

Loss of locality in gravitational correlators with a large number of insertionsSudip Ghosh^{*} and Suvrat Raju[†]*International Centre for Theoretical Sciences, Tata Institute of Fundamental Research,
Shivakote, Bengaluru 560089, India*

(Received 17 July 2017; published 28 September 2017)

We review lessons from the AdS/CFT correspondence that indicate that the emergence of locality in quantum gravity is contingent upon considering observables with a small number of insertions. Correlation functions, where the number of insertions scales with a power of the central charge of the CFT, are sensitive to nonlocal effects in the bulk theory, which arise from a combination of the effects of the bulk Gauss law and a breakdown of perturbation theory. To examine whether a similar effect occurs in flat space, we consider the scattering of massless particles in the bosonic string and the superstring in the limit, where the number of external particles, n , becomes very large. We use estimates of the volume of the Weil-Petersson moduli space of punctured Riemann surfaces to argue that string amplitudes grow factorially in this limit. We verify this factorial behavior through an extensive numerical analysis of string amplitudes at large n . Our numerical calculations rely on the observation that, in the large n limit, the string scattering amplitude localizes on the Gross-Mende saddle points, even though individual particle energies are small. This factorial growth implies the breakdown of string perturbation theory for $n \sim (\frac{M_{\text{pl}}}{E})^{d-2}$ in d dimensions, where E is the typical individual particle energy. We explore the implications of this breakdown for the black hole information paradox. We show that the loss of locality suggested by this breakdown is precisely sufficient to resolve the cloning and strong subadditivity paradoxes.

DOI: 10.1103/PhysRevD.96.066033

I. INTRODUCTION

It is generally recognized that quantum gravity cannot be an exactly local theory due to the difficulty of localizing operators in spacetime to an accuracy better than the Planck length, ℓ_{pl} . In this paper, we would like to present evidence that, for at least some observables, these nonlocal effects can spread over *macroscopic distance scales*. We will argue that the observables that are sensitive to these effects are correlation functions with a very “large” number of insertions. A significant part of our analysis will be devoted to quantifying what we mean by “large” here.

The initial motivation to consider such effects came from the results of [1–3]. These papers proposed a representation of the interior of large black holes in the AdS/CFT correspondence [4]. But this representation had the remarkable property that the operators that described the degrees of freedom in the interior of the black hole were complicated combinations of operators that described the exterior of the black hole. This implied that a suitably complicated combination of exterior operators would fail to commute with an interior operator even if the exterior and interior operators were separated by a distance that was large in Planck units.

However, it is natural to expect that if these nonlocal effects are real, then they should be visible even in empty space, in the absence of black holes. This is indeed the case. A simple example of these nonlocal effects was examined in a

controlled setting in [5]. The paper [5] considered an operator localized in the center of empty anti-de Sitter (AdS).¹ In AdS/CFT, operators in the bulk of AdS can be mapped to the CFT using the standard Hamilton-Kabat-Lifschytz-Lowe mapping [6]. The authors of [5] then showed that this operator could be explicitly rewritten as a complicated combination of other operators that were localized near the boundary of AdS on the same time slice. We review this construction in Sec. II B, where we show that the operator in the center of AdS, $\phi(0)$, can be written as

$$\phi(0) = \sum_{n,m} c_{nm} X_n \mathcal{P}_0 X_m^\dagger. \quad (1.1)$$

Here, X_n and X_m are simple polynomials in operators localized near the boundary, c_{nm} are c-number coefficients, and \mathcal{P}_0 , which we calculate explicitly in II B, is a complicated polynomial that involves the bulk graviton fluctuations near the boundary and has a degree as large as N , where N^2 is the central charge of the theory.

The central feature of (1.1) is that the operators that appear in the polynomials on the right-hand side are all localized at points that are *spacelike* to the operator that appears on the left-hand side. This equation obviously implies nonzero commutators at spacelike separation. For example, an operator that contains the momentum conjugate to $\phi(0)$ will fail to commute with the operators on the right-hand side of (1.1). However, the relation (1.1) provides us with a stronger statement: it tells us that the information in the center of AdS

^{*}sudip.ghosh@icts.res.in
[†]suvrat@icts.res.in

¹We describe what we mean by an operator “localized” at a point in greater detail in Sec. II A.

can be entirely recovered by making a suitably complicated measurement, on the same time slice, near the boundary of AdS. Thus, (1.1) provides a toy model of the phenomenon of black hole complementarity [7].

There are two physical effects that allow the relation (1.1) to hold. One of them is the bulk Gauss law, which tells us that the energy of a local operator can be measured through a Hamiltonian that is entirely defined through a surface integral at infinity. The Gauss law itself leads to small nonvanishing commutators between operators at spacelike separations. In AdS, these commutators are suppressed by factors of $\frac{1}{N}$. However, the key to (1.1) is that by taking complicated enough combinations of localized operators, we can enhance these $\frac{1}{N}$ effects to an $O(1)$ effect. This requires the *breakdown* of $\frac{1}{N}$ perturbation theory. Thus, the nonlocal effects that we are looking for arise from a combination of a kinematical effect—the Gauss law—and a dynamical effect—the breakdown of gravitational perturbation theory.

In this paper, our main focus is on exploring whether similar effects exist in flat space and, if so, on the implications that such effects might have for black hole evaporation. Of course, the kinematic ingredient that was present in AdS—the Gauss law—operates in flat space as well. In flat space, the Gauss law leads to commutators between local operators that are suppressed by a power of $(\frac{E}{M_{\text{pl}}})$, where E is a measure of the energy of the configuration of operators and $M_{\text{pl}} = \frac{1}{\ell_{\text{pl}}}$ is the Planck scale.

So, in this paper, we will focus on the dynamical effect that was required above: the breakdown of the perturbation theory. In flat space, the analogue of the breakdown of the $\frac{1}{N}$ expansion is the breakdown of gravitational perturbation theory for correlation functions where the number of insertions becomes very large, even if these insertions are well separated.

The fact that the breakdown of the perturbation theory corresponds to a loss of bulk locality in theories of gravity is also natural from the path-integral viewpoint. The reason that gravity is approximately local for simple observables, even though the path integral sums over all metrics, is because the path integral is dominated by a saddle point in many circumstances. This saddle point provides a dominant metric, and when we refer to approximate locality, we are referring to locality as defined by this metric.

The breakdown of perturbation theory is a sign that the saddle point approximation is no longer valid for a particular observable. So it is natural that a perturbative breakdown in gravity corresponds to either a loss or a change in the notion of locality.²

²We caution the reader that, in doing the quantum-gravity path integral, we hold the asymptotic geometry fixed. Therefore, in this entire discussion, loss of locality only refers to a loss of “bulk locality” and not any change in the causal relationship of points on the asymptotic boundary to each other. This is true in AdS/CFT as well where the boundary theory remains exactly local even though the bulk theory is not exactly local.

To study this perturbative breakdown in a precise setting, we focus on S-matrix elements rather than correlation functions since scattering amplitudes are naturally gauge invariant. We also study scattering in string theory—both the bosonic theory and superstring theory—rather than pure gravity. This helps us ensure that the breakdown of perturbation theory that we describe here is not cured by stringy effects. However, it is also true—perhaps somewhat surprisingly—that the technical analysis of the breakdown of perturbation theory turns out to be easier in string theory than in pure gravity for reasons that we explain in Sec. V D.

Our results are as follows. We consider the scattering of massless particles in the bosonic string and type II superstring theory with external polarization tensors chosen so that these particles correspond to linear combinations of the graviton and the dilaton. We take the limit where the number of external particles, $n \rightarrow \infty$, but where the energy per particle is taken to zero in string units $E\sqrt{\alpha'} = \frac{1}{n^\gamma}$ with $0 < \gamma < \frac{1}{d-2}$, and the dimensionless string coupling constant, g_s , is also taken to be very small. Then we argue that string perturbation theory breaks down for $n \propto (\frac{1}{g_s})^{\frac{1}{1-(d-2)\gamma}}$. This threshold for the breakdown of perturbation theory can also be rewritten as $n \propto (\frac{M_{\text{pl}}}{E})^{d-2}$.

