X. Bekaert, S. Cnockaert, C. Iazeolla and M.A. Vasiliev, *Nonlinear higher spin theories in various dimensions*, hep-th/0503128 [INSPIRE].

[11] H. Osborn and A.C. Petkou, *Implications of conformal invariance in field theories for general dimensions*, *Annals Phys.* **231** (1994) 311 [hep-th/9307010] [INSPIRE].

[12] J. Erdmenger and H. Osborn, *Conserved currents and the energy momentum tensor in conformally invariant theories for general dimensions*, *Nucl. Phys.* **B 483** (1997) 431 [hep-th/9605009] [INSPIRE].

[13] G. Arutyunov and S. Frolov, *Three point Green function of the stress energy tensor in the AdS/CFT correspondence*, *Phys. Rev.* **D 60** (1999) 026004 [hep-th/9901121] [INSPIRE].

[14] D.M. Hofman and J. Maldacena, *Conformal collider physics: energy and charge correlations*, *JHEP* **05** (2008) 012 [arXiv:0803.1467] [INSPIRE].

[15] I. Heemskerk, J. Penedones, J. Polchinski and J. Sully, *Holography from conformal field theory*, *JHEP* **10** (2009) 079 [arXiv:0907.0151] [INSPIRE].

JHEP02(2016)020

[16] A.L. Fitzpatrick, E. Katz, D. Poland and D. Simmons-Duffin, *Effective conformal theory and the flat-space limit of AdS*, *JHEP* **07** (2011) 023 [arXiv:1007.2412] [INSPIRE].

[17] J.M. Maldacena, *Non-Gaussian features of primordial fluctuations in single field inflationary models*, *JHEP* **05** (2003) 013 [astro-ph/0210603] [INSPIRE].

[18] A. Strominger, *The dS/CFT correspondence*, *JHEP* **10** (2001) 034 [hep-th/0106113] [INSPIRE].

[19] E. Witten, *Quantum gravity in de Sitter space*, hep-th/0106109 [INSPIRE].

[20] J.M. Maldacena and G.L. Pimentel, *On graviton non-Gaussianities during inflation*, *JHEP* **09** (2011) 045 [arXiv:1104.2846] [INSPIRE].

[21] P. McFadden and K. Skenderis, *Cosmological 3-point correlators from holography*, *JCAP* **06** (2011) 030 [arXiv:1104.3894] [INSPIRE].

[22] P. Benincasa and F. Cachazo, *Consistency conditions on the S-matrix of massless particles*, arXiv:0705.4305 [INSPIRE].

[23] P.C. Schuster and N. Toro, *Constructing the tree-level Yang-Mills S-matrix using complex factorization*, *JHEP* **06** (2009) 079 [arXiv:0811.3207] [INSPIRE].

[24] D.A. McGady and L. Rodina, *Higher-spin massless S-matrices in four-dimensions*, *Phys. Rev.* **D 90** (2014) 084048 [arXiv:1311.2938] [INSPIRE].

[25] R. Penrose, *On Schwarzschild causality — a problem for "Lorentz covariant" general relativity*, in *Essays in general relativity*, F.J. Tipler ed., Elsevier The Netherlands (1980), pg. 1.

[26] S. Gao and R.M. Wald, *Theorems on gravitational time delay and related issues*, *Class. Quant. Grav.* **17** (2000) 4999 [gr-qc/0007021] [INSPIRE].

[27] G. Velo and D. Zwanziger, *Noncausality and other defects of interaction lagrangians for particles with spin one and higher*, *Phys. Rev.* **188** (1969) 2218 [INSPIRE].

[28] R.W. Fuller and J.A. Wheeler, *Causality and multiply connected space-time*, *Phys. Rev.* **128** (1962) 919 [INSPIRE].

[29] W. Israel, *Thermo field dynamics of black holes*, *Phys. Lett.* **A 57** (1976) 107 [INSPIRE].

[30] J.M. Maldacena, *Eternal black holes in anti-de Sitter*, *JHEP* **04** (2003) 021 [hep-th/0106112] [INSPIRE].

