PHYSICAL REVIEW D **81,** 025002 (2010)

The gravitational S matrix

Steven B. Giddings* and Rafael A. Porto†
Department of Physics, University of California, Santa Barbara, California 93106, USA
and PH-TH, CERN, 1211 Geneve 23, Switzerland
(Received 22 September 2009; published 5 January 2010)

We investigate the hypothesized existence of an S matrix for gravity and some of its expected general properties. We first discuss basic questions regarding the existence of such a matrix, including those of infrared divergences and description of asymptotic states. Distinct scattering behavior occurs in the Born, eikonal, and strong gravity regimes, and we describe aspects of both the partial wave and momentum space amplitudes, and their analytic properties, from these regimes. Classically the strong gravity region would be dominated by the formation of black holes, and we assume its unitary quantum dynamics is described by corresponding resonances. Masslessness limits some powerful methods and results that apply to massive theories, though a continuation path implying crossing symmetry plausibly still exists. Physical properties of gravity suggest nonpolynomial amplitudes, although crossing and causality constrain (with modest assumptions) this nonpolynomial behavior, particularly requiring a polynomial bound in complex s at fixed physical momentum transfer. We explore the hypothesis that such behavior corresponds to a nonlocality intrinsic to gravity, but consistent with unitarity, analyticity, crossing, and causality.

DOI: 10.1103/PhysRevD.81.025002 PACS numbers: 11.55.Bq, 04.70.Dy, 11.80.Et

I. INTRODUCTION

In a quantum-gravitational theory where spacetime, locality, etc. may not be fundamental concepts, an important question is what quantities are amenable to quantitative analysis. In this paper, we will assume that flat space, or something which it closely approximates, is an allowed configuration of the theory. We will moreover assume that there is an action of its symmetry group, namely, the Poincare group, both on this configuration and on perturbations about it. This suggests that we can consider states incident from infinity, with given momenta and energies, and study their scattering. The resulting quantum amplitudes should be summarized in an S matrix.

One would like to understand what properties are expected of such an S matrix. For a quantum theory, unitarity is a given. Analyticity in momenta and crossing symmetry encode important physical features of S matrices in quantum field theory (QFT), like causality [1]. Gravity offers some new features whose role needs to be understood. Masslessness is first and causes infrared singularities; these we however envision regulating by working in spacetime dimension $D > 4$, or by proper formulation of inclusive amplitudes. Another is growth of the range of gravity with energy, as is seen, for example, in growth of the Schwarzschild radius of a black hole formed in a high-energy collision. An important question is how these new features can be reconciled with the others. One would also like to understand how these and other physical properties either do or do not manifest themselves in a gravitational S matrix—particularly locality and causality. The latter properties are especially interesting, given that a certain lack of locality could be part of a mechanism for information to escape a black hole, and thus explain the mysteries surrounding the information paradox. Yet locality is manifest in low-energy descriptions of nature and is a cornerstone of QFT; it is also nontrivially related to causality, which plays an important role in consistency of a theory.

In this paper, we carry out some preliminary investigation of these matters, with particular focus on the ultrahigh energy regime. We will make the maximal analyticity hypothesis [1], where one assumes that the only singularities that appear in the scattering amplitudes are those dictated by unitarity. Our investigations will then focus on the question of what can be learned by combining unitarity, analyticity, crossing, and causality together with expected general features of gravity. In spite of the plausibly nonlocal behavior of the gravitational amplitudes that we will explore, we have found no evidence of a lack of harmony between such nonlocality and these basic properties. We thus entertain the possibility that an S-matrix representation of such nonlocal dynamics exists, which retains the essential physical features.

The next section will further describe the S-matrix hypothesis, and some issues that must be confronted in its formulation, particularly questions of infrared divergences and asymptotic completeness, and summarize aspects of exclusive amplitudes and their partial wave expansion. Section III contains a summary of the different scattering regimes (broadly, Born, eikonal, and strong gravity), and aspects of the physics of each. Section IV focuses on the strong gravity regime, where one expects significant contributions from processes classically described as black hole formation. We parametrize the corresponding intermediate states as resonances, and investigate their implications for the form of the partial wave amplitudes. Section V further develops the description of these ampli-

*giddings@physics.ucsb.edu
†rporto@physics.ucsb.edu

tudes, summarizing our knowledge of the contributions to the phase shifts and their imaginary parts from the different regimes. Section VI overviews some properties of amplitudes in momentum space, some of which can be inferred from those of partial wave amplitudes. In particular, for both forms of amplitudes, we find strong indications of nonpolynomial behavior. Section VII investigates aspects of analyticity and crossing; the latter is less transparent than in a theory with a mass gap. Nonetheless, there is an argument for crossing, and this together with causality (plus Hermitian analyticity and a smoothness assumption) in turn leads to constraints on nonpolynomial growth. Section VIII closes with further discussion of nonpolynomiality, and its connection with the question of locality of the theory.

The study of ultra-Planckian collisions in gravity has a long history. In string theory, this includes [2–6], and other prominent early references are [7–10]. An important question is whether string theory resolves the puzzles of this regime. In particular, the information paradox suggests a breakdown of locality in this context; while string theory is apparently nonlocal due to string extendedness, it has been argued [11,12] that this effect does not appear to enter in a central way in the regimes of interest. In fact, the strong gravitational regime, where classically black holes form, apparently corresponds to a breakdown of the gravitational loop expansion. Reference [13] has argued for a possible resummation of string amplitudes that continues into this regime, but we view the apparent need for nonlocal mechanics as well as the absence of clearly relevant stringy effects as suggesting that a new ingredient is instead required for a fundamental description of this regime [11]. Though a perturbative string description appears insufficient for a complete description, it has been argued that nonperturbative dual formulations such as AdS/CFT [14] will address these problems. While there has been some progress toward extracting a flat space gravitational S matrix from AdS/CFT [15–19], some puzzles remain [17,20] about whether this is possible; one expects similar issues in matrix theory [21]. Whether or not it is, we take a more general viewpoint, extending the work of [22]: whatever theory provides this S matrix, we would like to characterize its features, and some of those may be rather special in order to describe gravity. Moreover, it may be that, as suggested in [23], the need to describe such features is in fact a critical clue to the dynamics of a quantum theory of gravity.

II. THE HYPOTHESIS OF THE GRAVITATIONAL S MATRIX

It is natural to expect that the problem of high-energy gravitational scattering in asymptotically flat space can be properly formulated in terms of the S matrix. Here, however, one must grapple with some preliminary issues.

The first issue is that we do not know a precise description of the quantum numbers of these states. For example, they could be states of string theory, some other completion of supergravity, or some other theory of gravity. However, in any case, we expect that the asymptotic states include those corresponding to widely separated individual incident particles, e.g. electrons, neutrinos, etc., in order to match our familiar description of nature. Or, if the theory were string theory, incident states are string states. We might have states with other quantum numbers as well. An example of the latter that is sometimes useful to consider is scattering in Minkowski space that is reached by compactification from higher dimensions; there, one may have incident particles or strings with conserved Kaluza-Klein charge. In any of these cases, a nice feature of gravity is that it universally couples to all energy, so we view it as plausible that some important features of gravitational scattering, particularly at high energy, are independent of this detailed description of the asymptotic states.

A second issue is that, in a perturbative description of gravitons propagating in flat space, gravity suffers from infrared divergences in four dimensions, arising from soft gravitons, and as a consequence one must generalize from the S matrix to inclusive amplitudes. While it does not seem inconceivable that this is of fundamental importance, we will assume that it is not. One reason for this is that QED suffers a similar problem, with the simple resolution through inclusive generalization of the S matrix, summing over soft photon states. Moreover, we note that this problem is not present if one works with higher-dimensional gravity. Specifically, for spacetime dimension $D > 4$, soft graviton divergences are not present. (For $D \geq 7$, the total cross section is finite.) We have already motivated considering higher-dimensional theories, by including the possibilities of string theory or supergravity, or we may simply think of this as dimensional regularization—in any case, to avoid this issue we will typically work in $D > 4$.

Another issue that plausibly comes closer to being fundamental regards the question of asymptotic completeness of states. The asymptotic completeness condition[1] states that the Hilbert space of the theory is equivalent to a Fock space of asymptotic free particles. However, there are apparent limitations to such a Fock space description. An example is the locality bound [23,25,26] and its N-particle generalizations [11]. Specifically, if one considers two particles in wave packets, which we, for example, can take to be Gaussian with central positions and momenta x, y and p, q, these have a field theory description in terms of a Fock space state $\phi_{x,p}\phi_{y,q}|0\rangle$. However, such a description must break down when we violate the bound

$$|x - y|^{D-3} > G|p + q|, \tag{2.1}$$

where $G \sim G_D$, the D-dimensional Newton constant. In

[1] See, e.g., Chapter 7 of [24].

this regime, gravity becomes strong, and so limits a Fock space description of the system; this limitation in principle extends to arbitrarily large distances. One may yet be able to construct an asymptotic description of all states in terms of free-particle states, using further evolution—if one evolves a state violating (2.1) backward in time, it generically ceases violating the bound, and would be expected to resolve itself into well-separated free particles asymptotic from infinity. Thus, with such a limiting procedure, and a weak form of local Lorentz invariance (in order to describe asymptotic particles with relative boosts), one plausibly describes asymptotics in terms of *certain* Fock space states.

In short, we will hypothesize the existence of a gravitational S matrix, or its inclusive generalization in $D = 4$. While we do not have a complete description of the asymptotic states, we will assume that they include states closely approximating particles that are initially widely separated, and moreover are allowed to have very large relative momenta. This starting point amounts to making certain assumptions about a weak notion of locality (asymptotically separated particles) and local Lorentz invariance (large relative boosts allowed for widely separated particles). However, we will *not* necessarily assume that stronger forms of locality and local Lorentz invariance are fundamental in the theory.

For practical purposes, it is often convenient to imagine that the asymptotic states correspond to spinless particles of mass m, plus gravitons. With such a collection of asymptotic states $|\alpha\rangle_{\rm in}$, $|\alpha\rangle_{\rm out}$ (taken to be Heisenberg-picture states), we expect an S matrix of the form

$$S_{\alpha\beta} = {}_{\rm out}\langle\alpha|\beta\rangle_{\rm in} = \langle\alpha|S|\beta\rangle. \tag{2.2}$$

As usual, we separate off the nontrivial part as $S = 1 + i\mathcal{T}$.

A. Exclusive amplitudes

Much of this paper's discussion will focus on the simplest nontrivial amplitude of the theory—that for exclusive $2 \to 2$ scattering. Here, the transition matrix element T (in the plane-wave limit) is then defined by

$$\begin{aligned}\langle p_3, p_4|\mathcal{T}|p_1, p_2\rangle = \mathcal{T}_{p_3p_4,p_1p_2} &= (2\pi)^D\delta^D(p_1 + p_2 - p_3 - p_4) \\ &\quad\times T(s, t),\end{aligned} \tag{2.3}$$

and is a function of the Mandelstam parameters

$$s = -(p_1 + p_2)^2 = E^2, \qquad t = -(p_1 - p_3)^2, \qquad u = -(p_1 - p_4)^2. \tag{2.4}$$

We expect that important features of the theory are encoded in this amplitude and its analyticity properties. Since the graviton is massless, amplitudes are singular at $t = 0$, and likewise in other channels; for example, the Born approximation to t-channel exchange gives

$$T_{\rm tree}(s, t) = -8\pi G_D s^2/t. \tag{2.5}$$

We will consider other aspects of analyticity in Sec. VII.

B. Partial wave expansion

Unitarity and some other physical features of the amplitude are most clearly formulated by working with the D-dimensional partial wave expansion, which is [7]

$$T(s, t) = \psi_\lambda s^{2-D/2}\sum_{l=0}^{\infty}(l + \lambda)C_l^\lambda(\cos\theta)f_l(s). \tag{2.6}$$

Here $\lambda = (D - 3)/2$,

$$\psi_\lambda = 2^{4\lambda+3}\pi^\lambda\Gamma(\lambda), \tag{2.7}$$

and C_l^λ are Gegenbauer polynomials, with arguments given by the center-of-mass (c.m.) scattering angle,

$$\cos\theta = 1 + \frac{2t}{s - 4m^2}. \tag{2.8}$$

Note that

$$t = (4m^2 - s)\sin^2(\theta/2), \qquad u = (4m^2 - s)\cos^2(\theta/2). \tag{2.9}$$

The inverse relationship to (2.6) gives the partial wave coefficients $f_l(s)$ in terms of the matrix element,

$$\begin{aligned}f_l(s) = \frac{s^{(D-4)/2}}{\gamma_D C_l^\lambda(1)}\int_0^\pi d\theta \sin^{D-3}\theta C_l^\lambda(\cos\theta) \\ \times T[s, (4m^2 - s)\sin^2(\theta/2)],\end{aligned} \tag{2.10}$$

with

$$\gamma_D = 2\Gamma\left(\frac{D-2}{2}\right)(16\pi)^{(D-2)/2}. \tag{2.11}$$

The unitarity condition

$${\rm Im}\, f_l(s) \geq |f_l(s)|^2, \tag{2.12}$$

for real $s \geq 0$ can be solved in terms of two real parameters, the phase shift $\delta_l(s)$, and the absorptive coefficients $\beta_l(s) \geq 0$:

$$f_l(s) = \frac{i}{2}[1 - e^{2i\delta_l(s) - 2\beta_l(s)}]. \tag{2.13}$$

It is important to understand the convergence properties of the partial wave expansion (2.6). For a theory with a mass gap, the expansion can be shown to converge in the Lehmann ellipse [27], which extends into the unphysical regime $t > 0$, $\cos\theta > 1$. This extension is useful for further constraining amplitudes, e.g. through the Froissart-Martin [28,29] bound.