To derive this bound, we first derive some simple bounds on the growth of tree amplitudes in any perturbative theory in Sec. III. These bounds state that if tree amplitudes grow factorially in the number of external particles then, provided the energy per particle does not fall too rapidly, the perturbation theory eventually breaks down for a large enough number of external particles. In Secs. IV and V, we then argue that tree-level scattering amplitudes, both in the bosonic string and the superstring, do grow factorially.

Our arguments in Sec. IV rely on the growth of the volume of moduli space of punctured Riemann surfaces. This is a subject that has attracted some recent attention in the mathematical literature. We are able to utilize these results by arguing that at large n , the size of the string amplitudes is bounded below by the volume of moduli space.

Our arguments in Sec. V are independent and rely on a numerical study of string scattering at large n . Here, we exploit the fact that at large n , the integral over moduli space localizes to a set of saddle points that are solutions of the so-called “scattering equations.” By solving these equations numerically, we are able to numerically estimate the growth of amplitudes. This numerical estimate precisely matches the estimate from the volume of moduli space that we derive in Sec. IV.

The arguments above then suggest that scattering experiments with a large enough number of external particles may be sensitive to nonlocal effects in the bulk. In Sec. VI, we explore the implications of such effects for the information paradox. In particular, we consider two versions of the information paradox—the “cloning paradox” and the “strong subadditivity paradox.”

The cloning paradox is based on the observation that the Hawking radiation that emerges from a black hole at late times carries information about the infalling matter. However, the geometry of the evaporating black hole suggests that we can draw a spacelike nice slice that intersects both the infalling matter and the outgoing Hawking radiation. This seems to suggest that the same information is present in two places, which would violate the linearity of quantum mechanics.

However, when we carefully examine the observables that are required to extract information from the outgoing radiation, it turns out that we need to measure S -point correlators of the outgoing Hawking quanta, at intervals of size $\frac{1}{E}$, where S is the entropy of the black hole and E is energy of a typical Hawking quanta. Our analysis of the breakdown of perturbation theory tells us that gravitational perturbation theory breaks down *precisely* for such correlators.

This suggests an elegant resolution to the cloning paradox. It is *not* the case that two different operators can extract the same information from the state. Rather the simple operator that acts on the infalling matter should be *equated* to a complicated operator that acts on the outgoing Hawking radiation. Even though it seems that these operators are distinct, because they seem to act at different points in space, they may still be related through a relation of the form (1.1). Therefore, our resolution to the cloning paradox is that the linearity of quantum mechanics is preserved, and it is our notion of locality that must be modified.

The strong subadditivity paradox is closely related to the cloning paradox. As we review in Sec. VI, this paradox relies on splitting up the black hole spacetime into three regions on a spacelike slice. Plausible arguments about the von Neumann entropies of each of these three regions then suggest that these entropies violate the strong subadditivity of entropy. However, the von Neumann entropy is a fine-grained probe of a state, and we argue in VIB that measuring this quantity is equivalent to measuring a S -point correlator in the Hawking radiation. Therefore, nonlocal effects may be important for such quantities. In particular, we should not expect that the Hilbert space of the theory factorizes into a tensor product of the Hilbert spaces corresponding to different local regions on a spacelike slice. Thus the strong subadditivity paradox may also be resolved by recognizing limits on locality in gravity.

In Sec. VID, we show how the simple setup of empty AdS considered in Sec. IIB can also be used to produce toy models of both the cloning and the strong subadditivity paradoxes. In this setting, the resolution to both of these paradoxes is absolutely clear and involves, as we expect, a loss of bulk locality rather than any modification of quantum mechanics.

A summary of the main results in this paper was provided in [8]. The scattering of a large number of particles was also studied in [9] although the motivation

and perspective of these papers was different from ours. The significance of $O(S)$ point correlators for the information paradox, and the fact that they might deviate strongly from naive expectations, was also discussed in [10].

Our conventions in the rest of the paper are as follows. We set $\alpha' = 2$. We will use the string coupling constant $4\pi^2 g_s^2 = \kappa^2 (\frac{\alpha'}{2})^{\frac{2-d}{2}}$. With our choice of units, this simply becomes $g_s = \frac{\kappa}{2\pi}$, where κ is the d -dimensional gravitational coupling. We also use $\kappa^2 = 8\pi G = \ell_{\text{pl}}^{d-2}$.

II. LOCALITY IN GRAVITY AND PERTURBATION THEORY

In this section, we describe the relation between approximate notions of locality in gravity and the validity of perturbation theory. First, we clarify what we mean by an approximately local operator in a theory of gravity. Then we review the results of [5]. This analysis provides an explicit example of nonlocality in quantum gravity in a controlled setting. Then we abstract some lessons from this example and show how appreciable nonlocal effects in correlators result from a *combination* of the effect of the Gauss law and the breakdown of perturbation theory. We then provide some additional arguments for nonlocality based on the path integral. Finally, we emphasize the distinction between asymptotic notions of locality, which we expect to be exact, and bulk locality, which we expect is only approximate.

A. Approximately local operators in gravity

Since this paper is devoted to the study of nonlocal effects in gravity, it is important to clarify what we mean by an approximately local operator. For simplicity, we will consider scalar operators. Under a diffeomorphism, $x^\mu \rightarrow x^\mu + \eta^\mu$, a scalar operator transforms as $\phi(x) \rightarrow \phi(x) - \eta^\mu \partial_\mu \phi(x)$. Therefore, unless we provide additional information, this operator does not provide us with gauge invariant information.

The simplest way to resolve this issue is just to fix gauge. Alternately, it is possible to use a relational prescription that fixes the location of the operator with reference to an asymptotic boundary. An example of such a relational prescription is given in Sec. IV of [2] or Sec. 3.1.1 of [3].

Both alternatives then lead to operators that have the following important property. If x_1, x_2 are spacelike separated then in the limit where $M_{\text{pl}} \rightarrow \infty$ and within low point correlators, these operators satisfy

$$\langle 0 | [\phi(x_1), \phi(x_2)] \phi(x_3) \dots \phi(x_{2n}) | 0 \rangle = 0, \quad \text{as } M_{\text{pl}} \rightarrow \infty, \quad (2.1)$$

where n does not scale with any power of M_{pl} and $|x_{ij}| \gg \frac{1}{M_{\text{pl}}}, \forall i, j$.

The property (2.1) defines what we mean by a local operator in this paper. Note that the commutator of such operators does *not* vanish in general when M_{pl} is finite. Second, as we describe in great detail in the rest of this paper, various subtle effects arise when we keep M_{pl} finite and scale the number of insertions in a correlator with a power of M_{pl} . Consequently, the notion of locality described above is only *approximate*. We will sometimes refer to such operators as “quasilocal” operators.

B. Small algebras and locality in AdS/CFT

If nonlocal effects are present in nature, we would expect that they should be observable even in empty space without black hole horizons which are, after all, teleological objects. The AdS/CFT correspondence provides us with a simple setting to study such effects, and it is indeed true that these effects are present even in empty AdS. This was shown in [5]. We review this example below, and it will serve as a prototype for the nonlocal effects that we will later invoke while considering the information paradox in flat space.