[31] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781 [arXiv:1306.0533] [INSPIRE].

[32] P.C. Aichelburg and R.U. Sexl, *On the gravitational field of a massless particle*, *Gen. Rel. Grav.* **2** (1971) 303 [INSPIRE].

[33] T. Dray and G. 't Hooft, *The gravitational shock wave of a massless particle*, *Nucl. Phys.* **B 253** (1985) 173 [INSPIRE].

[34] G.T. Horowitz and A.R. Steif, *Space-time singularities in string theory*, *Phys. Rev. Lett.* **64** (1990) 260 [INSPIRE].

[35] D.N. Kabat and M. Ortiz, *Eikonal quantum gravity and Planckian scattering*, *Nucl. Phys.* **B 388** (1992) 570 [hep-th/9203082] [INSPIRE].

[36] S.B. Giddings, *The gravitational S-matrix: Erice lectures*, *Subnucl. Ser.* **48** (2013) 93 [arXiv:1105.2036] [INSPIRE].

JHEP02(2016)020

[37] M. Ciafaloni and D. Colferai, *Rescattering corrections and self-consistent metric in Planckian scattering*, *JHEP* **10** (2014) 085 [arXiv:1406.6540] [INSPIRE].

[38] S. Stieberger, *Open & closed vs. pure open string disk amplitudes*, arXiv:0907.2211 [INSPIRE].

[39] H. Kawai, D.C. Lewellen and S.-H. Henry Tye, *A relation between tree amplitudes of closed and open strings*, *Nucl. Phys.* **B 269** (1986) 1 [INSPIRE].

[40] B. Zwiebach, *Curvature squared terms and string theories*, *Phys. Lett.* **B 156** (1985) 315 [INSPIRE].

[41] A. Buchel, J. Escobedo, R.C. Myers, M.F. Paulos, A. Sinha and M. Smolkin, *Holographic GB gravity in arbitrary dimensions*, *JHEP* **03** (2010) 111 [arXiv:0911.4257] [INSPIRE].

[42] T. Jacobson and R.C. Myers, *Black hole entropy and higher curvature interactions*, *Phys. Rev. Lett.* **70** (1993) 3684 [hep-th/9305016] [INSPIRE].

[43] T. Liko, *Topological deformation of isolated horizons*, *Phys. Rev.* **D 77** (2008) 064004 [arXiv:0705.1518] [INSPIRE].

[44] S. Sarkar and A.C. Wall, *Second law violations in Lovelock gravity for black hole mergers*, *Phys. Rev.* **D 83** (2011) 124048 [arXiv:1011.4988] [INSPIRE].

[45] R. Britto, F. Cachazo, B. Feng and E. Witten, *Direct proof of tree-level recursion relation in Yang-Mills theory*, *Phys. Rev. Lett.* **94** (2005) 181602 [hep-th/0501052] [INSPIRE].

[46] M.T. Grisaru, H.N. Pendleton and P. van Nieuwenhuizen, *Supergravity and the S-matrix*, *Phys. Rev.* **D 15** (1977) 996 [INSPIRE].

[47] D.M. Eardley and S.B. Giddings, *Classical black hole production in high-energy collisions*, *Phys. Rev.* **D 66** (2002) 044011 [gr-qc/0201034] [INSPIRE].

[48] V.S. Rychkov, *Black hole production in particle collisions and higher curvature gravity*, *Phys. Rev.* **D 70** (2004) 044003 [hep-ph/0401116] [INSPIRE].

[49] S.B. Giddings and V.S. Rychkov, *Black holes from colliding wavepackets*, *Phys. Rev.* **D 70** (2004) 104026 [hep-th/0409131] [INSPIRE].

[50] S. Weinberg and E. Witten, *Limits on massless particles*, *Phys. Lett.* **B 96** (1980) 59 [INSPIRE].

[51] D. Amati, M. Ciafaloni and G. Veneziano, *Superstring collisions at Planckian energies*, *Phys. Lett.* **B 197** (1987) 81 [INSPIRE].