Masslessness of gravity alters this behavior. Let us first ask when the partial wave coefficients (2.10) are well defined. Specifically, at long distance/small angle, we have the Born approximation, (2.5). This gives a pole at zero angle, $T \sim 1/\theta^2$, and correspondingly the integral (2.10) only converges for $D > 4$. While other long-distance effects, like soft graviton emission, could modify the amplitude (2.5), we do not expect them to alter this convergence behavior.

In general, a series of the form (2.6) converges in an ellipse with foci at $\cos\theta = \pm 1$. The existence of the singularity in T at $\theta = 0$ indicates that the partial wave expansion does not converge past $\cos\theta = 1$. Thus, the Lehmann ellipse has collapsed into a line segment along the real axis. Note that one does expect $\mathrm{Im}T(\theta = 0)$ to be finite for $D \geq 7$. This follows from the optical theorem (see the Appendix); as we have noted, the Born cross section given by (2.5) is not infrared divergent for $D \geq 7$. However, this finiteness does not indicate that the expansion of $\mathrm{Im}T$ can be continued past this point—higher derivatives of $\mathrm{Im}T$ are expected to in general diverge at $\theta = 0$.

The failure of convergence of the partial wave expansion in the regime $t > 0$ is an impediment to using some of the powerful methods that have been successfully applied in theories with a mass gap. Nonetheless, we suggest that the study of partial wave amplitudes can still be useful for inferring features of scattering. While we are, in particular, interested in features of the analytic continuation of $T(s, t)$ to complex values of s and t, where convergence of the expansion is problematic, we can exploit the *inverse* relation (2.10). Regardless of the convergence of the partial wave expansion, we have argued that (2.10) is convergent for $D \geq 5$. Thus, if physical considerations imply statements about the behavior of $f_l(s)$, these in turn imply properties of the integrand of (2.10), and specifically of $T(s, t)$.

III. SCATTERING REGIMES

In different regions of s and t, or E and l, we expect differing physical behaviors of amplitudes. A more pictorial way to think of these different regimes is as a function of energy and impact parameter b of the collision—these are after all often variables controlled experimentally. While the transformation to impact parameter representation suffers from some complexities, our main focus will be on collisions in the ultrahigh-energy limit, $E \gg M_D$, where $M_D^{D-2} = (2\pi)^{D-4}/(8\pi G_D)$ gives the D-dimensional Planck mass. There, for many purposes, we expect the classical relation

$$l \sim Eb/2, \tag{3.1}$$

which should approximately hold more generally, to serve as a useful guide to the physics, though we expect precise statements to be more easily made in terms of the conserved quantities E and l.

Figure 1 illustrates some of the regimes that we expect to be relevant for ultrahigh-energy scattering, in terms of energy and impact parameter. We will particularly focus on the Born regime, the eikonal regime, and the strong gravity, or "black hole" regime.

A. Born and eikonal

The best-understood regime is the Born regime, corresponding to large impact parameters/small angles. Here, the elastic scattering amplitude, corresponding to single graviton exchange, has been given in (2.5); one may also consider corrections due to soft graviton emission [4,22,30].

As the impact parameter decreases, or the energy increases, diagrams involving exchange of more gravitons become important. The leading contributions at large impact parameter are the ladder and crossed ladder diagrams, which can be summed to give the eikonal approximation to the amplitude [2,3,9,31,32].[2] This can be written in terms of the *eikonal phase*, which arises from a Fourier transformation converting the tree-level amplitude into a function of a variable naturally identified as the impact parameter:

$$\chi(x_\perp, s) = \frac{1}{2s}\int \frac{d^{D-2}q_\perp}{(2\pi)^{D-2}} e^{-i\mathbf{q}_\perp\cdot\mathbf{x}_\perp} T_{\rm tree}(s, -q_\perp^2) = \frac{4\pi}{(D-4)\Omega_{D-3}}\frac{G_D s}{x_\perp^{D-4}}, \tag{3.2}$$

where $q_\perp$ is the transverse momentum transfer and where

$$\Omega_n = \frac{2\pi^{(n+1)/2}}{\Gamma[(n+1)/2]} \tag{3.3}$$

is the volume of the unit n sphere. The eikonal approximation to the amplitude is then

$$iT_{\rm eik}(s, t) = 2s\int d^{D-2}x_\perp e^{-i\mathbf{q}_\perp\cdot\mathbf{x}_\perp}(e^{i\chi(x_\perp, s)} - 1), \tag{3.4}$$

expressing the amplitude in an impact parameter form. From (3.4), one sees where eikonal corrections to the Born amplitude become important, namely, when the eikonal phase χ becomes of order 1. Indeed, [22] showed that at the corresponding point via (3.1), the partial wave phase shifts become of order unity, and thus the eikonal amplitudes unitarize the amplitudes of the Born approximation. (Contributions due to soft graviton emission were also estimated in [22].) In terms of impact parameter, this transition region is given by

$$b \sim (G_D E^2)^{1/(D-4)}, \tag{3.5}$$

as is illustrated in Fig. 1. It is alternatively described as the region where the momentum transfer is of order of the inverse impact parameter,

$$\sqrt{-t} \sim \frac{1}{b}. \tag{3.6}$$

[2]One may inquire about UV divergences of loop diagrams. However, these are short distance effects, for which we assume there is some UV regulation; for example, string theory might serve this purpose, or even supergravity, if it is perturbatively finite [33].

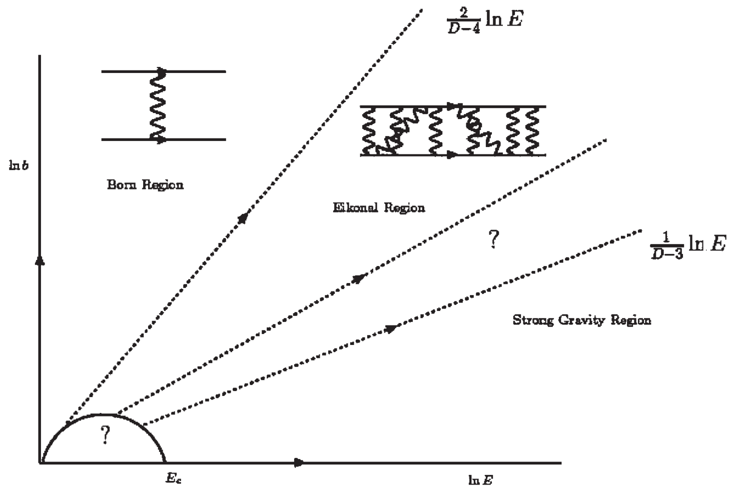

FIG. 1. Scattering regimes in an impact parameter picture; the question marks denote possible model dependence discussed in Sec. III C.

In general, eikonal approxmations are expected to capture semiclassical physics. In the high-energy gravitational context, the semiclassical geometry is the collision of two Aichelburg-Sexl shock waves, and various evidence supports the correspondence between (3.4) and this picture [2–5,8]. In particular, the saddle point of (3.4) gives a classical scattering angle

$$\theta_c \sim \frac{1}{E}\frac{\partial}{\partial b}\chi \sim \left[\frac{R(E)}{b}\right]^{D-3}, \tag{3.7}$$

matching that of a test particle scattering in the Aichelburg-Sexl geometry. Here, we have introduced the Schwarzschild radius corresponding to the c.m. energy,

$$R(E) = \frac{1}{M_D}\left(\frac{k_D E}{M_D}\right)^{1/(D-3)}, \tag{3.8}$$

where

$$k_D = \frac{2(2\pi)^{D-4}}{(D-2)\Omega_{D-2}}. \tag{3.9}$$

One finds [5] that corrections to the ladder series become important when $\sqrt{-t} \sim E$, or alternatively when the scattering angle reaches $\theta \sim 1$. Equation (3.7) shows that this happens at the impact parameter comparable to the Schwarzschild radius, $b \sim R(E)$, as pictured in Fig. 1. A schematic argument for this follows from power counting. Consider a diagram arising from a graviton tree attached to the external lines. Each graviton vertex gives a factor $\sqrt{G_D}$. Those connecting to external lines are accompanied by a $\sqrt{s}$. The remaining dimensions come from internal (loop) momenta. For the processes in question, these have characteristic value[3] $k \sim 1/b$. This counting then produces a power series in $(R/b)^{D-3}$. A leading such correction, the H diagram, which has been discussed in [4,5], is illustrated in Fig. 2. One can alternatively understand this expansion by thinking of the external lines as classical sources; using standard power-counting techniques [34], one can easily show that the H diagram is $\mathcal{O}[(G_D E)^2/r^{2(D-3)}]$ compared to one graviton exchange, if the distance between the sources is r [35]. Using $G_D E \sim R^{D-3}$ and taking $r \sim b$ then yields the same expansion parameter. In terms of the semiclassical geometry, at impact parameters $b \sim R$, one forms a trapped surface [36,37], and hence a black hole.

B. Strong gravity

Since corrections to the eikonal amplitudes give terms that differ from the eikonal amplitudes by powers of $[R(E)/b]^{D-3}$, the region where a classical black hole forms apparently corresponds to a manifest breakdown of the perturbative expansion; it is not even asymptotic. We can also parametrize this in terms of a critical angular momentum, given by

$$l \sim L(E) = ER(E)/2. \tag{3.10}$$

One might be tempted to believe that a quantum treatment of the evolution can still be given by performing an expansion in fluctuations about a shifted background—that of the semiclassical black hole. However, the problem of the singularity guarantees this is not a complete description. Moreover, even evolution on spatial "nice slices" that avoid the singularity is problematical, given that a standard field theory treatment of it leads to the information paradox.[4] This suggests that the boundary of this regime represents a correspondence boundary, analogous to that, for example, between classical and quantum mechanics, beyond which local quantum field theory does not give a

[3]Indeed, in the eikonal regime, the dominant term in the exponential series of (3.4) occurs at order $N \sim G_D s/b^{D-4}$, corresponding to a characteristic momentum $k \sim \sqrt{-t}/N \sim 1/b$ in each internal line of the $N-1$-loop Feynman ladder diagram.

[4]For reviews, see [38,39].

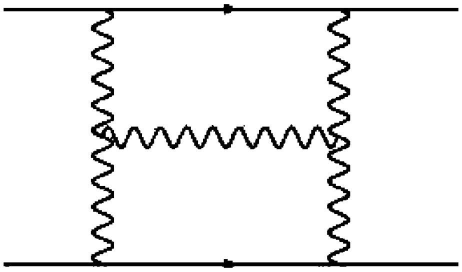

FIG. 2. The H diagram, which provides a leading correction to the eikonal amplitudes as scattering angles approach $\theta \sim 1$.

complete description of the dynamics [40]. In particular, the unitary evolution which we are assuming, in which the quantum information must escape the black hole while it is still comparable to its original radius [41], suggests that the nonperturbative dynamics unitarizing the physics is not local with respect to the semiclassical geometry—a sort of "nonlocality principle" [23,26]. [This then fits with the proposed parametrization of part of the correspondence boundary given by the locality bound [11,25,26]: namely, local field theory fails for multiparticle states whose wave functions are concentrated inside a radius of size $R(E)$, where E is their combined c.m. energy.]

While we do not have the means to calculate quantum amplitudes in this regime,[5] we can infer some of their properties if we believe that the semiclassical picture of formation of a black hole and its subsequent evaporation provides a good approximate description of the physics when addressing certain coarse-grained questions. Specifically, Ref. [22] parametrized certain features of the corresponding S matrix, and we will improve on the corresponding "black hole ansatz" in subsequent sections.