Consider empty global AdS_{d+1} , where we set the AdS radius $\ell_{\text{ads}} = 1$ so that the metric is

$$ds^2 = -(1+r^2)dt^2 + \frac{dr^2}{1+r^2} + r^2 d\Omega_{d-1}^2.$$

We also consider a band of length T on its boundary, where $T < \pi$ so that the band is smaller than the light-crossing time of AdS. This gives rise to the setup shown in Fig. 1. We denote the band itself by \mathcal{B} , by which we mean the set of all points on the boundary with time coordinate between $-\frac{T}{2}$ and $\frac{T}{2}$. A causal diamond \mathcal{D} , in the bulk, is out of causal contact with \mathcal{B} . We also have the “causal wedge” $\bar{\mathcal{D}}$ in the

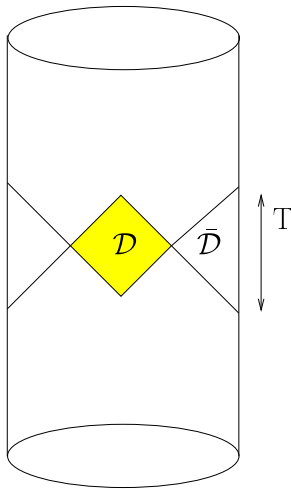


FIG. 1. A band of length $T < \pi$ on the boundary. A causal diamond \mathcal{D} in the middle of AdS is out of causal contact with the band.

bulk. Here, by causal wedge, we mean the region in the bulk such that from each point in this region, it is possible to send *both* a future directed light ray and a past-directed light ray to the band \mathcal{B} . In the literature, the term “causal wedge” is often only applied to bulk regions that are dual to an entire causal diamond on the boundary, but this terminology also makes sense in the setting considered here. The significance of the causal wedge is that perturbative fields propagating in the causal wedge can be related to simple operators in \mathcal{B} using the equations of motion.

The bulk AdS may have various propagating, weakly coupled fields. We will take ϕ to be one such field, dual to an operator \mathcal{O} of dimension Δ . The standard AdS/CFT dictionary then relates the boundary value of ϕ to the value of \mathcal{O} through

$$\lim_{r \rightarrow \infty} r^\Delta \phi(t, r, \Omega) = \mathcal{O}(t, \Omega).$$

If the field ϕ is weakly coupled in the bulk, then the operator $\mathcal{O}(t, \Omega)$ is a generalized free field in the conformal field theory that is dual to this bulk theory.

We would now like to consider the “algebra” of “low-order polynomials” in these generalized free fields. This means that we consider the set of polynomials in generalized free fields

$$\begin{aligned} \mathcal{A} = \{ & \mathcal{O}(t_1, \Omega_1), \mathcal{O}(t_2, \Omega_2)\mathcal{O}(t_3, \Omega_3), \dots \\ & \times \mathcal{O}(t_m, \Omega_m)\mathcal{O}(t_{m+1}, \Omega_{m+1}) \dots \mathcal{O}(t_{m+n}, \Omega_{m+n}) \}, \end{aligned}$$

where $(t_i, \Omega_i) \in \mathcal{B}$. We put a cutoff on the number of operators that such a polynomial can contain by imposing $n \ll N$, where $N^2 \equiv (\frac{\ell_{\text{ads}}}{\ell_{\text{pl}}})^{d-1}$. Here, when we write $n \ll N$, we mean that in the limit where $N \rightarrow \infty$, n does not scale as any finite power of N . Note that, as a result of this cutoff, \mathcal{A} is, strictly speaking, just a set and not an algebra since it is not closed under multiplication. Nevertheless, we will continue to use the phrase “algebra” below although this caveat should be kept in mind.

Now, by solving the bulk equations of motion, the set of operators \mathcal{A} can be related to the algebra of truncated polynomials in the bulk fields ϕ in the region $\bar{\mathcal{D}}$ in Fig. 1. The explicit mapping is given in [5]. While respecting the constraints of bulk locality, we clearly cannot relate operators in \mathcal{D} to operators in \mathcal{B} since all points in \mathcal{D} are spacelike to all points in \mathcal{B} .

Nevertheless, it was shown in [5] that it is possible to construct operators in \mathcal{D} using operators in the band \mathcal{B} , provided that we consider *complicated* polynomials of generalized free fields that are *not* elements of \mathcal{A} . This can be done in two steps as follows. To be specific, consider an operator $\phi(r=0, t=0, \Omega)$ localized in the diamond \mathcal{D} at some point on the sphere. We will denote this operator by $\phi(0)$ to lighten the notation below. Now, this operator can be approximated as

$$\phi(0) \approx \sum_{n,m \ll N} c_{nm} |n\rangle \langle m|, \quad (2.2)$$

where the states $|n\rangle$ correspond to energy eigenstates in AdS and c_{nm} are c-number coefficients. At energies below the Planck scale, N , these energy eigenvalues are approximately quantized in units of the AdS scale. So far, in (2.2), we have not done anything except expand the operator in a complete set of states and place a cutoff since we are not interested in the matrix elements of $\phi(0)$ for ultra-Planckian energies.

Now, we note that the states $|n\rangle$ can be written as

$$|n\rangle = X_n |0\rangle, \quad (2.3)$$

where $X_n \in \mathcal{A}$. This means that all low-energy eigenstates in anti-de Sitter space can be created by acting on the vacuum with the algebra of *simple* polynomials of generalized free-field operators in the band, \mathcal{B} , or by simple polynomials of bulk fields in $\bar{\mathcal{D}}$. The reader may think of this as an application of the Reeh-Schlieder theorem [11] as applied to the set of simple bulk operators in the region $\bar{\mathcal{D}}$. However, in [5], this theorem was directly proved by using the properties of generalized free fields in a CFT and by considering the algebra of such fields in the band, \mathcal{B} . Explicit expressions for the operators, X_n , are also available in [5].

Although (2.3) is a surprising statement, so far there has been no violation of locality since (2.3) would hold even in a theory without gravity. However, we now note that we can also write

$$P_0 = |0\rangle \langle 0| = \lim_{\alpha \rightarrow \infty} e^{-\alpha H}.$$

Here, H is the Hamiltonian and given by

$$H = \int T^{00}(\Omega) d^{d-1} \Omega - E_0, \quad (2.4)$$

where

$$E_0 = \langle \Omega | \int T^{00}(\Omega) d^{d-1} \Omega | \Omega \rangle.$$

The key point is that, in a theory of gravity in the bulk, H is an operator in \mathcal{B} and the boundary value of a bulk operator in $\bar{\mathcal{D}}$. This property is evidently true in AdS/CFT since the stress tensor is a generalized free field in \mathcal{B} and is dual to the fluctuations of the graviton in the bulk. However, this fact is independent of AdS/CFT and relies on a fundamental property that emerges from the canonical quantization of gravity: the Hamiltonian, in any theory of quantum gravity, is a boundary term.

In fact, the projector above can be approximated using a *very complicated polynomial*. In particular, we can write

$$P_0 \approx \mathcal{P}_0 = \sum_{p=0}^{p_c} \frac{(\alpha_c H)^p}{p!}, \quad (2.5)$$

by choosing a particular large value of $\alpha = \alpha_c$ and expanding the exponential in a power series and cutting that off at p_c . We can see that a choice of $\alpha_c = \log(N)$ is sufficient to ensure that for the lowest possible nonzero energy eigenvalue, E_1 , we have $\mathcal{P}_0 |E_1\rangle = \mathcal{O}(\frac{1}{N})$. Second, an exponential e^x can be expanded in a power series as long as we keep $\mathcal{O}(x)$ terms. Therefore, if we choose $p_c = N \log(N)$, we ensure that this polynomial approximates the true projector on the vacuum, $\mathcal{P}_0 \approx P_0$, as long as it is inserted only in states with energy eigenvalues much smaller than N .

Putting these observations together, we obtain the following *formula*:

$$\phi(0) = \sum_{n,m \ll N} c_{nm} X_n \mathcal{P}_0 X_m^\dagger. \quad (2.6)$$

We thus see that we have succeeded in representing an operator in the center of AdS in terms of a complicated polynomial of operators that are uniformly spatially separated from the original operator. This is an important example, since it serves as a concrete prototype for the nonlocal effects that we expect are important for black hole evaporation.

We return to this toy model in VI showing how it also provides a toy model of various examples of the information paradox, which can then be resolved in this setting.

C. Breakdown of perturbation theory and locality

The example above shows how operators at one point in anti-de Sitter space can be represented entirely in terms of operators at other points. We can abstract two important elements from this analysis. The root of the nonlocality visible in the formula above lies in the Gauss law. The fact that the energy of a local insertion in the bulk can be measured at infinity leads to the fact that the Hamiltonian is purely a boundary term in gravity. This is what allowed us to construct the projector on the vacuum as a complicated polynomial of boundary operators in (2.5).