[52] D. Amati, M. Ciafaloni and G. Veneziano, *Classical and quantum gravity effects from Planckian energy superstring collisions*, *Int. J. Mod. Phys.* **A 3** (1988) 1615 [INSPIRE].

[53] D. Amati, M. Ciafaloni and G. Veneziano, *Can space-time be probed below the string size?*, *Phys. Lett.* **B 216** (1989) 41 [INSPIRE].

[54] D. Amati, M. Ciafaloni and G. Veneziano, *Planckian scattering beyond the semiclassical approximation*, *Phys. Lett.* **B 289** (1992) 87 [INSPIRE].

[55] D. Amati, M. Ciafaloni and G. Veneziano, *Effective action and all order gravitational eikonal at Planckian energies*, *Nucl. Phys.* **B 403** (1993) 707 [INSPIRE].

[56] L. Susskind, *Structure of hadrons implied by duality*, *Phys. Rev.* **D 1** (1970) 1182 [INSPIRE].

[57] D.N. Page, S. Surya and E. Woolgar, *Positive mass from holographic causality*, *Phys. Rev. Lett.* **89** (2002) 121301 [hep-th/0204198] [INSPIRE].

JHEP02(2016)020

[58] L. Cornalba, M.S. Costa, J. Penedones and R. Schiappa, *Eikonal approximation in AdS/CFT: from shock waves to four-point functions*, *JHEP* **08** (2007) 019 [hep-th/0611122] [INSPIRE].

[59] M. Brigante, H. Liu, R.C. Myers, S. Shenker and S. Yaida, *The viscosity bound and causality violation*, *Phys. Rev. Lett.* **100** (2008) 191601 [arXiv:0802.3318] [INSPIRE].

[60] M. Hotta and M. Tanaka, *Shock wave geometry with nonvanishing cosmological constant*, *Class. Quant. Grav.* **10** (1993) 307 [INSPIRE].

[61] G.T. Horowitz and N. Itzhaki, *Black holes, shock waves and causality in the AdS/CFT correspondence*, *JHEP* **02** (1999) 010 [hep-th/9901012] [INSPIRE].

[62] D.M. Hofman, *Higher derivative gravity, causality and positivity of energy in a UV complete QFT*, *Nucl. Phys.* **B 823** (2009) 174 [arXiv:0907.1625] [INSPIRE].

[63] L. Cornalba, M.S. Costa, J. Penedones and R. Schiappa, *Eikonal approximation in AdS/CFT: conformal partial waves and finite N four-point functions*, *Nucl. Phys.* **B 767** (2007) 327 [hep-th/0611123] [INSPIRE].

[64] L. Cornalba, M.S. Costa and J. Penedones, *Eikonal approximation in AdS/CFT: resumming the gravitational loop expansion*, *JHEP* **09** (2007) 037 [arXiv:0707.0120] [INSPIRE].

[65] R.C. Myers, M.F. Paulos and A. Sinha, *Holographic studies of quasi-topological gravity*, *JHEP* **08** (2010) 035 [arXiv:1004.2055] [INSPIRE].

[66] K. Sen and A. Sinha, *Holographic stress tensor at finite coupling*, *JHEP* **07** (2014) 098 [arXiv:1405.7862] [INSPIRE].

[67] L. Di Pietro and Z. Komargodski, *Cardy formulae for SUSY theories in $d = 4$ and $d = 6$*, *JHEP* **12** (2014) 031 [arXiv:1407.6061] [INSPIRE].

[68] L.F. Alday and J.M. Maldacena, *Comments on operators with large spin*, *JHEP* **11** (2007) 019 [arXiv:0708.0672] [INSPIRE].

[69] A.L. Fitzpatrick, J. Kaplan, D. Poland and D. Simmons-Duffin, *The analytic bootstrap and AdS superhorizon locality*, *JHEP* **12** (2013) 004 [arXiv:1212.3616] [INSPIRE].