Of course, investigating the internal dynamics seen, e.g. by observers falling into black holes, and reconciling that with outside observations such as described by an S matrix, remains a challenging problem. Reference [40] has argued for flaws in the "nice-slice argument" for information loss, of two origins. First, attempts to measure the nice-slice state at a level of precision appropriate to investigate information loss lead to large backreaction on the state. Secondly, fluctuations e.g. in the Hawking radiation are argued to lead to fluctuations in the nice-slice state after long times. We expect that sharper investigations should follow from use of protolocal observables [43], but ultimately the full nonperturbative dynamics of gravity is plausibly necessary in order to give both a complete picture of infalling observers and of reconciliation of their observations with a unitary S matrix.

C. Other regimes

Before turning to a further description of the strong gravity regime, it is important to note that at impact parameters larger than $b \sim R(E)$, other features of the dynamics can become relevant. Indeed, some have argued that this indicates other dynamics besides strong gravity is a dominant feature of high-energy scattering. To give an example, in the context of string theory, with string mass M_{st}, it is possible to make long strings with length $l \sim E/M_{st}^2$. In fact, such processes are highly suppressed, but Amati *et al.* [2] pointed out that such amplitudes receive other important string corrections through "diffractive excitation" beginning at impact parameters of size $b_t \sim M_D^{-1}(E/M_{st})^{2/(D-2)}$. Indeed, Veneziano [44] proposed that this effect may provide important corrections to a picture of black hole formation; if true, this would likely obscure a strong gravity interpretation of the regime $b \lesssim R(E)$.

References [11,12] investigated these effects more closely. Indeed, as pointed out in [11], a simple picture of the origin of these effects is string excitation arising from the tidal impulse of the gravitational field of the other colliding string. Moreover, Ref. [12] investigated the evolution of the corresponding string states. For impact parameters $b_t \gg b \gg R(E)$, the asymptotic state of the string is indeed highly excited as a result of this tidal string deformation. However, for impact parameters $b \lesssim R(E)$, the time scales of horizon formation and string excitation differ significantly. Roughly, in a semiclassical picture the trapped surface forms before the tidal excitation causes significant extension of the string. Thus, one seemingly produces a configuration described as a pair of excited strings inside a black hole; in this context there is no clear reason to believe that string extendedness would lead to significant modification of the black hole description of the dynamics. Likewise, there is not a clear mechanism for string effects to provide the necessary nonlocality with respect to the semiclassical picture, to allow information escape.

Indeed, one can imagine a similar dynamics being relevant for collisions of other composite objects—hydrogen atoms, protons, etc. Specifically, when tidal forces reach a size sufficient to excite the internal degrees of freedom of the object, asymptotic states will be excited states. Thus, there can be model-dependent tidal-excitation effects. However, once impact parameters reach the regime $b \lesssim R(E)$ (and for sufficiently large E), such effects are not expected to prevent black hole formation. Since these model-dependent tidal-excitation effects do not appear to

[5]Reference [13] suggested analytic continuation of the perturbative sum giving the amplitude in the region $b > R$. However, one might at best expect such a sum to approximately reconstruct the semiclassical geometry, as in [42]. Then, in particular, it is not clear how the resulting prescription would give unitary amplitudes that escape the usual reasoning behind the information paradox, which as we have summarized, apparently requires new dynamical ingredients. Indeed, this paper elaborates on the view that local QFT cannot fully capture the physics of the strong gravitational regime semiclassically associated with black hole formation.

contribute fundamental features to the story, we will largely ignore them in the following discussion.

Another regime that has been of much interest in string theory discussions is that near the string energy, $E \sim M_{st}$, where one might expect to initially see weakly coupled string excitations. This region lies in the lower left corner of Fig. 1. One expects such excitations to merge into black holes at a ''correspondence point'' [45] where $R(E_c) \sim 1/M_{st}$. Our focus will be on higher energies.

IV. THE STRONG GRAVITATIONAL REGIME

We currently lack a complete quantum description of the strong gravitational regime. However, we will assume that the quantum description of this regime must be compatible with certain features following from a semiclassical picture of black hole formation. If one accepts such a viewpoint, and moreover assumes that the microphysical evolution is unitary, these combined assumptions potentially provide interesting constraints on the dynamics—particularly in view of the preceding statements that unitary evolution is apparently incompatible with evolution that is local with respect to the semiclassical geometry.

A. Black hole formation

We begin by recalling basic features of black hole formation in a high-energy collision, which has been extensively studied as a phenomenological feature of models with a low Planck scale [46,47].[6]

Consider a high-energy collision of two particles, with c.m. energy $E \gg M_D$. Let us moreover assume that the wave functions of these particles are Gaussian wave packets with characteristic size Δx, and that these collide with an impact parameter $b \lesssim R(E)$; for large E, we may take $\Delta x \ll R(E)$.

In the classical description of this process, a trapped surface will form in the geometry [36,49], signaling formation of a black hole, and as a result of the small curvatures, one expects a corresponding statement in a semiclassical approximation to the quantum dynamics [37]. Not all of the collision energy is trapped in the black hole, which is initially rather asymmetrical, and radiation (soft gravitons, gauge fields, etc.) will escape to infinity during the ''balding'' process in which it settles down to a Kerr black hole[7] of mass M. The time scale for balding is of order $\tau_{\rm form} \sim R(E)$, and for impact parameters sufficiently below $R(E)$, the amount of energy lost is an $\mathcal{O}(1)$ fraction, but not large (e.g. $\lesssim 40\%$), thus $M \approx E$.

Subsequently, the black hole will radiate, initially preferentially radiating states that lower its spin. The characteristic energy of radiated particles is the Hawking temperature, $T_H \sim 1/R(M)$, and roughly one quantum is emitted per time $\tau \sim R(M)$.

B. Black holes as resonances

We will thus think of the black holes that form after $\tau_{\rm form}$ as resonances [22]. Since the width for such a state to decay (typically into a lower-energy black hole) is $\Gamma(M) \sim 1/R(M)$, this is a limit to the sharpness with which we can define the energy of the black hole. However, black holes with $M \gg M_D$ are sharp resonances in the sense that

$$\frac{\Gamma}{M} \sim \frac{1}{RM} \sim \frac{1}{S(M)} \ll 1, \tag{4.1}$$

where $S(M)$ is the Bekenstein-Hawking entropy.

We will assume that the number of possible black hole resonances is given by this entropy. To be more precise, let us assume that the number of black hole microstates with energies in a range $(M, M + \Delta M)$ is

$$\Delta \mathcal{N}(M) = B(M) e^{S(M)} R(M) \Delta M, \tag{4.2}$$

where $B(M)$ is a possible prefactor that is dimensionless and is expected to have much more slowly varying energy dependence than the exponential. Thus the density of black hole states is of the form

$$\rho(M) = RBe^{S(M)}, \tag{4.3}$$

and the total number of states with energy $\leq M$ is $\mathcal{N}(M) \simeq B(M) \exp\{S(M)\}$. The spacing between the states is clearly much smaller than their widths. Let us label the states in the interval $(M, M + 1/R)$ as

$$|M; I\rangle, \tag{4.4}$$

where $I = 1, \ldots, \Delta\mathcal{N}(M) \sim \exp\{S(M)\}$. We may further refine the description to project on angular momentum eigenstates, with angular momenta l. In that case, the entropy entering the preceding formulas is expected to be

$$S(M, l) = \frac{4\pi E R(M, l)}{D - 2}, \tag{4.5}$$

where $R(E, l)$ is given by [50]

$$R^{D-5}\left[R^2 + \frac{(D-2)^2 l^2}{4M^2}\right] = \frac{16\pi G_D M}{(D-2)\Omega_{D-2}}. \tag{4.6}$$

For small l, this gives an expansion of the form

$$S(M, l) = S(M, 0)\left(1 - \text{const}\frac{l^2}{L^2}\right). \tag{4.7}$$

C. Black hole spectrum and evolution

Let us explore in more detail the quantum states formed in a collision, which could be either a two-particle collision with a c.m. energy $\approx E$, or an n-body collision. Note that one can also form a black hole of mass M by producing a

[6]For a review with some further references, see [48].

[7]In models with gauge charges not carried by light particles, the black hole can also carry charge.

higher-mass black hole in a collision with $E \gg M$, and then waiting for that black hole to evaporate to M.

Consider general initial multiparticle (but not black hole) states; these can be labeled by energy, momentum, generalized partial waves, and asymptotic species and spin content. Let us work in the c.m. frame, and ignore the effects of particle spin. Some subset of the states, denoted $|E; a\rangle_{\rm in}$, will form a black hole; examples are the two-particle states described above, which classically do so, and thus are expected to have probability essentially unity for black hole formation.

This means that a state[8] $|E; a\rangle_{\rm in}$ can be rewritten in terms of states that at a time just after formation correspond to a combined state of black hole and balding radiation; let us choose an orthonormal basis $|E'; i\rangle_{\rm rad}$ for the latter, and thus write

$$|E; a\rangle_{\rm in} = \sum_{M,I,i} \mathcal{A}(E, M)_{aIi} |M; I\rangle |E - M; i\rangle_{\rm rad}; \qquad (4.8)$$

here we neglect the possibility of a small component on states that are *not* black holes. In principle we can project on a definite state of the radiation, yielding a pure black hole state:

$$_{\rm rad}\langle E - M; i|E; a\rangle_{\rm in} = \sum_I \mathcal{A}(E, M)_{aIi} |M; I\rangle. \qquad (4.9)$$

In a generic black hole basis we expect the amplitudes $\mathcal{A}(E, M)_{aIi}$ to be of order $e^{-S(M)/2}$, corresponding to the fact that from (4.3) we expect there to be $\mathcal{O}(e^S)$ states. The space of states in (4.9) can be combined to form an orthonormal basis for a subspace of black hole states, denoted $|M; A\rangle$, and labeled by the initial and radiation state labels. However, this basis will not span the space of all black hole states, since (4.9) yields too few states. Indeed, note that there are arguments (extending [51]) that only of order

$$\exp\left\{E \frac{(D - 2)(D - 1)}{D(D - 3)}\right\} \qquad (4.10)$$

states can be formed from collapse of matter of energy E; thus a should have such a range. If one also accounts for the balding radiation, as above, there are more states that can be accessed through their entanglement with this radiation. Typical radiated quanta have energies $\sim 1/R$, and given the radiated energy $E - M$, this yields an entropy $\sim R(E)(E - M) \propto E^{(D-2)/(D-3)}$. This exponentiates to give the number of states over which the index i can range. However, this is still far fewer than the $\exp S(M)$ black hole states, since typically $M > E/2$. Thus, the number of states that are ‘‘accessible’’ in the collision at energy E is far less than the number of possible states of the black hole. We can label a basis for the remaining complementary black hole state space as $|M; \bar{A}\rangle$. One expects that one approach to accessing these states is to form a black hole of mass $M' > M$ in a higher energy collision, and then allow it to evaporate down to mass M. In doing so, internal states of the black hole become entangled with the state of the Hawking radiation, like in the preceding discussion of balding radiation.[9] For large enough M', this gives $e^{S(M)}$ independent accessible states. For many purposes, it is simplest to forget the balding radiation, which as we have explained does not appear to play a particular central role, and in a slight abuse of notation, think of the labels A as corresponding to the initial states from which the black hole formed.

We can likewise label the possible n-body out states, representing the final decay products of a given black hole, as $|E, a\rangle_{\rm out}$. In a similar spirit to the preceding discussion, we could choose a basis of black hole states labeled by this out-state description. Again, we expect the matrix elements between the preceding basis and this one to generically have size $\exp\{-S(M)/2\}$. Correspondingly, the amplitude for a given initial black hole state to decay into a given final state of the Hawking radiation will be of generic size

$$|_{\rm out}\langle M, a|M, I\rangle| \sim e^{-S(M)/2}. \qquad (4.11)$$

The quantum description of black holes as a decaying multistate system has analogies to other such systems, like $K0 - \overline{K0}$ mesons. In the assumed unitary dynamics, an initial black hole state $|M; I\rangle$ can mix both with other states with the same energy, and with states that are in the continuum, which consist of a lighter black hole together with radiated quanta. One might expect, via a Weisskopf-Wigner [52] approximation, that evolution in the Hilbert space of black hole states with mass $\sim M$ is governed by an effective Hamiltonian:

$$i \frac{d}{dt} |M; I\rangle = H |M; I\rangle. \qquad (4.12)$$

Though conceivably more general dynamics is needed,[10] this exhibits possible features of black hole evolution. Because of the decay, the Hamiltonian is not Hermitian in this subspace, and in general takes the form

$$H_{IJ} = M_{IJ} - \frac{i}{2} \Gamma_{IJ}, \qquad (4.13)$$

where M_{IJ} and Γ_{IJ} are Hermitian matrices. In general, these will not commute.