It is important to realize that, naively, the Gauss law seems to lead to very small nonlocal commutators that are suppressed by factors of $\frac{1}{N}$. The stress tensor, as it appears in (2.4) has a two-point function³ that is proportional to N^2 . In particular, the canonically normalized bulk graviton field is

³The precise form of this function is not relevant for our discussion but is given by [12]

$$\langle T_{ij}(x) T_{lm}(0) \rangle = \frac{N^2}{x^{2d}} I_{ij,lm}(x),$$

where the tensor $I_{ij,lm}(x)$ is given by

$$I_{ij,lm} = \frac{1}{2} (I_{il}(x) I_{jm}(x) + I_{im}(x) I_{jl}(x)) - \frac{1}{d} \delta_{ij} \delta_{lm},$$

and $I_{ij}(x) = \delta_{ij} - 2 \frac{x_i x_j}{x^2}$.

dual to $h_{ij} \leftrightarrow \frac{T_{ij}}{N}$. Therefore, the nonlocal commutator $i[H, \phi(t, r, \Omega)] = \dot{\phi}(t, r, \Omega)$ leads to nonlocal commutators between the canonically normalized graviton field and other bulk fields that are suppressed by $\frac{1}{N}$. It is important to note that the effect above, where we were able to completely rewrite a bulk operator in terms of other operators near the boundary, was obtained by suitably “enhancing” this $\frac{1}{N}$ effect to an order 1 effect. This can only happen when $\frac{1}{N}$ perturbation theory breaks down. This is why it is important that the polynomials that appear in (2.6) contain $O(N)$ insertions.

We thus find two underlying features in our toy model that lead to the nonlocality that is visible there. These are

- (1) Nonlocal perturbative commutators due to the Gauss law.
- (2) The enhancement of these small commutators due to the breakdown of perturbation theory.

Precisely the same analysis applies in flat space. As we review below, the Gauss law leads to commutators between two quasilocal operators that are suppressed by a power of $\frac{E}{M_{\text{pl}}}$, where E is a measure of the energy of the operators involved. The breakdown of perturbation theory may enhance these commutators to give rise to physically significant effects.

Recall that the Hamiltonian in gravity, in asymptotic flat space, can also be written as a boundary term. If we choose n^k to be the unit normal to the sphere at infinity, then we have

$$H = \frac{1}{16\pi G} \int_{\infty} d^{d-2} \Omega n^k (g_{ik,i} - g_{ii,k}),$$

where the repeated indices are summed over the spatial directions [13]. The Hamiltonian generates translations of asymptotic time.

Now consider studying a quasilocal operator in the interior of spacetime, $\phi(t, \vec{x})$, where we have separated the time t from spatial coordinates, \vec{x} . To define what we mean by the coordinates (t, \vec{x}) , we need to fix gauge or use a relational prescription. But provided that our prescription for localizing the operator satisfies the property that large diffeomorphisms that translate the asymptotic time coordinate by δt also lead to translations of the local time coordinate $t \rightarrow t + \delta t$, then it is guaranteed that $i[H, \phi(t, \vec{x})] = \frac{d\phi(t, \vec{x})}{dt}$. This commutator is nonlocal since the Hamiltonian can be defined by integration on a surface that is entirely spacelike to the point (t, \vec{x}) .

This is simply the Gauss law in action again. In fact, it is well-known that, in gravity, as a result of the Gauss law, there are *no* exactly gauge-invariant local operators. Nevertheless, it is possible to define quasilocal observables, since the nonlocality induced by the commutators above is small, as we now explain.

Note that, in terms of the canonically normalized fluctuations of the metric, $g_{\mu\nu} = \eta_{\mu\nu} + \sqrt{8\pi G} \tilde{h}_{\mu\nu}$, the expression for the Arnowitt-Deser-Misner Hamiltonian can be written as

$$H = \frac{1}{2\sqrt{8\pi G}} \int_{\infty} d^{d-2} \Omega n^k (\tilde{h}_{ik,i} - \tilde{h}_{ii,k}),$$

where we recognize $\sqrt{8\pi G} = \ell_{\text{pl}}^{\frac{d-2}{2}}$. Therefore, purely on dimensional grounds, we expect that if we smear the metric fluctuations to obtain a unit-normalized operator and consider its commutator with another unit-normalized operator, then this commutator will be suppressed by $(\frac{E}{M_{\text{pl}}})^{\frac{d-2}{2}}$, where E is a measure of the energy of the configuration of the two operators. Commutators between other field operators (not the metric) may be suppressed by further factors.

The precise commutators depend on how we define our quasilocal operators. Equivalently, the precise nonlocal commutators induced by the Gauss law depend on a choice of gauge. But, a concrete example was analyzed in [14], and we can use their results to verify our expectations. The authors of [14] worked in $d = 4$ so we will use this value below and then indicate the generalization to arbitrary d . With the choice made in [14], the authors found that the commutator between two quasilocal scalar operators outside the light cone was given by

$$[\phi(t, \vec{x}), \phi(t', \vec{y})] = \frac{-iG}{8} (\dot{\phi}(t, \vec{x}) \partial_i \phi(t', \vec{y}) + \partial_i \phi(t, \vec{x}) \dot{\phi}(t', \vec{y})) \times \frac{x^i - y^i}{((t-t')^2 + (\vec{x}-\vec{y})^2)^{\frac{1}{2}}}.$$

We emphasize that the precise form of the commutator above depends on the *choice* made in [14] of how to localize the operator. We remind the reader that this is similar to quantum electrodynamics, where our choice of how we place the Wilson lines on a local charged field controls the commutators of that field with local currents.

To estimate the size of this commutator evaluated in [14], we smear both fields to generate operators with a unit-normalized two-point function. We choose

$$\phi(f) = \int \phi(t, \vec{x}) f(t, \vec{x}) dt d^3 \vec{x};$$

$$\phi(g) = \int \phi(t, \vec{x}) g(t, \vec{x}) dt d^3 \vec{x},$$

with the constraint that $\langle \phi(f) \phi(f) \rangle = \langle \phi(g) \phi(g) \rangle = 1$. This leads to simple constraints on the functions f and g ,

$$\int \frac{f(t, \vec{x}) f(t', \vec{y}) d^3 \vec{x} d^3 \vec{y} dt dt'}{(t-t')^2 + (\vec{x}-\vec{y})^2} = 1,$$

$$\int \frac{g(t, \vec{x}) g(t', \vec{y}) d^3 \vec{x} d^3 \vec{y} dt dt'}{(t-t')^2 + (\vec{x}-\vec{y})^2} = 1.$$

These conditions are calculated at leading order since we have put in the leading order two-point function for ϕ but they can be corrected order by order in perturbation theory if required. We also demand that the two functions have no points of common support: $f(t, \vec{x}) \neq 0 \Rightarrow g(t, \vec{x}) = 0$ and $g(t, \vec{x}) \neq 0 \Rightarrow f(t, \vec{x}) = 0$. The expectation value of the commutator then becomes

$$\langle [\phi(f), \phi(g)] \rangle = i \left(\frac{E}{M_{\text{pl}}} \right)^2, \quad d = 4$$

where the “energy” of this configuration of operators is defined through

$$E^2 \equiv \int \frac{(t-t')(\vec{x}-\vec{y})^2}{((t-t')^2 + (\vec{x}-\vec{y})^2)^{\frac{d}{2}}} f(t, \vec{x}) g(t', \vec{y}) d^3 \vec{x} d^3 \vec{y} dt dt', \quad (2.7)$$

in $d = 4$. The expression is not covariant due to various choices made in defining the operators; these choices can be thought of as a choice of gauge. We have used the term “energy” for this quantity because it is a measure of the inverse distance scale in this configuration of operators.

In arbitrary d , we expect an entirely analogous result to hold

$$\langle [\phi(f), \phi(g)] \rangle = i \left(\frac{E}{M_{\text{pl}}} \right)^{d-2},$$

with E defined in analogy to (2.7), up to numerical prefactors, and with the exponent of $\frac{d}{2}$ in the denominator replaced by $\frac{2d-1}{2}$.