[70] Z. Komargodski and A. Zhiboedov, *Convexity and liberation at large spin*, *JHEP* **11** (2013) 140 [arXiv:1212.4103] [INSPIRE].

[71] O. Nachtmann, *Positivity constraints for anomalous dimensions*, *Nucl. Phys.* **B 63** (1973) 237 [INSPIRE].

[72] G.J. Galloway, K. Schleich, D. Witt and E. Woolgar, *The AdS/CFT correspondence conjecture and topological censorship*, *Phys. Lett.* **B 505** (2001) 255 [hep-th/9912119] [INSPIRE].

[73] S.H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067 [arXiv:1306.0622] [INSPIRE].

[74] PLANCK collaboration, P.A.R. Ade et al., *Planck* 2013 *results. XXII. Constraints on inflation*, *Astron. Astrophys.* **571** (2014) A22 [arXiv:1303.5082] [INSPIRE].

[75] D. Anninos, T. Hartman and A. Strominger, *Higher spin realization of the dS/CFT correspondence*, arXiv:1108.5735 [INSPIRE].

[76] BICEP2 collaboration, P.A.R. Ade et al., *Detection of B-mode polarization at degree angular scales by BICEP2*, *Phys. Rev. Lett.* **112** (2014) 241101 [arXiv:1403.3985] [INSPIRE].

JHEP02(2016)020

[77] M.J. Mortonson and U. Seljak, *A joint analysis of Planck and BICEP2 B-modes including dust polarization uncertainty*, *JCAP* **10** (2014) 035 [arXiv:1405.5857] [INSPIRE].

[78] R. Flauger, J.C. Hill and D.N. Spergel, *Toward an understanding of foreground emission in the BICEP2 region*, *JCAP* **08** (2014) 039 [arXiv:1405.7351] [INSPIRE].

[79] M.S. Costa, J. Penedones, D. Poland and S. Rychkov, *Spinning conformal correlators*, *JHEP* **11** (2011) 071 [arXiv:1107.3554] [INSPIRE].

[80] X.O. Camanho and J.D. Edelstein, *Causality constraints in AdS/CFT from conformal collider physics and Gauss-Bonnet gravity*, *JHEP* **04** (2010) 007 [arXiv:0911.3160] [INSPIRE].

[81] 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** (1980) 343 [INSPIRE].

[82] F.A. Berends and R. Gastmans, *Quantum electrodynamical corrections to graviton-matter vertices*, *Annals Phys.* **98** (1976) 225 [INSPIRE].

[83] S. Dubovsky, A. Nicolis, E. Trincherini and G. Villadoro, *Microcausality in curved space-time*, *Phys. Rev.* **D 77** (2008) 084016 [arXiv:0709.1483] [INSPIRE].

[84] D.J. Gross and E. Witten, *Superstring modifications of Einstein's equations*, *Nucl. Phys.* **B 277** (1986) 1 [INSPIRE].

[85] R.R. Metsaev and A.A. Tseytlin, *Order α' (two loop) equivalence of the string equations of motion and the σ-model Weyl invariance conditions: dependence on the dilaton and the antisymmetric tensor*, *Nucl. Phys.* **B 293** (1987) 385 [INSPIRE].

[86] R.R. Metsaev and A.A. Tseytlin, *Curvature cubed terms in string theory effective actions*, *Phys. Lett.* **B 185** (1987) 52 [INSPIRE].

[87] H.J. de Vega and N.G. Sanchez, *Quantum string scattering in the Aichelburg-Sexl geometry*, *Nucl. Phys.* **B 317** (1989) 706 [INSPIRE].

[88] R.C. Tolman, *The theory of relativity of motion*, Berkeley Univ. Press, U.S.A. (1917).

[89] S. Coleman, *Acausality*, in *Erice* 1969, *Ettore Majorana School on Subnuclear Phenomena*, New York U.S.A. (1970), pg. 282 [INSPIRE].

[90] V.E. Didenko and E.D. Skvortsov, *Elements of Vasiliev theory*, arXiv:1401.2975 [INSPIRE].

JHEP02(2016)020