D. Exclusive processes

If one considers, in particular, an exclusive process with two-particle initial and final states $|p_1, p_2\rangle_{\rm in}$, $|p_3, p_4\rangle_{\rm out}$, such as pictured in Fig. 3, one thus expects the intermediate black hole states to contribute to the S matrix as

[8] A more careful treatment uses narrow wave packets.

[9] One can in principle ‘‘purify’’ such states by projection on definite states of the Hawking radiation, as with the preceding projection of balding radiation.

[10] In particular, we do not expect H to necessarily be a Hamiltonian constructed from a local Lagrangian.

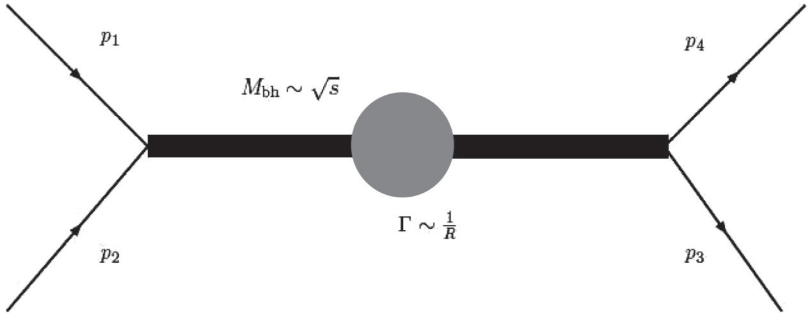

FIG. 3. Schematic of a black hole as a resonance in $2 \rightarrow 2$ scattering.

$$ {}_{\rm out}\langle p_3, p_4|p_1, p_2\rangle_{\rm in} = (2\pi)^D \delta^D\Big(\sum p\Big)\sum_{IJ}\langle p_3, p_4|J\rangle \times \left(\frac{1}{E-H}\right)_{JI}\langle I|p_1, p_2\rangle. \tag{4.14}$$

[Note that in the bases adapted to in or out states, described in the preceding section, the indices are expected to only range over $\sim S(E)$ values.] If M_{IJ} and Γ_{IJ} do not commute, H_{IJ} cannot be diagonalized by a unitary transformation, but we will assume it can be diagonalized by a more general linear transformation. The eigenstates $|M;\check{I}\rangle$ are then not orthogonal;

$$\langle M;\check{I}|M;\check{J}\rangle = g_{\check{I}\check{J}} \tag{4.15}$$

for some $g_{\check{I}\check{J}} \neq \delta_{\check{I}\check{J}}$. In such a basis (4.14) becomes[11]

$$ {}_{\rm out}\langle p_3, p_4|p_1, p_2\rangle_{\rm in} = (2\pi)^D \delta^D\Big(\sum p\Big)\sum_{\check{I}\check{J}}\langle p_3, p_4|\check{I}\rangle \times \frac{1}{E-H_{\check{I}}} g^{-1}_{\check{I}\check{J}}\langle\check{J}|p_1, p_2\rangle, \tag{4.16}$$

where $H_{\check{I}} = M_{\check{I}} - i\Gamma_{\check{I}}/2$ are eigenvalues. This will produce a sum of terms of Breit-Wigner form contributing to the amplitude. However, the sum itself will not, in general, take the Breit-Wigner form.

In the case where the particles being collided are the narrowly focused wave packets that we have described, one plausibly expects the corresponding amplitude to be of size

$$|{}_{\rm out}\langle a|b\rangle_{\rm in}| \sim e^{-S(E)/2}. \tag{4.17}$$

The reason for this is that for such wave packets the amplitude to form a black hole is essentially unity, and the amplitude for it to decay back to a two-particle state is of size given by (4.11). Note that our discussion suggests a resolution to questions raised [10] about the relation of intermediate black holes to Breit-Wigner behavior. One has $\mathcal{O}(1)$ amplitude to form *some* black hole state; in a generic basis for black hole states, this is a superposition with $\mathcal{O}(e^{-S/2})$ coefficients, although, as indicated in the preceding section, one can choose a special basis where black hole states are labeled by the initial states that created them. Thus, the amplitude to form a generic black hole state from a two-particle state is $\sim e^{-S/2}$, as is the amplitude for a generic black hole state to decay back into a two-particle state.

One might ask whether there could be any larger contributions to the $2 \rightarrow 2$ amplitude, due to processes that avoid black hole formation. For example, our Gaussian wave packets will have tails at large impact parameter. However, these have probability of size $\exp\{-(R/\Delta x)^2\}$ at $b \sim R$. The width Δx is constrained by $\Delta x > 1/E$, but this constraint produces a quantity merely of size $\gtrsim \exp\{-S^2\}$.

While we cannot at present rule out other such effects, none have been identified. Another test of this statement comes from scattering of a particle of high-energy E off a *preexisting* black hole in the relevant range $b \ll R$; here the amplitude $\mathcal{R}$ for reflection is also exponentially suppressed [53]:

$$R \sim e^{-4\pi ER}. \tag{4.18}$$

It is thus plausible that the amplitude for the *classically predicted* [36,37,49] black hole formation process only receives corrections that are exponentially suppressed at least to the level (4.17).

V. PARTIAL WAVE AMPLITUDES

In this section we restrict attention to $2 \rightarrow 2$ scattering, in a partial wave basis, and investigate consequences of the preceding picture and related considerations. For simplicity, we focus on scattering of one species of spinless particles. The initial two-particle states will be labeled by just their energy and angular momentum l, and the scattering amplitude is of form

$$\mathcal{S}_l(E) = e^{-2\beta_l(E)+2i\delta_l(E)}. \tag{5.1}$$

A. Strong gravitational regime

As outlined above, for impact parameters $b \ll R(E)$, or correspondingly angular momenta $l \ll L(E)$, the ampli-

[11]The form of this equation may alternately be simplified through the definition of a dual basis, $\langle\check{I}_d| = g^{-1}_{\check{I}\check{J}}\langle\check{J}|$.

tude for such a state to form a black hole with total angular momentum $l_{\rm BH} \approx l$ is expected to be of order unity.

1. Absorption

In the $2 \to 2$ process that goes through the black hole channel, $l_{\rm BH} = l$. From (4.11), we note that the amplitude for the given resonance $|E, l\rangle$ to decay *back* to a two-particle state is $\sim e^{-S(E,l)/2}$.

As in the preceding section, it is plausible that processes avoiding black hole formation in the regime $l \ll L$ are exponentially suppressed at least to this level. Arguments for that build on the preceding ones, together with the properties of partial-wave wave packets.

For example, consider a wave packet with definite angular momentum in the relative coordinates between the two particles:

$$\psi_{lm}(x) = \int dE \frac{J_{l+\lambda}(Er)}{(Er)^\lambda} e^{-iEt} Y_{lm}(\Omega) f(E), \tag{5.2}$$

where $J_{l+\lambda}$ is a Bessel function, $Y_{lm}(\Omega)$ are $D-2$ dimensional spherical harmonics, and $f(E)$ is a Gaussian wave packet with width ΔE. Asymptotics of Bessel functions for large order and argument (see Eq. 8.41.4 of [54]) then show that for $l \ll Er$,

$$J_{l+\lambda}(Er) \to \sqrt{\frac{2}{\pi Er}} \cos\left[Er - \frac{\pi(l+\lambda)}{2} - \frac{\pi}{4}\right], \tag{5.3}$$

with subleading corrections consisting of terms suppressed by powers of $(l+\lambda)/Er$ times sine or cosine functions of the same form. Thus combining (5.2) and (5.3) gives a wave packet that is Gaussian in $t \pm r$ with width $\Delta r \sim 1/\Delta E$, and subleading terms are similarly Gaussian.

A related argument comes from the relation between the partial wave representation and impact parameter representation [55]. Specifically, if $f(b, s)$ is the amplitude in impact parameter representation, then at high energies one finds the corresponding partial wave amplitude [56,57]

$$f_l(s) = f(2(l+\lambda)/E, s) + \frac{A}{s} \frac{d^2 f(b,s)}{db^2}\bigg|_{b=(2(l+\lambda))/E} + \cdots, \tag{5.4}$$

where A is a numerical coefficient, indicating that in the high-energy limit, localization in angular momentum corresponds to localization in impact parameter, as expected.[12]

A final argument comes from the behavior of partial waves scattering from a preexisting black hole; [53] argues that their reflection amplitude in the limit $ER \gg 1$ is of size (4.18).

Based on these, and on the discussion of Sec. IV, we thus conjecture that in the regime $l \ll L(E)$, the $2 \to 2$ amplitude is indeed exponentially small in the entropy, and arises mainly due to such a strong gravity channel. These statements suggest additional rationale for the black hole ansatz of [22], where in this regime

$$|\mathcal{S}_l(E)| = e^{-2\beta_l} \sim \exp\{-S(E,l)/2\}. \tag{5.5}$$

Notice that this behavior has *two* characteristic features. The first is the exponential *strength* of the absorption. The second is the long *range* of the absorption, which is characterized by the growth of $L(E)$ with energy. Even should the preceding arguments regarding the strength of the exponential suppression be evaded, we expect the feature of significant absorption at long range to persist.

2. Phase shifts

We have suggested that the amplitude is essentially unity for a given initial two-particle state with $l \ll L(E)$ to enter the strong gravitational regime. In $2 \to 2$ scattering, one might therefore expect that in each energy range $(E, E + 1/R)$ we form one of the $\Delta\mathcal{N}(E, l)$ black hole states[13] with the corresponding energy and angular momentum. This would correspond to a density of accessible states

$$\rho_{\rm acc}(E, l) \approx R(E). \tag{5.6}$$

(This value would be less relevant for $2 \to N$ scattering, where, as we have argued, more states may be accessible and entangle with the balding radiation.) Notice that this would imply that the *total* number of such accessible black hole states of angular momentum l and energy $<E$ is given by

$$\mathcal{N}_{\rm acc}(E) = \int_0^E \rho_{\rm acc}(E, l) dE \approx S(E, l). \tag{5.7}$$

Consider the parametrization (4.14) of the contributions of intermediate black hole states. If the matrix H_{IJ} were diagonal in the "in"-state basis $|M; A\rangle$, discussed in Sec. IV, then we would expect a contribution to the amplitude of Breit-Wigner form:

$$e^{2i\delta_l(E)} \approx e^{2i\delta_b}\left(1 - \frac{i\Gamma}{E - E_{\rm BH} + i\Gamma/2}\right), \tag{5.8}$$

where δ_b is a "background" value. Then, the phase $\delta_l(E)$ would increase by π as we pass through each such accessible (or strongly coupled) resonance, and correspondingly, the combined effect of resonances at increasing energies would give

$$\delta_l^{\rm diag}(E) = \pi\mathcal{N}_{\rm acc}(E, l) \approx \pi S(E, l), \tag{5.9}$$

as with Levinson's theorem for single-channel scattering. Note also that such a result would yield a decay time $d\delta_l/dE \sim R(E)$, compatible with the width $\Gamma \sim 1/R$.

[12] The series (5.4) may be regulated by considering incoming wave packets instead of plane waves.

[13] As noted, this state is a superposition of states of a generic basis with coefficients of size $\mathcal{O}(e^{-S/2})$.

However, we see no reason to expect H_{IJ} to be diagonal, and so consider phase shifts of a more general form, which we parametrize as

$$\delta_l(E) = \pi k(E, l) S(E, l), \tag{5.10}$$

where $k(E, l)$ varies more weakly with energy than $S(E, l)$. One might expect $k(E, l) > 0$ (corresponding to time delay) due to the attractive nature of gravity. Indeed, in scattering off a preexisting black hole the gravitational field introduces a positive phase shift relative to scattering from the angular momentum barrier. We will investigate additional constraints on $k(E, l)$ in subsequent sections.

To summarize, combining (5.9) and (5.10) suggests that the partial wave amplitudes in the strong gravity regime take the form

$$f_l^{\rm SG}(s) \approx \frac{i}{2}\left\{1 - \exp\left(-\frac{1}{2} S(E, l)[1 - 4\pi i k(E, l)]\right)\right\}. \tag{5.11}$$

Notice that this expression differs from that of [22]; that analysis did not take into account the role of inelasticity and accessibility of resonance channels. Thus (5.11) comprises an improvement of the black hole ansatz of [22].