The fact that, for separations much larger than the Planck length, we have $E \ll M_{\text{pl}}$ tells us that the nonlocality induced by the Gauss law is small and explains why we observe approximate locality in nature.

However, the perturbative parameter for gravitational perturbation theory in flat space is also $(\frac{E}{M_{\text{pl}}})^{d-2}$.⁴ This suggests that the limits in which perturbation theory breaks down in flat space may also be interesting from the point of view of the loss of locality. We caution the reader that unlike the case of empty AdS above, we will not be able to demonstrate this effect explicitly in flat space. However, we believe that it is extremely likely that, at least in some settings, a combination of the fundamental nonlocality induced by the Gauss law, and the breakdown of gravitational perturbation theory leads to large-scale nonlocal effects. We now describe why this is also natural from a path-integral viewpoint.

⁴In any specific calculation, it may be more convenient to choose a definition of the coupling constant that differs from this by an $O(1)$ numerical prefactor. However, what is important here is that if all the distances are scaled by λ , the gravitational interactions fall by λ^{2-d} .

D. Locality in path integrals and perturbation theory

A consideration of the quantum-gravity path integral provides another heuristic argument for the claim that the breakdown of locality is concomitant with the breakdown of perturbation theory. In the Euclidean theory, we can consider some quasilocal observables $\phi(t_i^E, \vec{x}_i)$ at some value of Euclidean time, t_i^E and position \vec{x}_i . As mentioned above, to define these observables precisely, we need to choose a gauge or a relational prescription. We can then imagine inserting these operators into a path integral to compute a Euclidean correlation function. For example,

$$\begin{aligned} \langle \phi(t_1^E, \vec{x}_1) \dots \phi(t_n^E, \vec{x}_n) \rangle \\ = \int e^{-S} \phi(t_1^E, \vec{x}_1) \dots \phi(t_n^E, \vec{x}_n) \mathcal{D}g_{\mu\nu} \mathcal{D}\phi, \end{aligned} \quad (2.8)$$

where we integrate over all bulk metrics with some specified asymptotic boundary conditions.

Now, if we want this multilocal correlator to conform to some notions of locality, then we need a notion of when two points (t_1^E, \vec{x}_1) and (t_2^E, \vec{x}_2) are close to each other. Such a notion is predicated on a metric. However, in the path integral, we only specify asymptotic boundary conditions for the metric.

Nevertheless, it is possible to define an approximate notion of locality when the path integral is dominated by a *saddle point*. In the saddle point approximation, some particular metric $g_{\mu\nu}^0$ dominates the path integral, and this metric allows us to define the distance between two points. We may continue to use this metric to specify our notion of locality, provided that the correlator in (2.8) can be computed in an asymptotic series expansion about this saddle point.

However, if perturbation theory breaks down in the computation, this is a sign that the saddle point approximation to the path-integral has broken down. In this case, either the quantity (2.8) is dominated by another metric saddle point, $\tilde{g}_{\mu\nu}^0 \neq g_{\mu\nu}^0$, or else perhaps it cannot be computed in a saddle point approximation at all. In either case, our original notion of locality, which was predicated on the metric $g_{\mu\nu}^0$ is invalidated.

Therefore, from a path-integral point of view, it is natural for the breakdown of the perturbation theory in gravity to be a signal of the loss of locality. This analysis also helps us emphasize that, in a nongravitational theory, where the metric does not fluctuate there is no such link between perturbative breakdown and the loss of locality. It is *only in a theory of gravity* that locality is tied to the dominance of a particular background metric as a saddle point in the path integral, which in turn is tied to the validity of perturbation theory.

E. Boundary vs bulk locality

The path-integral analysis also helps us clarify another issue. It is very important that the effects we describe here

do *not* lead to a violation of boundary causality. For example, in AdS/CFT, the boundary theory satisfies micro-causality, and locality in the boundary theory is *not* lost even if we consider arbitrarily high-point correlators. Rather, our claim is that very high-point correlators may not have a simple *bulk* local interpretation.

The situation is similar in flat space. In doing the path integral in (2.8), we keep the asymptotic geometry fixed, and we do not expect to violate asymptotic notions of locality, as defined in [15]. For example, we may consider the situation where we take the points (t_i, x_i) to either future or past null infinity: \mathcal{I}^\pm . On both \mathcal{I}^\pm , we can specify these points through a null coordinate—which we denote by u_i and v_i , respectively—and a point on the sphere at infinity Ω_i . Then the points (u_1, Ω_1) and (u_2, Ω_2) are spacelike to each other if $\Omega_1 \neq \Omega_2$. In this situation, for instance, we expect that

$$\langle [\phi(u_1, \Omega_1), \phi(u_2, \Omega_2)] \phi(u_3, \Omega_3) \dots \phi(u_n, \Omega_n) \rangle = 0,$$

even as $n \rightarrow \infty$.

So, we expect that this commutator vanishes even if it is inserted in a correlator with an arbitrarily large number of insertions.

Our point, in this paper, is simply that if we try and define quasilocal operators that are *not* asymptotic operators, then it is possible that even approximate notions of locality for such operators may break down with the breakdown of perturbation theory. With this motivation, we now turn to a detailed study of the breakdown of string perturbation theory.

III. BOUNDS ON PERTURBATION THEORY

We will start our analysis of the breakdown of perturbation theory by reviewing some simple bounds on the rate of growth of tree amplitudes in perturbation theory. Consider the scattering of $\frac{n}{2}$ particles. (We assume n is even.) Unitarity of the S matrix tells us that

$$\sum_f \int d\Pi_f |M(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{f\})|^2 = 2\text{Im}[M(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_1 \dots k_{\frac{n}{2}}\})].$$

Here, $d\Pi_f$ is the measure on phase space, and the sum over f schematically indicates the sum over all possible final states.

Since the left-hand side is a sum over positive terms, we can restrict the sum to the case where the final state also consists of only $\frac{n}{2}$ particles to obtain an inequality. This leads to

$$\int d\Pi_{\frac{n}{2}} |M(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_{\frac{n}{2}+1} \dots k_n\})|^2 \leq 2\text{Im}[M(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_1 \dots k_{\frac{n}{2}}\})]. \quad (3.1)$$

In a theory with coupling constant g , we can expand the amplitude as

$$M(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_{\frac{n}{2}+1} \dots k_n\}) = \sum_{\ell=0}^{\infty} g^{n-2+2\ell} M^\ell(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_{\frac{n}{2}+1} \dots k_n\}),$$

where ℓ is the loop order. Later, the relevant coupling constant will turn out to be the string coupling g_s but for now, we do not need to specify any particular value for g . While perturbation theory is valid, the inequality (3.1) must then hold *order by order* in perturbation theory. Within this perturbative expansion, the first few terms simply lead to some positivity constraints but the first nontrivial term in the inequality (3.1) is

$$\int d\Pi_{\frac{n}{2}} g^{2n-4} |M^0(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_{\frac{n}{2}+1} \dots k_n\})|^2 \leq 2g^{2n-4} \text{Im}[M^{\frac{n-2}{2}}(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_1 \dots k_{\frac{n}{2}}\})],$$

where on the right-hand side, we have a high-order loop amplitude with $\frac{n-2}{2}$ loops. Now, the validity of perturbation theory requires that loop amplitudes be smaller than tree amplitudes. From this criterion, we obtain the relation

$$g^{2n-4} \text{Im}[M^{\frac{n-2}{2}}(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_1 \dots k_{\frac{n}{2}}\})] \leq g^{n-2} |M^0(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_1 \dots k_{\frac{n}{2}}\})|.$$

We emphasize that this is a very *weak* condition because we have included factors of the coupling constant in the inequality. A much higher power of the coupling constant appears on the left-hand side. This relation can fail to hold only if a huge factor in the amplitude overcomes this large power of the coupling constant. When this happens, the perturbation theory breaks down. Combining the relations above, we find that

$$\int d\Pi_{\frac{n}{2}} g^{2n-4} |M^0(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_{\frac{n}{2}+1} \dots k_n\})|^2 \leq 2g^{n-2} \text{Im}[M^0(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_1 \dots k_{\frac{n}{2}}\})].$$

Defining

$$M^{\text{tr}}(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_{\frac{n}{2}+1} \dots k_n\}) = g^{n-2} M^0(\{k_1 \dots k_{\frac{n}{2}}\} \rightarrow \{k_{\frac{n}{2}+1} \dots k_n\}),$$

which is just the tree level amplitude including all powers of the coupling, we see that this relation becomes

$$\int d\Pi_{\frac{n}{2}} |M^{\text{tr}}(\{k_1 \dots k_{\frac{n}{2}} \rightarrow \{k_{\frac{n}{2}+1} \dots k_n\})|^2 \leq 2 |M^{\text{tr}}(\{k_1 \dots k_{\frac{n}{2}}\})|. \quad (3.2)$$

At the point where the bound is violated, we expect that perturbation theory breaks down, and all orders in the perturbative answer become as important as the tree-level answer to the S matrix.