B. Born and eikonal

One can likewise infer properties of the partial waves in the longer-distance regimes, where the Born or eikonal approximations are expected to be valid. In particular, Ref. [22] computed the eikonal phase shift,

$$\delta_l^{\rm eik}(E) = \frac{\sqrt{\pi}(D-2)\Gamma[(D-4)/2]}{8\Gamma[(D-1)/2]} \frac{L(E)^{D-3}}{l^{D-4}} \sim \frac{E^{D-2}}{l^{D-4}}, \tag{5.12}$$

and checked that the eikonal amplitude unitarizes the Born amplitude, which is the leading term in an expansion in δ_l, as expected. Thus the transition from Born to eikonal regimes occurs in the small-angle regime $l \sim E^{(D-2)/(D-4)}$. Notice that the phase shifts are indeed positive definite, as expected from the attractive nature of gravity.[14] The correspondence between the eikonal amplitudes and the semiclassical picture [2–5,8] suggests the utility of the eikonal description until $l \sim L$.

For decreasing impact parameter/increasing scattering angle, different effects can contribute to absorption. A generic effect is soft graviton bremsstrahlung. This was estimated in [22] to give a contribution of size

$$\beta_l^{\rm br} \sim L(E)^{3D-9}/l^{3D-10} \sim \frac{E^{3D-6}}{l^{3D-10}}. \tag{5.13}$$

[14] This is the case provided $D > 4$. The four-dimensional case suffers from Coulomb-like singularities, requiring the usual *inclusive* amplitudes, avoided in this paper by working in higher dimensions.

Note that this matches onto the energy dependence of (5.5) at $l \sim L$, which also fits with a picture where a non-negligible fraction of the collision energy can be emitted in the balding radiation.

As noted in Sec. III, there may be other less-generic effects, e.g. due to excitation of internal degrees of freedom of the colliding bodies. In string theory, such an effect is the diffractive excitation or "tidal string excitation" explored in [2–5,11,12]. But, as noted, we do not expect such effects to prevent amplitudes from matching onto those of the strong gravitational regime.

C. Combined pictures

We can thus suggest combined pictures describing the weak and strong coupling regimes. The results (5.5) and (5.13) suggest energy and angular momentum dependences of the absorptive coefficients β_l as pictured in Figs. 4 and 5.

While the phase shift is well studied in the eikonal regime, as we have indicated, we have less information in the strong gravity regime, but expect an increase

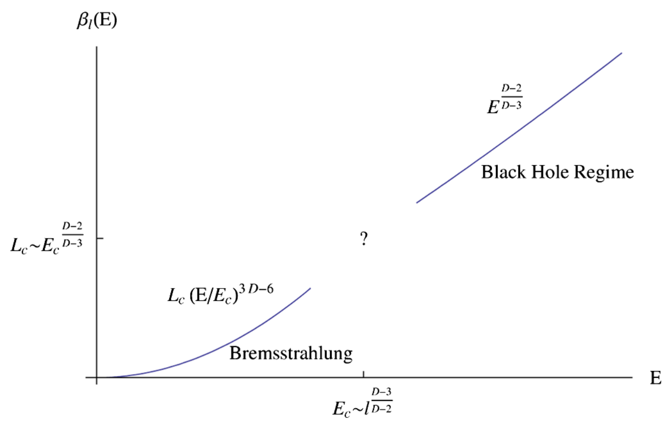

FIG. 4 (color online). Absorption coefficients at a fixed angular momentum as a function of the c.m. energy.

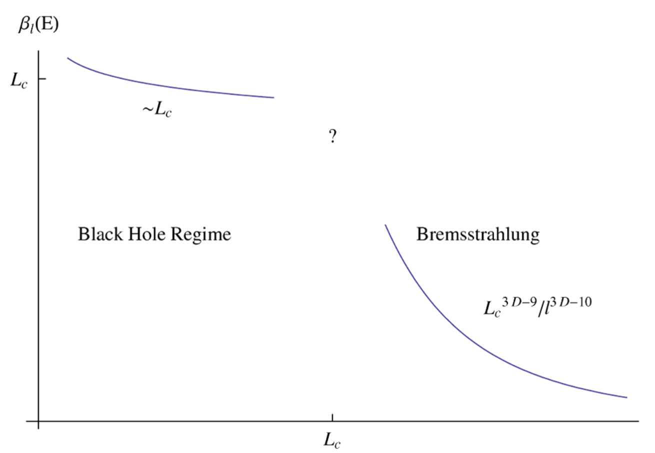

FIG. 5 (color online). Absorption coefficients at a fixed c.m. energy as a function of angular momentum, with $L_c \equiv L(E)$.

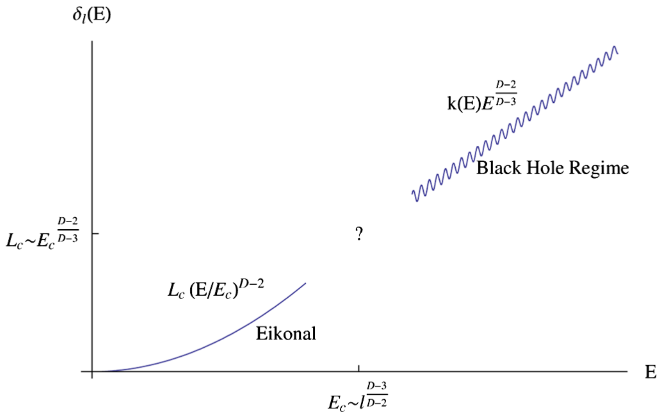

FIG. 6 (color online). Phase shift for fixed angular momentum as a function of the c.m. energy.

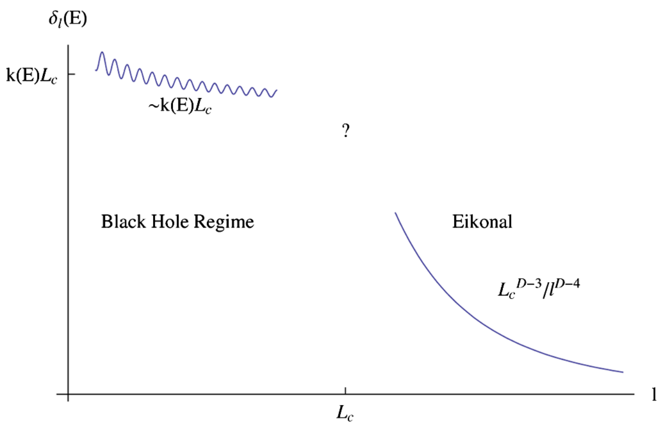

FIG. 7 (color online). Phase shift for a fixed c.m. energy as a function of angular momentum, with $L_c \equiv L(E)$.

bounded by $\delta_l(E) \sim E^{(D-2)/(D-3)}$ as in (5.10). Sketches of energy and angular momentum dependence are given in Figs. 6 and 7.

VI. MOMENTUM SPACE AMPLITUDES

We now ask what properties of momentum space amplitudes can be inferred from the preceding discussion. In Sec. II, we noted the collapse of the Lehmann ellipse, and, in particular, that convergence of the partial wave expansion cannot extend past $t = 0$ to positive t. Likewise, continuation of s to complex values with fixed real $t < 0$ would correspond to complex $\cos\theta$, outside the convergence region. These and related limitations restrict our ability to prove results that follow in massive theories. However, we have argued that the expression for the partial wave coefficients, (2.10), is expected to be well defined and finite. This means that properties of the $f_l(s)$ are those of the corresponding integral, and this in turn constrains the behavior of $T(s, t)$.

Additional information about the momentum space amplitudes comes directly from their eikonal approximation, (3.4). At very small angles, this expression reduces to the Born amplitude, (2.5). The match between the Born and eikonal regimes occurs near $\chi \sim 1$, corresponding to $t \sim -s^{-2/(D-4)}$ or

$$\theta_{B/E} \sim \frac{1}{E^{(D-2)/(D-4)}}. \tag{6.1}$$

The asymptotics of the eikonal amplitude at larger angles follows from performing the integral over angles in (3.4), which yields

$$\begin{aligned} iT_{\rm eik}(s,t) = {}& -2is(2\pi)^{(D-2)/2} q_\perp^{-(D-4)/2} \\ &\times \int_0^\infty dx_\perp x_\perp^{(D-2)/2} J_{(D-4)/2}(q_\perp x_\perp) \\ &\times (e^{i\chi(x_\perp,s)} - 1). \end{aligned} \tag{6.2}$$

Then, combining the Bessel function asymptotics (5.3) with a saddle-point approximation of the integral gives an asymptotic amplitude of the form

$$T_{\rm eik} \sim \exp\{i[s(-t)^{(D-4)/2}]^{1/(D-3)}\}. \tag{6.3}$$

This exhibits some interesting features—such as nonpolynomiality—that we will return to in the next section.

One may also inquire about implications for T of the strong gravity behavior outlined in the preceding section. Recall that the physical features of that behavior were (1) significant scattering, and moreover absorption, to an angular momentum that grows with energy as $l \sim L(E)$, (2) strong absorption for large E and $l \ll L(E)$, and (3) potentially rapid growth in the phase, (5.10).

For $\delta_l = \beta_l = 0$, (2.13) gives $f_l = 0$, so the first feature implies nonvanishing f_l to $l \sim L(E)$; significant absorption moreover implies that $f_l \sim i/2$. These become conditions on the integral

$$\int_0^\pi d\theta \sin^{D-3}\theta \frac{C_l^\lambda(\cos\theta)}{C_l^\lambda(1)} T[s, t(s,\theta)] = \frac{\gamma_D f_l(s)}{s^{(D-4)/2}}, \tag{6.4}$$

where $t(s, \theta)$ is given by (2.9). However, a direct statement about T in the strong gravity regime $s \sim -t$ is not easily inferred from the significance of the right side of (6.4), since the integral, in particular, receives a contribution from the Born regime. For $\theta < \theta_{B/E}$ and $l < L$, one has $l\theta \ll 1$ and can use the small-angle approximation

$$C_l^\lambda(1 - \theta^2/2) \simeq C_l^\lambda(1)\left(1 - \frac{l(l+2\lambda)\theta^2}{2(2\lambda+1)}\right). \tag{6.5}$$

The Born contribution to (6.4) is thus of size

$$\int_0^{\theta_{B/E}} d\theta \theta^{D-3} \frac{E^2}{\theta^2} \sim \frac{1}{E^{D-4}}. \tag{6.6}$$

This shows that one expects a contribution to partial wave amplitudes from both the Born and eikonal regions that is significant at angular momenta $l \lesssim L(E)$.

Indeed, a related fact is that the cross section due to this small-angle scattering is expected to be large as compared to that of the strong gravity region,

$$\sigma_{\rm SG} \approx \pi [R(E)]^{D-2} \sim E^{(D-2)/(D-3)}. \tag{6.7}$$

For $D > 6$, where the small-angle contribution converges, it can be estimated using the impact parameter where Born and eikonal match, giving [2]

$$\sigma_{B/E} \sim E^{2(D-2)/(D-4)}. \tag{6.8}$$

Large growth of β_l and δ_l with energy implies that $f_l - i/2$, or df_l/ds, are small, and rapidly oscillating. Equation (6.4) thus indicates that $T(s, t)$ correspondingly has rapid falloff and oscillations. Moreover, we see that exponential falloff of $f_l - i/2$ would indicate precise cancellations between the contributions of $T(s, t(s, \theta))$ in the Born, eikonal, and strong gravity regimes; as we have discussed, physical aspects of the scattering such as the analogy with scattering from a fixed black hole suggest such falloff.

A sharper statement arises if one considers continuation of (5.11) into the complex s plane. This form for $f_l(s)$ suggests that generically it would grow exponentially somewhere in the complex s (or E) plane. In particular, for small enough k, one finds exponential growth in the s upper half plane (UHP) $0 < \arg s < \pi$: for constant k, this would occur for

$$k < \frac{1}{4\pi} \tan \frac{\pi}{2(D-3)}, \tag{6.9}$$

and likewise for the example of a decreasing power, $k \propto E^{-p}$. By (6.4), this corresponds to exponential, thus not polymomially bounded, growth in $T(s, t)$ for complex t. While with the specific functional form (5.11), a phase that is too small leads to growth that is not polynomially bounded, it is conceivable that a more complicated analytic structure of the exact amplitude avoids this conclusion.[15]

VII. ANALYTICITY AND CROSSING

We have investigated aspects of unitarity, particularly via the partial wave expansion; we now turn to analyticity and crossing.

Consider scattering of two massive particles of mass m coupled to gravity. We might imagine these to be an e^+e^- pair, although to avoid complications of spin we will treat the scalar case. Another specific context to contemplate, if in a string theory context, is scattering of a $D0 - \overline{D0}$ pair.

First, consider behavior for fixed real $t < 0$, as a function of s. The two-particle cut in the s channel begins at $s = 4m^2$. However, one can also have such a pair annihilate to two or more gravitons (in the absence of a net conserved charge), implying multiple cuts beginning at $s = 0$.[16] Likewise, there are multiple u-channel cuts beginning at $u = 0$. Given

$$s + t + u = 4m^2, \tag{7.1}$$

we find that the u-channel cuts, for fixed t, originate at

$$s = 4m^2 - t, \tag{7.2}$$

and are taken to extend along the negative s axis. Thus, these cuts overlap those from $s = 0$—there are branch cuts running all along the real s axis, with no gap between them, unlike the massive case. These features of massless theories weaken some of the constraints present in massive theories.

We likewise expect singular behavior at $t = 0$; we have noted the Coulomb pole there, but one might find a more general singularity (e.g. branch point) when higher-order processes are accounted for. As we have already described, this prevents the usual continuation along the real axis from $t < 0$ to $t > 0$ that is a useful tool in massive theories.

A. Crossing symmetry

For real $s_0 > 4m^2$, the physical amplitude with $s = s_0$, $t < 0$ is assumed to arise from the analytic function $T(s, t)$ with $s = s_0 + i\epsilon$ in the limit $\epsilon \to 0^+$. By the maximal analyticity hypothesis, T only has singularities dictated by unitarity, so it can be continued throughout the s UHP; likewise for fixed s, one can continue in t, avoiding singularities.

In a massive theory, at small $t < 0$, one can continue in s across the real axis, through the gap between the cuts. This allows one to define the amplitude for $s = s_1 - i\epsilon$, for large negative real s_1, which by (7.1) corresponds to u-channel kinematics. Crossing symmetry is the assumption that a single function $T(s, t, u)$, with variables satisfying (7.1), defines amplitudes in all channels through such continuation.

Clearly this specific continuation fails in the massless case, given the lack of a gap between the cuts. However, it appears possible to still obtain crossing, through use of a different path.

The BEG path

Such a path was given by Bros, Epstein, and Glaser (BEG) in [62], as follows. First, begin at large $s_0 > 0$, and hold $u = u_0 < 0$ fixed. One can continue through the

[15] Though, with added assumptions like Hermitian analyticity/dispersion relations, one may possibly generalize methods of [58,59] to show that the exponential falloff in (5.11) implies a lower bound on the phase, e.g. $\delta \gtrsim \log s$, given a polynomial bound in the UHP; also, certain analyticity assumptions together with this falloff might possibly be used to prove violation of polynomial bounds in some region, with methods like in [60,61]. We leave these for future investigation. (Notice that in QFT we do not expect such a strong absorptive behavior, thus polynomial boundedness is expected to lead to a phase bounded above by $\log s$.)

[16] One might also contemplate the possibility of worse behavior, e.g. $\sim e^{-1/s^p}$ for some p.

upper s plane to $e^{i\pi}s_0$. Here, t will approach the positive real axis with a $-i\epsilon$; we can denote this as the t^- channel. Next, beginning at this point, keep $s < 0$ fixed and continue $t \to e^{-i\pi}t$. This is analogous to the preceding continuation, and takes t^- to u^+—here the positive real u axis is approached from above. The combined path thus continues from the physical s channel s^+ to the physical u channel u^+, permitting crossing.[17]

B. Crossing and polynomial boundedness

Analyticity and crossing constrain possible nonpolynomial behavior, as we will now discuss; the reader may wish to refer to figure Fig. 8. This observation follows from the Phragmen-Lindelöf theorem: If an analytic function is bounded along two straight lines sustaining an angle $\frac{\pi}{\alpha}$, e.g. $|T(|s|)| < M$ on the lines, and if $T(s)$ grows at most like $e^{|s|^\beta}$ with $\beta < \alpha$ in any other direction, then in fact $T(s)$ is bounded by M in the whole sector sustained by the two lines.

Choose, for example, $\alpha = 1$. Let us assume that the amplitude is quite weakly bounded, $|T(s, t < 0)| < e^{|s|}$. Note that this bound is easily satisfied both by the eikonal behavior (6.3), and by behavior that could arise from growth of the strong gravity region, either from the large absorption coefficients $|\beta_l(s)| \sim |s|^{(D-2)/(2(D-3))} \ll |s|$, or the large range $R(E) \sim E^{1/(D-3)}$ which suggests behavior [22] (see the next section), $T(s, t < 0) \sim e^{R(E)\sqrt{t}}$. Therefore, by the theorem, if we had a nonpolynomial growth in the UHP that would also require a nonpolynomial growth in a straight line $i\epsilon$ above the real axis from $-\infty$ to $+\infty$.

The region $[0, +\infty)$ corresponds to the s-channel amplitude. However, properties of the Gegenbauer polynomials combined with the optical theorem (see the appendix) show that $\mathrm{Im}T(s, t < 0) < \mathrm{Im}T(s, 0) \sim s\sigma_T(s) < s^N$. (The polynomial bound at $t = 0$ is directly connected to existence of a forward dispersion relation [22], following from causality, to be discussed in the next section.) Moreover, we have the high-energy expression

$$\int_0^\pi d\theta \sin^{D-3}\theta |\mathrm{Re}T|^2 \propto \int d\Omega_{D-2} |\mathrm{Re}T|^2 < \int d\Omega_{D-2} |T|^2 \propto s^{3-D/2}\sigma_{2\to 2} < s^{3-D/2}\sigma_T, \qquad (7.3)$$

where proportionality is modulo numerical coefficients, and therefore the real part of the amplitude also must be polynomially bounded, provided it is sufficiently smooth. (Recall that in the strong gravitational regime the real part of the amplitude is indeed subdominant due to strong absorption.)

[17]Note that one must also include a small path segment from $(s, t, u) = (-s_0 + i\epsilon, 4m^2 - u_0 + s_0 - i\epsilon, u_0)$ to $(-s_0, 4m^2 - u_0 + s_0 - i\epsilon, u_0 + i\epsilon)$. We assume this is permitted by sufficient holomorphy in this neighborhood, as in [63], though more systematic investigation is conceivably warranted.

In massive theories, the $(-\infty, 0]$ region is related to the u-channel amplitude by complex conjugation.[18] This follows from the property of *Hermitian analyticity* or extended unitarity, which is the requirement $T(s^*, t^*) = T(s, t)^*$. Notice that this implies $f_l(s^*) = f_l(s)^*$ for the partial wave coefficients. If we work at negative values of transfer momentum, e.g. $t < 0$, Hermitian analyticity also connects the discontinuity across the cuts due to threshold singularities to the imaginary part of the amplitude by

$$\mathrm{Disc}\, T(s, t) = 2i\, \mathrm{Im}T(s + i\epsilon, t). \qquad (7.4)$$

With a mass gap, Hermitian analyticity follows from reality of the amplitude below threshold, along with the Schwarz reflection property. In massless theories the status of Hermitian analyticity remains unclear, although it seems to hold at any order in perturbation theory. If Hermitian analyticity holds in gravity, it thus also forbids nonpolynomial growth along $(-\infty, 0]$, and so by the above theorem, in the UHP of s.

A conservative conjecture is that gravity respects both crossing symmetry and Hermitian analyticity, and that amplitudes thus satisfy such a polynomial bound. We can check this in the asymptotics of the eikonal, (6.3), which does so for $D > 4$, as does the preceding strong gravity expression.

Nonpolynomiality of amplitudes is however generally expected to give unbounded behavior in other regions of s, t, and u. Indeed, one can directly see indications for such behavior given the partial wave coefficients (5.11). For example, if $k(E, l) \sim E^{-p}$ for some $p > 0$, then the strong gravity f_l's given by (5.11) will have polynomially unbounded behavior somewhere in the UHP $\mathrm{Im}(s) > 0$. Then, (6.4) implies that $T[s, t(s, \theta)]$ must likewise be unbounded. Notice, though, that this is for fixed θ rather than t; thus unboundedness at large $|s|e^{i\phi}$ corresponds to $t \sim -|s|e^{i\phi}$. As discussed, even $k(E, l) = \mathcal{O}(1)$ does not necessarily eliminate this behavior, though positive k—corresponding to time delay—decreases the region of nonbounded behavior in the UHP. Likewise, $k < 0$, corresponding to a time advance, increases the domain of this behavior. One also observes unbounded behavior from the eikonal phases, (5.12).

It is interesting that a polynomial bound in the physical region $\mathrm{Im}(s) > 0$, $t < 0$ (and correspondingly in other channels) follows from the very general assumptions that we have described, together with the assumption of causality in the form of the forward polynomial bound. We next turn to the investigation of connections between polynomiality and locality.

[18]A rough argument for this follows from the relation between the continuations $s \to -s$ and $E \to -E$; the latter corresponds to taking the complex conjugate of the amplitude.

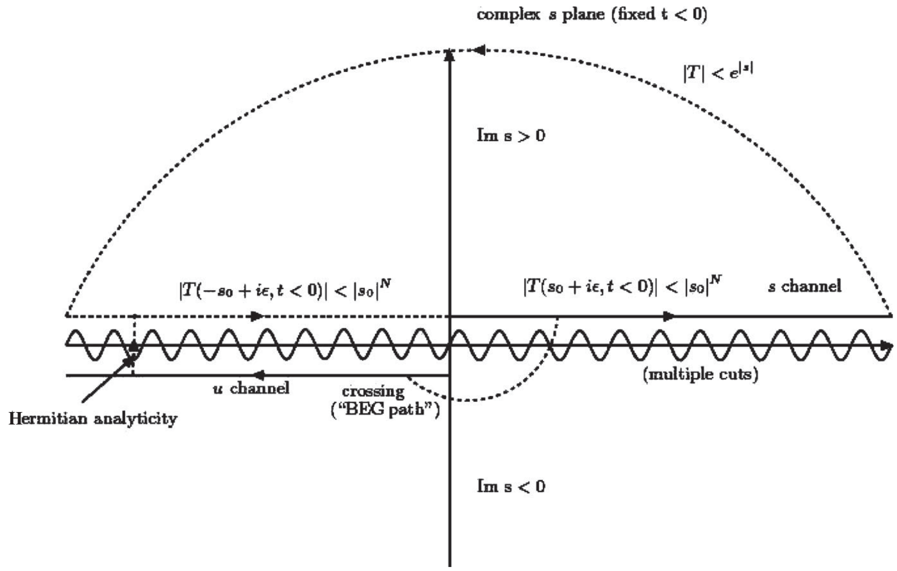

FIG. 8. The complex s plane, indicating some of the relationships entering into the Phragmen-Lindelöf argument for a polynomial bound.

VIII. LOCALITY VS NONPOLYNOMIALITY

The status of locality in gravity is a very important question, given that it is one of the cornerstones of a local quantum field theory description of nature. Locality is also one of the assumptions leading to the information paradox, and conversely, certain violations of locality inherent to nonperturbative gravity have been proposed as the mechanism for information to escape an evaporating black hole [11,23,25,40].[19]

If one is restricted to an S-matrix description of dynamics, one can ask how specifically locality is encoded in that description. In particular, nonpolynomial behavior in the momenta, such as we have described, is suggestive of nonlocal behavior[20]; a first heuristic for this is the observation that nonpolynomial interactions take the form e^{∂^n} in position space, which is clearly not local.

For massive theories, sharper statements can be made. In particular, commutativity of observables outside the light cone can be used to show that the forward amplitude is polynomial bounded [67], $|T(s, 0)| < s^N$. With a mass, such statements can be extended [68] both to $t < 0$ and to complex values of t, including $t > 0$.

Diffeomorphism invariance forbids local observables in gravity. It has been proposed that local observables are approximately recovered from certain relational *protolocal observables*; initial exploration of them in effective field theory is described in [43,69,70]. However, as yet no sharp criterion for locality can be formulated in terms of these observables, and indeed it has been argued [25,43] that there are fundamental obstacles to such precise locality.[21]

Nonetheless, bounds on amplitudes can also be understood from a physical perspective, in connection with causality. This becomes particularly clear with forward scattering.

Consider first $0 + 1$ dimensional scattering, with initial and final amplitudes related by an S matrix,

$$\psi_f(t) = \int_{-\infty}^{\infty} dt' S(t - t')\psi_i(t'). \tag{8.1}$$

Causality states that if the source ψ_i vanishes for $t' < 0$, the response ψ_f does as well. In the complex energy plane, this arises as a result of $S(E)$ having the appropriate analytic structure, and, in particular, the needed contour deformation arguments require that $S(E)$ be polynomially bounded in the UHP for E. For example, $S(E) = e^{-iE\tau}$ would produce an acausal time *advance* by τ.

The arguments for higher-dimensional forward scattering can be formulated in analogous fashion; a wave packet that scatters at zero angle should not reach infinity more rapidly than one that does not scatter, implying a polynomial bound, and corresponding dispersion relations.[22] Whereas in the massive case such a bound also implies bounds for $t \neq 0$, the collapse of the Lehmann ellipse that we have noted in the massless case obstructs such arguments.