To give a concrete form to this inequality, we also need the phase-space factor. We will consider massless particles so that the phase space factor is simply given by

$$d\Pi_{\frac{n}{2}} = \frac{(2\pi)^d \delta(\frac{nE}{2} - \sum |k_l|) \delta^{d-1}(\sum k_l)}{(n/2)!} \prod_l \frac{d^{d-1} k_l}{(2\pi)^{d-1} 2|k_l|}. \quad (3.3)$$

Note the factor of $(\frac{n}{2})!$ which appears in the denominator. This simply arises from the conventional normalization of scattering amplitudes. Here, E is the center-of-mass energy per particle and $l, t \in \{\frac{n}{2} + 1, \dots, n\}$.

It is not too difficult to work out the total volume of phase space, which we do in Appendix C. The result is

$$\int d\Pi_{\frac{n}{2}} = v \frac{E^{\frac{(d-2)n-d}{2}}}{(n/2)!}. \quad (3.4)$$

Here,

$$v = \frac{2\pi \Gamma(\frac{d}{2} - 1)^{\frac{n}{2}} (\frac{n}{2})^{n(\frac{d}{2}-1)-d}}{(4\pi)^{(n-2)\frac{d}{2}} \Gamma(\frac{(d-2)(n-2)}{4}) \Gamma(\frac{(d-2)n}{4})}. \quad (3.5)$$

In this paper, we are concerned with the situation where, at large n , tree amplitudes grow as

$$M^{\text{tr}}(k_1 \dots k_{\frac{n}{2}} \rightarrow k_{\frac{n}{2}+1} \dots k_n) = \frac{n!}{\Lambda^{\frac{(d-2)n-d}{2}}}. \quad (3.6)$$

Here, Λ is a physically important energy scale, and its appearance on the right-hand side can be understood through dimensional considerations. Momentum eigenstates are normalized as $\langle k|k' \rangle = (2\pi)^{d-1} (2|k|) \delta^{d-1}(k-k')$. The amplitude is given by the overlap of an in-state with $n/2$ particles with an out-state with $n/2$ particles. In our analysis, we do *not* include the overall momentum conserving delta function in the amplitude. Hence, the mass dimension of the amplitude is $d - \frac{n(d-2)}{2}$, and the power of Λ ensures that the right-hand side also has the correct dimension. We will return to the significance of Λ below.

When (3.6) holds, we see that the inequality (3.2) is violated at a value of n that satisfies

$$v \frac{(n!)^2 E^{\frac{(d-2)n-d}{2}}}{\Lambda^{(d-2)n-2d}} \frac{1}{(n/2)!} = \frac{2n!}{\Lambda^{\frac{(d-2)n-d}{2}}},$$

or

$$\frac{v}{2} \frac{(n!)}{(n/2)!} \left(\frac{E}{\Lambda}\right)^{\frac{(d-2)n-d}{2}} = 1.$$

Keeping only the leading terms and using Stirling's approximation for the factorial: $\log(n!) \approx n \log(n) - n + O(\log(n))$, we find that perturbation theory breaks down at a value of n that satisfies

$$\frac{(2-d) \log \frac{E}{\Lambda}}{\log(n)} = 1 + O\left(\frac{1}{\log(n)}\right).$$

We see that although the prefactor v grows exponentially in n , it turns out to be irrelevant in the final answer because the dominant terms grow factorially with n . More precisely, we have $\log(v)/(n \log n) \rightarrow 0$ in the large n limit.

The breakdown of perturbation theory at large n is not specific to string theory or gravity. In fact, it is well-known in ordinary quantum field theories. For example, amplitudes grow factorially even in the $\lambda\phi^4$ theory in four dimensions. Moreover, we note that two particles with momentum components of $O(E)$ can be added to the amplitude at the cost of an additional propagator that contributes a term of $O(\frac{1}{E^2})$ and a single coupling constant factor λ . Therefore, the energy scale that enters (3.6) is $\Lambda = E/\sqrt{\lambda}$. Note that Λ depends on the energy per particle. These arguments suggest that perturbation theory breaks down for $n = O(1/\lambda)$ in $d = 4$ and this is indeed the result that was found in [16].

In Secs. IV and V, we will now argue that tree amplitudes in string theory also display at least a factorial growth, where the energy scale in (3.6) is $\Lambda = M_{\text{pl}} = \ell_{\text{pl}}^{-1}$. The fact that this energy scale appears can be seen easily since the coupling constant in string theory is g_s . In the units that we have adopted here, we have $2\pi g_s = \ell_{\text{pl}}^{\frac{d-2}{2}}$. A factor of g_s^{n-2} appears with each n point tree-level string amplitude. When dimensions are restored, this is equivalent to a factor of $\ell_{\text{pl}}^{\frac{(d-2)(n-2)}{2}}$. We then find that string perturbation theory breaks down for values of n and E that satisfy

$$\frac{\log(E \ell_{\text{pl}})}{\log(n)} = \frac{1}{2-d} + O\left(\frac{1}{\log(n)}\right). \quad (3.7)$$

If we take the energy per particle to scale with n as $E \propto \frac{1}{n^\gamma}$ then this threshold can also be written as

$$\frac{\log(g_s^2)}{\log(n)} = (d-2)\gamma - 1 + O\left(\frac{1}{\log(n)}\right).$$

It is important that we take $\gamma > 0$ (for reasons that we explain in Sec. IV D) and from the relations above, we see that we must also take $\gamma < \frac{1}{d-2}$.

A. Calculation with compact extra dimensions

Before we proceed to string amplitudes, we would like to describe some simple extensions of the bound above. One situation that is of physical interest is when the string theory lives in d dimensions but some of these dimensions are compactified. For simplicity, we take m of the extra dimensions to be compactified on a torus, where each side has radius ρ ; it is easy to generalize our calculation to more general compact manifolds. We consider a regime where $\frac{1}{E} \gg \rho \gg \sqrt{\alpha'}$ but we do not scale ρ with n . We also define $p = d - m$, the number of noncompact dimensions.

In general, as we will see below, the estimate for the amplitude is not altered by the compactification. This is because that the estimate for the growth of the amplitude relies on the volume of moduli space that depends on the structure of the world sheet and not on whether the spacetime is compactified. In Sec. V, we will also estimate the amplitude through a sum of the solutions of the scattering equations. But the fact that the number of solutions to the scattering equations grows factorially is independent of whether some of the spacetime dimensions are compact or not.