Consider, however, a physical picture of nonforward scattering, as described in e.g. [74]; see Fig. 9. If the

[19]For earlier proposals of a role for nonlocal effects, see [51,64,65].

[20]Although, formulations of local field theory with mild nonpolynomial behavior have been proposed [66].

[21]For further discussion, see [71].

[22]The relations between causality, analyticity, and a well-defined UV completion are interesting and subtle. Indeed, other strong restrictions on which IR behavior can be consistently completed into a causal UV theory, given existence of forward dispersion relations, are described in [72,73].

scattering has a range R, a wave packet can shorten its path by an amount up to $R|q|/E$ with respect to a path going through the origin, with a corresponding time advance. Thus, we would expect asymptotic behavior

$$S \sim e^{-i\sqrt{-t}R}, \tag{8.2}$$

which is *not* bounded. Note, however, that such a picture is appropriate to a *repulsive* potential. If one instead considers scattering in gravity, e.g. in the background of a high-energy particle, whose gravitational field is approximately Aichelburg-Sexl (see Fig. 10), the scattering angle is negative, and the particle receives a time delay, corresponding to positive phase shift, appropriate to an attractive force. If of finite range R, this corresponds to behavior

$$S \sim e^{i\sqrt{-t}R}. \tag{8.3}$$

In this way, long-range behavior of this kind, which in the absence of a better definition we will also call nonlocal, does not obviously conflict with causality. The danger of a conflict appears even less in an attractive case which produces only time delays; correspondingly one has a polynomial bound for $R \propto E^p$ in this case when E undergoes a small enough positive phase rotation. Thus, plausibly, nonlocality with time delays is consistent with the existence of a polynomial bound in the physical region, $t < 0$, $\mathrm{Im}(s) > 0$. The preceding section also argued that crossing, Hermitian analyticity, and causality imply such a bound. While the large phase shifts and strong absorption up to large impact parameters that we have inferred on physical grounds might have violated such a polynomial bound in the physical region, we have found no evidence for such behavior. It remains possible that an exponential growth may emerge at fixed (real) scattering angle, other than $\theta = 0$. This however does not seem to contradict any fundamental property we know, but is another possible signal of nonlocal behavior.[23]

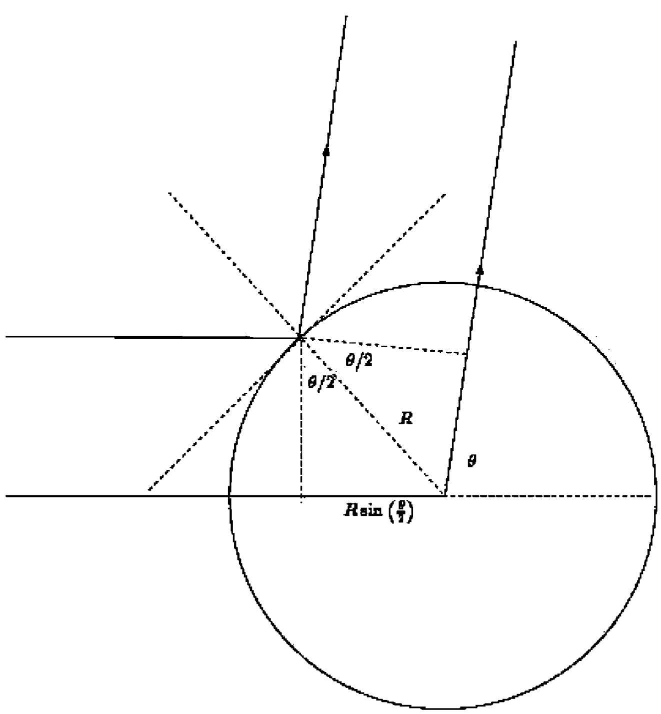

FIG. 9. Illustration of scattering by a repulsive interaction of range R; the scattered wave at angle θ has a path that is shorter by $2R\sin\frac{\theta}{2}$ relative to a wave traveling unscattered through the origin, and thus has a relative time advance.

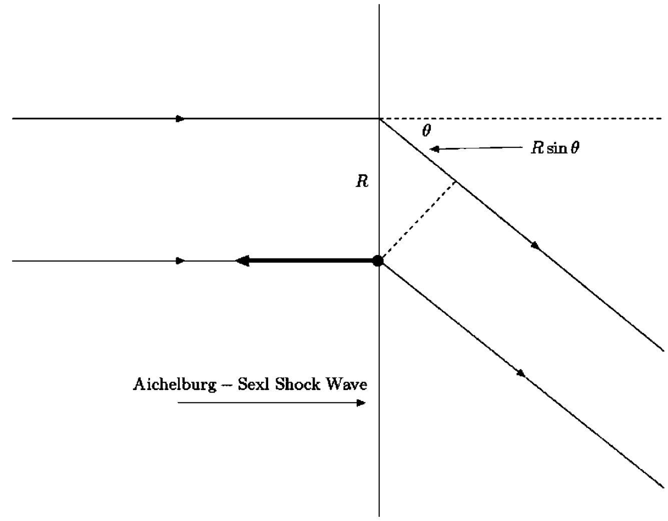

FIG. 10. Illustration of scattering of a particle by the gravitational field of an ultrarelativistic source; the scattering angle is negative, corresponding to attraction, and this results in a path for the scattered wave that is longer by $R\sin\theta \sim 2R\sqrt{tu}/s$ as compared to a wave that passes through the scattering center.

In saying this, we should address arguments of [22] suggesting behavior combining (8.2) with (8.3), where $R = R(E)$, which would be naturally interpreted in terms of a time advance. However, this arose from a sharp cutoff in the partial wave sum and does not account for the phase shifts. If one avoids $\theta = 0$, where causality requires cancellations of nonpolynomial behavior [22], we can write

$$T(s,t) \propto \sum_{l=0}^{\infty} (l+\lambda) C_l^{\lambda}(\cos\theta) e^{2i\delta_l(s) - 2\beta_l(s)} \tag{8.4}$$

[the sum of $i/2$ generates a $\delta(\cos\theta - 1)$]. Plausibly, the exact phase shifts and absorptive coefficients yield only time-delayed behavior, and a bound in the s UHP.

In the preceding section, we argued that the effective range of the interaction grows with E; $R \sim E^p$, with $p = 1/(D-3)$ for the strong gravity region, and the rough estimate $p = 2/(D-4)$, from (6.8), for the eikonal amplitudes. It is interesting to compare this behavior to what is commonly regarded as another indicator of unitary local behavior, the Froissart bound, which states

$$R \le R_f = a \log E \tag{8.5}$$

[23] As noted, one might also consider the possibility, which we have not been able to rule out, that amplitudes, while nonpolynomial, may have sufficiently complicated analytic structure to stay polynomially bounded in other regions as well.

for constant a. In a massive theory, there is a direct connection between this bound and polynomial boundedness. Heuristically, this is seen via

$$e^{\pm R_f\sqrt{-t}} \sim E^{\pm a\sqrt{-t}}, \tag{8.6}$$

which is polynomial behavior. More sharply, the polynomial bound is used directly in the proof of the Froissart bound [29,75]. However, this proof proceeds via the partial wave expansion in the region $t > 0$, which we have argued is divergent for gravity.

It is tempting to conjecture that there is such a direct connection between power-law growth of the cross section in gravity and nonpolynomiality, perhaps through appropriate regulation of the partial wave expansion. Indeed, as discussed in [22] and above, the appearance of strong absorption to $L \sim E^{(D-2)/(D-3)} \gg E\ln E$ implies nonpolynomial behavior of a truncated partial wave sum.[24] However, as we have argued, we expect the full sum to be polynomial bounded in the s UHP, even if it is not polynomial. One issue arising from massless modes is that we cannot neglect the tail of the partial wave expansion, as one does, for example, in theories with a mass gap, where f_l decays exponentially for $l \gg E\log E$. In our gravitational context, these large impact parameter contributions are central in producing the IR singularities at $t = 0$. Indeed, masslessness also plays an important role in the form of the amplitudes in the eikonal regime (where $l \gg L$), which appears to dominate the cross section at large energies. Since the partial wave expansion does not converge at $t > 0$, the Froissart bound can be violated without collateral damage. We may associate this with a sort of IR/UV mixing, in the sense that the singularities in the IR (correspondingly the long-range character of gravity) permit a much faster growth in the cross section deep in the UV without conflicting with any other fundamental property. Notice that the eikonal amplitudes already provide us with such an example, without explicit reference to the strong gravity region.

One thus finds that masslessness, and, in particular, singular behavior at $t = 0$, nonpolynomiality, and polynomial growth of cross sections are intricately entwined. One might question whether all novel features follow from masslessness alone. However, given that one does not find power-law growth $R \sim E^p$ in gauge theory, gravity appears distinctive, due in part to the power-law growth of its coupling with energy. One might conjecture that a massless theory like QED is on the borderline of locality, but gravity is in a real sense not local, as, for example, evidenced by its growth of range. Such a conjecture is certainly permitted without a sharper characterization of locality.

It is interesting to consider one known approach to regulating IR behavior in gravity, namely, working in an anti–de Sitter (AdS) background. With AdS curvature $\mathcal{R} \sim \mu^2$, the graviton effectively has a mass $\sim\mu$. Correspondingly, growth of black hole radius with energy stops being power law once $R \sim 1/\mu$, and one, in particular, finds evidence for Froissart-like behavior, $R \propto \log E$, for scattering above this energy [76]. One might likewise expect restoration of polynomial scattering amplitudes. However, the matter of extracting the S matrix in AdS remains an open question [20], despite some recent progress [18,19].

It is very interesting that no fundamental inconsistency has yet arisen between the conditions of unitarity, analyticity, crossing symmetry, causality, and nonlocality in the sense described, despite the existence of nontrivial constraints arising from their combination; it is also moreover interesting that gravitational amplitudes could well run the gauntlet among these conditions. This would also be in harmony with arguments that local field theory breaks down in contexts described by the locality bound [23,25,26], and with more general statements that the nonperturbative physics that unitarizes gravity (and specifically leads to unitary black hole decay) is not intrinsically local [23], yet retains certain analytic features and aspects of causality—particularly those necessary for consistency. In any case, further exploration of properties of consistent quantum-mechanical amplitudes for gravity is certainly of great interest.

ACKNOWLEDGMENTS

We wish to thank N. Arkani-Hamed, H. Epstein, M. Green, D. Gross, A. Martin, J. Polchinski, M. Srednicki, R. Stora, D. Trancanelli, G. Veneziano, and E. Witten for valuable discussions. We greatly appreciate the stimulating hospitality of the CERN theory group over the course of part of this work. This work was supported in part by the Department of Energy under Contract No. DE-FG02-91ER40618, and by Grant No. RFPI-06-18 from the Foundational Questions Institute (fqxi.org).

APPENDIX: OPTICAL THEOREM IN D DIMENSIONS

From the unitarity of the S matrix we have

$$T_{\alpha\beta} - T^*_{\beta\alpha} = i\sum_N \int (2\pi)^D d\Phi_N T_{\alpha N} T^*_{\beta N}, \tag{A1}$$

where we take α, β to be the initial and final two-body states with $p_\alpha \equiv p_1 + p_2$, $p_\beta = p_3 + p_4$, and the sum runs over all possible N-particle states allowed by the symmetries and conservation of energy and momentum.

[24]Note that such strong absorption directly corresponds to a cross section with growth (6.7). This follows from taking $\beta_l \gg 1$ for $l \ll L$ in (2.6) evaluated at $\theta = 0$; this, together with the large-l asymptotics $C_l^\lambda(1) \sim l^{2\lambda-1}/\Gamma(2\lambda)$ gives $T(\theta = 0) \sim is^{(4-D)/2}L^{D-2}$, and thus, by the optical theorem, (6.7). Of course, as we have noted, an even larger contribution to σ_T comes from the eikonal region.

Here we use the Lorentz invariant normalization of states,

$$\langle k|k'\rangle = (2\pi)^{D-1}2\omega_{\mathbf{k}}\delta^{D-1}(\mathbf{k}-\mathbf{k}') \tag{A2}$$

with $\omega_{\mathbf{k}}^2 = \mathbf{k}^2 + m^2$, and introduce the Lorentz invariant measure

$$\widetilde{dk} \equiv \frac{d^{D-1}k}{(2\pi)^{D-1}2\omega_{\mathbf{k}}}. \tag{A3}$$

If the intermediate N-particle state consists of momenta q_i, the N-body phase space is defined by

$$d\Phi_N = \delta^D\Big(p_\alpha - \sum_i^N q_i\Big)\prod_{i=1}^N \widetilde{dq}_i. \tag{A4}$$

Using these conventions we have for the dimensions of the $2 \to 2$ scattering amplitude, $[T(s,t)] = M^{4-D}$.

If we now restrict (A1) to forward scattering, e.g. $\alpha = \beta$, we can replace the left-hand side by $2i\,\mathrm{Im}T(s,0)$, and on the right-hand side we recognize the sum of the square of the amplitudes which enters in the definition of the total cross section. Recall that this is defined as

$$\sigma_T \equiv \sigma(\alpha \to \text{all}) = \Big[\frac{1}{4\sqrt{(p_1\cdot p_2)^2 - m_1^2m_2^2}}\Big](2\pi)^D\sum_N\int d\Phi_N|T_{\alpha N}|^2. \tag{A5}$$

Notice that the prefactor in square brackets goes to $1/(8E_1E_2)$ when $s \gg m_1^2, m_2^2$. We are now ready to state the optical theorem, which is nothing but a direct consequence of unitarity:

$$\mathrm{Im}\,T(s,0) = 2\sqrt{(p_1\cdot p_2)^2 - m_1^2m_2^2}\,\sigma_T(s) \to s\sigma_T(s). \tag{A6}$$

We can also relate the coefficients in the partial wave projections (2.6), where the optical theorem takes the form (in the $s \gg m^2$ limit) [7]

$$\mathrm{Im}f_l(s) = 8(2\pi)^{2D-2}\Big(\frac{s}{4}\Big)^{2-D/2} \times \sum_N \delta^D(p_N - p_\alpha)|f_l(s,\{N\})|^2, \tag{A7}$$

from which (2.12) follows. In this expression the $f_l(s,\{N\})$ are the partial wave projections of the generic intermediate states, considered modulo an overall rotation. The sum runs over all possible such subclasses of states [7]. Performing the sum over l on both sides reproduces the optical theorem.

As we emphasized in this paper, due to the masslessness of gravity we expect singularities at $t = 0$. We noticed before that the IR singularities can be removed by working in $D > 4$. From the definition of the cross section we promptly discover that we actually need even higher D for it to be well defined. This follows from the elastic cross section; (2.5) gives probability

$$|T|^2 \sim \frac{1}{\theta^4}. \tag{A8}$$

This Rutherford-like singularity is tamed for $D > 6$ by the integration over solid angle, with measure $\sin^{D-3}\theta$, giving a finite cross section. Once the cross sections are finite the optical theorem (A6) shows that $\mathrm{Im}T(s,0)$ is also finite. One may be tempted to push the partial wave expansion to $t > 0$, but this attempt fails once we realize that $t = 0$ is indeed also a threshold for graviton production, and the partial wave expansion will not converge past that point. The finiteness of $\mathrm{Im}T(s,0)$ is due to the fact that in higher dimensions the threshold behavior scales as a power of momentum, e.g. $\sim(-t)^\alpha$, rather than logarithmically as we are used to encountering in four-dimensional field theories. This is intimately linked to the softness of the IR divergences in $D > 4$ due to the promotion of the measure in the loop integrals from $\frac{d^4q}{(2\pi)^4}$ to $\frac{d^Dq}{(2\pi)^D}$. It is then easy to see that the expansion of the derivatives of $T(s,t)$ at $t = 0$ will not converge and we cannot analytically continue the partial wave decomposition to positive values of t.

A final comment is in order. The reader may be puzzled by the fact that the Born approximation in (2.5) seems to have a divergent imaginary part as $t \to 0$ from the $i\epsilon$ prescription. A careful analysis shows that is indeed not the case, and such singularity only arises in the plane-wave limit and disappears as soon as we take into account wave packets. The real part of the amplitude is large, but finite, and gives rise to a finite contribution in the cross section as in (A8).

[1] R.J. Eden, P.V. Landshoff, D.I. Olive, and J.C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, Cambridge, England, 2002).

[2] D. Amati, M. Ciafaloni, and G. Veneziano, Phys. Lett. B **197**, 81 (1987); Int. J. Mod. Phys. A **3**, 1615 (1988).

[3] D. Amati, M. Ciafaloni, and G. Veneziano, Phys. Lett. B **216**, 41 (1989).

[4] D. Amati, M. Ciafaloni, and G. Veneziano, Nucl. Phys. **B347**, 550 (1990).

[5] D. Amati, M. Ciafaloni, and G. Veneziano, Nucl. Phys. **B403**, 707 (1993).

[6] D. J. Gross and P. F. Mende, Phys. Lett. B **197**, 129 (1987);

Nucl. Phys. **B303**, 407 (1988).
[7] M. Soldate, Phys. Lett. B **186**, 321 (1987).
[8] G. 't Hooft, Phys. Lett. B **198**, 61 (1987).
[9] I. J. Muzinich and M. Soldate, Phys. Rev. D **37**, 359 (1988).
[10] T. Banks and W. Fischler, arXiv:hep-th/9906038.
[11] S. B. Giddings, Phys. Rev. D **74**, 106006 (2006) [arXiv: hep-th/0604072].
[12] S. B. Giddings, D. J. Gross, and A. Maharana, Phys. Rev. D **77**, 046001 (2008) [arXiv:0705.1816].
[13] D. Amati, M. Ciafaloni, and G. Veneziano, J. High Energy Phys. 02 (2008) 049 [arXiv:0712.1209].
[14] J. M. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998); Int. J. Theor. Phys. **38**, 1113 (1999) [arXiv:hep-th/ 9711200].
[15] J. Polchinski, arXiv:hep-th/9901076.
[16] L. Susskind, arXiv:hep-th/9901079.
[17] S. B. Giddings, Phys. Rev. D **61**, 106008 (2000) [arXiv: hep-th/9907129].
[18] M. Gary, S. B. Giddings, and J. Penedones, Phys. Rev. D **80**, 085005 (2009) [arXiv:0903.4437].
[19] I. Heemskerk, J. Penedones, J. Polchinski, and J. Sully, J. High Energy Phys. 10 (2009) 079 [arXiv:0907.0151].
[20] M. Gary and S. B. Giddings, Phys. Rev. D **80**, 046008 (2009) [arXiv:0904.3544].
[21] T. Banks, W. Fischler, S. H. Shenker, and L. Susskind, Phys. Rev. D **55**, 5112 (1997) [arXiv:hep-th/9610043].
[22] S. B. Giddings and M. Srednicki, Phys. Rev. D **77**, 085025 (2008) [arXiv:0711.5012].
[23] S. B. Giddings, Phys. Rev. D **74**, 106005 (2006) [arXiv: hep-th/0605196].
[24] N. N. Bogolubov, A. A. Logunov, A. I. Oksak, and I. T. Todorov, *General Principles of Quantum Field Theory* (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990).
[25] S. B. Giddings and M. Lippert, Phys. Rev. D **65**, 024006 (2001) [arXiv:hep-th/0103231]; **69**, 124019 (2004) [arXiv: hep-th/0402073].
[26] S. B. Giddings, Phys. Rev. D **74**, 106009 (2006) [arXiv: hep-th/0606146].
[27] H. L. Lehmann, Nuovo Cimento **10**, 579 (1958).
[28] M. Froissart, Phys. Rev. **123**, 1053 (1961).
[29] A. Martin, Phys. Rev. **129**, 1432 (1963).
[30] S. Weinberg, Phys. Rev. **140**, B516 (1965).
[31] H. L. Verlinde and E. P. Verlinde, Nucl. Phys. **B371**, 246 (1992) [arXiv:hep-th/9110017].
[32] D. Kabat and M. Ortiz, Nucl. Phys. **B388**, 570 (1992) [arXiv:hep-th/9203082].
[33] Z. Bern, J. J. Carrasco, L. J. Dixon, H. Johansson, and R. Roiban, Phys. Rev. Lett. **103**, 081301 (2009) [arXiv:0905.2326].
[34] W. D. Goldberger and I. Z. Rothstein, Phys. Rev. D **73**, 104029 (2006) [arXiv:hep-th/0409156].
[35] J. B. Gilmore and A. Ross, Phys. Rev. D **78**, 124021 (2008) [arXiv:0810.1328].
[36] D. M. Eardley and S. B. Giddings, Phys. Rev. D **66**, 044011 (2002) [arXiv:gr-qc/0201034].
[37] S. B. Giddings and V. S. Rychkov, Phys. Rev. D **70**, 104026 (2004) [arXiv:hep-th/0409131].
[38] S. B. Giddings, arXiv:hep-th/9412138; arXiv:hep-th/ 9508151.
[39] A. Strominger, arXiv:hep-th/9501071.
[40] S. B. Giddings, Phys. Rev. D **76**, 064027 (2007) [arXiv: hep-th/0703116].
[41] D. N. Page, Phys. Rev. Lett. **71**, 3743 (1993) [arXiv:hep-th/9306083].
[42] M. J. Duff, Phys. Rev. D **7**, 2317 (1973).
[43] S. B. Giddings, D. Marolf, and J. B. Hartle, Phys. Rev. D **74**, 064018 (2006) [arXiv:hep-th/0512200].
[44] G. Veneziano, J. High Energy Phys. 11 (2004) 001 [arXiv: hep-th/0410166].
[45] G. T. Horowitz and J. Polchinski, Phys. Rev. D **55**, 6189 (1997) [arXiv:hep-th/9612146].
[46] S. B. Giddings and S. D. Thomas, Phys. Rev. D **65**, 056010 (2002) [arXiv:hep-ph/ 0106219].
[47] S. Dimopoulos and G. L. Landsberg, Phys. Rev. Lett. **87**, 161602 (2001) [arXiv:hep-ph/ 0106219].
[48] S. B. Giddings, AIP Conf. Proc. **957**, 69 (2007) [arXiv:0709.1107].
[49] R. Penrose (unpublished).
[50] R. C. Myers and M. J. Perry, Ann. Phys. (N.Y.) **172**, 304 (1986).
[51] G. 't Hooft, arXiv:gr-qc/9310026.
[52] V. Weisskopf and E. P. Wigner, Z. Phys. **63**, 54 (1930); **65**, 18 (1930).
[53] N. Sanchez, J. Math. Phys. (N.Y.) **17**, 688 (1976).
[54] G. N. Watson, *A Treatise on The Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1966), 2nd ed.
[55] T. Adachi and T. Kotani, Prog. Theor. Phys. **39**, 430 (1968); **39**, 785 (1968).
[56] W. N. Cottingham and R. F. Peierls, Phys. Rev. **137**, B147 (1965).
[57] T. Adachi and T. Kotani, Prog. Theor. Phys. Suppl. **37–38**, 297 (1966).
[58] M. Sugawara and A. Tubis, Phys. Rev. **130**, 2127 (1963).
[59] T. Kinoshita, Phys. Rev. **154**, 1438 (1967).
[60] R. P. Boas, *Entire Functions* (Academic Press, New York, 1954).
[61] A. Martin, Nuovo Cimento **37**, 671 (1965).
[62] J. Bros, H. Epstein, and V. Glaser, Commun. Math. Phys. **1**, 240 (1965).
[63] J. Bros, H. Epstein, and V. Glaser, Nuovo Cimento Series X **31**, 1265 (1964).
[64] S. B. Giddings, Phys. Rev. D **46**, 1347 (1992) [arXiv:hep-th/9203059].
[65] L. Susskind, J. Math. Phys. (N.Y.) **36**, 6377 (1995) [arXiv: hep-th/9409089].
[66] A. M. Jaffe, Phys. Rev. **158**, 1454 (1967).
[67] M. Gell-Mann, M. L. Goldberger, and W. E. Thirring, Phys. Rev. **95**, 1612 (1954).
[68] A. Martin, Nuovo Cimento A **42**, 930 (1966).
[69] M. Gary and S. B. Giddings, Phys. Rev. D **75**, 104007 (2007) [arXiv:hep-th/0612191].
[70] S. B. Giddings and D. Marolf, Phys. Rev. D **76**, 064023 (2007) [arXiv:0705.1178].
[71] S. B. Giddings, Mod. Phys. Lett. A **22**, 2949 (2007) [arXiv:0705.2197].
[72] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, J. High Energy Phys. 10 (2006) 014 [arXiv:hep-th/0602178].
[73] J. Distler, B. Grinstein, R. A. Porto, and I. Z. Rothstein,

Phys. Rev. Lett. **98**, 041601 (2007) [arXiv:hep-ph/0604255].
[74] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[75] M. Chaichian, J. Fischer, and Yu. S. Vernov, Nucl. Phys. **B383**, 151 (1992).
[76] S. B. Giddings, Phys. Rev. D **67**, 126001 (2003) [arXiv:hep-th/0203004].