However, in the compact extra dimensions case, the volume of phase space given above in (3.4) is modified. The momentum in the compact extra dimension is quantized, and string theory also generates new winding-sector states. However, in the regime under consideration, the lowest mass of a state in the winding sector is $m^2 = \frac{\rho^2}{\alpha'^2} \gg \frac{1}{\alpha'}$; so these states are heavier than the lightest string excitations and can consistently be ignored. Including the Kaluza-Klein excitations, the new volume of phase space can be calculated as follows:

$$\int d\Pi_{\frac{d}{2}}^m = \sum_{n_{qt}} \int \left[(2\pi)^p \delta\left(\frac{nE}{2} - \sum_i k_i^0\right) \delta^{p-1} \times \left(\sum_i k_i^i\right) \frac{(2\pi\rho)^m \delta^m(\sum_i n_{qt})}{(n/2)!} \times \prod_i (2\pi) \delta\left(k_i^2 - \sum_{q=1}^m \frac{n_{qt}^2}{\rho^2}\right) \frac{d^p k_i}{(2\pi)^p (2\pi\rho)^m} \right],$$

where the components of the momentum, k_i in the compact directions are specified by n_{qt} , where $q = 1 \dots m$; the measure $d^p k_i$ runs over the noncompact dimensions including time and the $(p - 1)$ -dimensional delta function runs over the spatial noncompact dimensions. The delta function imposing momentum conservation in the compact directions is, of course, a discrete delta function. We have also placed a superscript m on the measure on the volume of phase space to indicate that m dimensions are compact.

We are not aware of any method of evaluating this integral and sum exactly. However, fortunately, in the limit under consideration where $\frac{1}{\rho} \gg E$, the sum is dominated by

the term with $n_{qt} = 0$. In this limit, the volume of phase space is given by

$$\int d\Pi_{\frac{d}{2}}^m = \frac{v_m}{(2\pi\rho)^{m\frac{d}{2}-m}} \frac{E^{\frac{(p-2)n}{2}-p} (1 + O(E\rho))}{(n/2)!},$$

where

$$v_m = \frac{2\pi\Gamma(\frac{p}{2} - 1)^{\frac{d}{2}} (n/2)^{n(\frac{p}{2}-1)-p}}{(4\pi)^{(n-2)\frac{p}{4}} \Gamma(\frac{(p-2)(n-2)}{4}) \Gamma(\frac{(p-2)n}{4})}.$$

Repeating the analysis above, we find that perturbation theory breaks down for

$$\frac{\log(E\ell_{pl,p})}{\log(n)} = \frac{1}{2-p} + O\left(\frac{1}{\log(n)}\right),$$

where $\ell_{pl,p}$ is the p -dimensional Planck length that is related to the d -dimensional Planck length through $\ell_{pl,p}^{p-2} (2\pi\rho)^m = \ell_{pl}^{d-2}$.

Thus, we see that our d -dimensional bound on the validity of string-perturbation theory naturally generalizes to a lower-dimensional bound, when some of the dimensions are compactified.

B. Other combinations of incoming and outgoing particles

In the analysis above, we considered a process where $\frac{n}{2}$ particles were incoming and $\frac{n}{2}$ were outgoing. Of course, it is also possible to consider other processes such as $2 \rightarrow n$ scattering. We do not consider these combinations in this paper to avoid any possible complications that may arise if one of the ingoing or outgoing particles has trans-Planckian energy. However, we emphasize that, assuming that the factorial growth outlined above continues to hold, these kinematical configurations would not yield a bound that is any stronger than the bound above. This is easy to see as follows. Consider scattering from $\alpha n \rightarrow \beta n$ particles, where α, β are some fractions. Then we see that the unitarity bound above is saturated at

$$\frac{v}{2} [((\alpha + \beta)n)!]^2 \left(\frac{E}{\Lambda}\right)^{(d-2)\beta n - d} = (2\alpha n)! (\beta n)!$$

Simplifying this expression and disregarding all subleading terms, we see once again that this leads to a breakdown of the perturbation theory at

$$\frac{(2-d) \log \frac{E}{\Lambda}}{\log(n)} = 1 + O(\log(n)),$$

which is precisely the same expression as the one above.

IV. ANALYTIC ESTIMATES OF THE GROWTH OF STRING AMPLITUDES

In this section, we will argue that n point tree-level scattering amplitudes of massless states in closed bosonic string theory as well as in type II superstring theories grow at least as fast as $n!g_s^{n-2}$ for large n . Our argument is based on the formulation of string scattering amplitudes as integrals over the moduli space of punctured Riemann surfaces. We then provide some evidence that these moduli-space integrals are dominated by the volume of moduli space, which allows us to utilize results from the mathematics literature on these volumes. Since the analysis for the bosonic string and the superstring is similar, we provide several details for the bosonic string and then briefly describe the generalization to the superstring.

A. Closed bosonic string amplitudes

Scattering amplitudes in closed bosonic string theory can be formulated as integrals over the moduli space of punctured Riemann surfaces. This representation may be somewhat unfamiliar to the reader, since the textbook formulation of string scattering leads to a formula for the amplitude where the positions of the vertex operators are integrated over a nonsingular world sheet [17]. So we first review the equivalence of the two prescriptions.

We claim that the string scattering amplitude may be written as

$$M(k_1 \dots k_n) = \mathcal{N} g_s^{n-2+2g} \int [dm] \frac{\det \langle \mu_\alpha, \phi_\beta \rangle}{(\det \langle \phi_\alpha, \phi_\beta \rangle)^{\frac{1}{2}}} (\det' P_1^\dagger P_1)^{\frac{1}{2}} \times (\det' \Delta)^{-\frac{d}{2}} Q_n P_n \bar{P}_n. \quad (4.1)$$

Here, the integral $[dm]$ runs over the moduli space of a Riemann surface with n punctures and genus g . ϕ_α are holomorphic quadratic differentials on the punctured Riemann surface, and μ_α are Beltrami differentials that parametrize infinitesimal motion on the moduli space. We denote this moduli space by $\mathcal{M}_{g,n}$, and when we need to refer to the Riemann surface itself, we use $M_{g,n}$. Apart from this measure, we also have the standard ghost determinant $(\det' P_1^\dagger P_1)^{\frac{1}{2}}$ and a power of the determinant of the scalar Laplacian, $(\det' \Delta)^{-\frac{d}{2}}$, which arises when we do the path integral over the world sheet fields. Here, Q_n , P_n , \bar{P}_n are terms that come from the correlation functions of vertex operators, which are placed at the punctures, and we give explicit expressions for these terms below when the vertex operators correspond to massless particles.

The main point that we would like to emphasize in this formula is that the integral over the positions of the vertex operators has been absorbed into the integral over the positions of the punctures of the Riemann surface. The equivalence of this prescription to the textbook Polyakov prescription is somewhat subtle because the punctured

Riemann surface cannot be mapped to the unpunctured surface by a nonsingular Weyl transformation.

Nevertheless, it was shown in [18,19] that the formulation of string scattering on the moduli space of the unpunctured Riemann surface, $\mathcal{M}_{g,0}$ and the moduli space of the punctured Riemann surface $\mathcal{M}_{g,n}$ indeed gives rise to equivalent answers. We start by dividing the holomorphic quadratic differentials into two sets: (ϕ_A, ϕ_i) , where the differentials ϕ_A are holomorphic on the surface $M_{g,0}$ and ϕ_i which are meromorphic on $M_{g,0}$ with poles at the positions of the punctures. We can choose these two sets of differentials to have no overlap so that $\langle \phi_A, \phi_i \rangle = 0$. Similarly, we can divide the Beltrami differentials into two sets: μ_i that move the punctures and μ_A that change the other moduli. It is not difficult to see that we can also choose $\langle \mu_i, \phi_A \rangle = \langle \mu_A, \phi_i \rangle = 0$.

This then leads to the expression

$$M(k_1 \dots k_n) = \mathcal{N} g_s^{n-2+2g} \int [dm'] \prod d^2 z_i \frac{\det \langle \mu_A, \phi_B \rangle}{(\det \langle \phi_A, \phi_B \rangle)^{\frac{1}{2}}} \times \frac{\det \langle \mu_i, \phi_j \rangle}{(\det \langle \phi_i, \phi_j \rangle)^{\frac{1}{2}}} (\det' P_1^\dagger P_1)^{\frac{1}{2}} (\det' \Delta)^{-\frac{d}{2}} Q_n P_n \bar{P}_n,$$

where we have divided the integral over the moduli, $[dm] = [dm'] \prod d^2 z_i$, into an integral over the moduli of the surface $M_{g,0}$ and an integral over the positions of the punctures. At this point, the operator P_1 , defined on a world sheet with metric g_{ab} , as

$$(P_1 \delta V)_{ab} = \frac{1}{2} (\nabla_a \delta V_b + \nabla_b \delta V_a - g_{ab} \nabla_c \delta V^c), \quad (4.2)$$

still acts only on those vector fields that *vanish* at the positions of the punctures.

The main result of [18], which was further clarified in [19], was that when the functional determinants are appropriately regulated, we have

$$\frac{\det \langle \mu_i, \phi_j \rangle}{\det \langle \phi_i, \phi_j \rangle^{\frac{1}{2}}} (\det' P_1^\dagger P_1)^{\frac{1}{2}} = (\det' \tilde{P}_1^\dagger \tilde{P}_1)^{\frac{1}{2}},$$

where \tilde{P}_1 is the same operator as (4.2) but with a domain that includes vector fields that do not vanish at the positions of the punctures. Therefore, on the right-hand side above, we have the usual determinant that would have resulted from integrating out the ghost fields on the surface $M_{g,0}$. This reduces the expression (4.1) to the more familiar expression, which only involves quantities on the unpunctured Riemann surface,

$$M(k_1 \dots k_n) = \mathcal{N} g_s^{n-2+2g} \int [dm'] \prod d^2 z_i \frac{\det \langle \mu_A, \phi_B \rangle}{\det \langle \phi_A, \phi_B \rangle^{\frac{1}{2}}} \times (\det' \tilde{P}_1^\dagger \tilde{P}_1)^{\frac{1}{2}} (\det' \Delta)^{-\frac{d}{2}} Q_n P_n \bar{P}_n.$$

However, the advantage of the expression (4.1) is that it allows us to make contact with various results in the mathematical literature. On the punctured surface, we can choose the so-called ‘‘hyperbolic metric’’ on the world sheet so that, everywhere on the world sheet, we have uniform scalar curvature: $R = -1$. Note that it is possible to make this choice even for the tree-level amplitude because, for the n -punctured sphere, the Gauss Bonnet theorem reads

$$\int \sqrt{g} R d^2 z = 2\pi\chi(M_{g,n}) = 2\pi(2 - 2g - n).$$

Therefore, even for $g = 0$, there is no obstruction to choosing $R = -1$. This just implies that the area of the world sheet is $\int \sqrt{g} d^2 z = 2\pi(n + 2g - 2)$.

We note that near a puncture at $z = 0$, the hyperbolic metric behaves like

$$ds^2 \xrightarrow{z \rightarrow 0} \frac{dzd\bar{z}}{|z|^2 \log(|z|)^2}. \quad (4.3)$$

With this choice of metric on the world sheet, the measure on moduli space turns into the Weil-Petersson measure [20]

$$d\mu_{\text{WP}} = [dm] \frac{\det\langle \mu_\alpha, \phi_\beta \rangle}{\det\langle \phi_\alpha, \phi_\beta \rangle^{\frac{1}{2}}},$$

where the inner product between the quadratic differentials is taken with respect to the hyperbolic metric.

We now specialize to tree-level scattering so that we set $g = 0$. At tree level, our expression for the string scattering amplitude now becomes

$$M^{\text{tr}}(k_1 \dots k_n) = \mathcal{N} g_s^{n-2} \int d\mu_{\text{WP}} (\det' P_1^\dagger P_1)^{\frac{1}{2}} (\det' \Delta)^{-\frac{d}{2}} Q_n P_n \bar{P}_n, \quad (4.4)$$

where $d\mu_{\text{WP}}$ is the Weil-Petersson measure on the moduli space of the n -punctured sphere.

The final ingredient that we need is the correlation function of vertex operators. For massless states in the closed bosonic string theory, at tree-level, these correlators are easy to write down explicitly. Moreover, they are Weyl invariant by themselves and so take on the same form when the metric on the world sheet is hyperbolic, as they do when the world sheet is flat. For the closed bosonic string, we recall that the vertex operators for massless states are

$$\mathcal{V}(z, \bar{z}) = (\epsilon \cdot \partial X)(\bar{\epsilon} \cdot \bar{\partial} X) e^{ik \cdot X},$$

where we have also specified the polarization vectors $\epsilon, \bar{\epsilon}$, and the antiholomorphic polarization vector is just the complex conjugate of the holomorphic polarization vector.

Physically, this means that we are considering the scattering of *linear combinations* of the graviton and the dilaton.

The relevant correlation function can then be written

$$\langle \mathcal{V}(z_1, \bar{z}_1) \dots \mathcal{V}(z_n, \bar{z}_n) \rangle = Q_n P_n \bar{P}_n,$$

where

$$Q_n = \exp\left(-\frac{1}{2} \sum_{i \neq j} k_i \cdot k_j G_{ij}\right),$$

$$P_n = \mathcal{L} \left\{ \exp\left(\sum_{i \neq j} \frac{1}{2} \epsilon_i \cdot \epsilon_j \partial_i \partial_j G_{ij} + k_i \cdot \epsilon_j \partial_j G_{ij}\right) \right\}. \quad (4.5)$$

\bar{P}_n is the antiholomorphic counterpart of P_n . G_{ij} is the world sheet Green’s function,

$$G_{ij} = \ln |z_i - z_j|^2. \quad (4.6)$$

The symbol \mathcal{L} above is shorthand for a rule that instructs us to expand the exponential in P_n and keep only the part that is linear in all the polarization tensors.

1. Infrared divergences

The moduli-space integral receives divergent contributions from the boundaries of moduli space, where some closed geodesic on the Riemann surface pinches off and its length goes to zero. In the case of tree-level amplitudes, this corresponds to the situation where two punctures collide.

These divergences can be regulated through a suitable $i\epsilon$ prescription as explained in [21]. Equivalently, as explained in [22], one may divide the moduli-space integral into two regions: (1) the region that covers those Riemann surfaces that can be obtained by plumbing lower-dimensional surfaces together using the ‘‘plumbing fixture’’ and (2) the rest of the moduli space, which can be understood as coming from a fundamental higher-point vertex. The divergences mentioned above come from the first region of moduli space. Here, one can get rid of them by using field theoretic techniques to rewrite the integral as a sum over contributions coming from the propagation of intermediate particles. The full amplitude can then be obtained by including the contribution from the second part of moduli space where the integral is finite.

However, both of these prescriptions will complicate our estimate of the size of the integral. So, here, we will follow a simple-minded procedure by regulating these divergences by placing a cutoff, ℓ_0 , on the length of the smallest simple closed geodesic on $M_{0,n}$. This cuts off moduli space near its boundaries, and we will work in this cutoff moduli space.

We caution the reader that it is possible that this procedure is not justified. For example, it may happen that the contributions from the edges of moduli space

cancel off the contributions from the bulk of moduli space that we will focus on below. These cancellations are possible even though, in our analysis, we have chosen polarization vectors for the external particles that will make the integrand on the bulk of moduli space positive. But if we use the $i\epsilon$ prescription of [21], this instructs us to adopt a contour in a complexified version of moduli space near the edges, and then the integrand is no longer positive. For these reasons, it would be nice to repeat the arguments below without this cutoff.

Keeping these caveats in mind, we now examine each of the terms that appear in (4.4).

B. Volumes of Weil Petersson moduli spaces

We will argue below that the main contribution to the growth of amplitude comes simply from the volume of moduli space

$$V_{g,n} = \int d\mu_{\text{WP}}.$$

Volumes of the Weil-Petersson moduli spaces of punctured Riemann surfaces have been studied in the mathematics literature. For the n -punctured sphere, this volume was first calculated in [23]. Then, numerical techniques were used to advance conjectures for the growth of these volumes at large genus [24]. In [25], an analytic recursion relation was developed to calculate the Weil-Petersson volumes for any value of n, g . The asymptotic growth of this volume was then studied in [26].

These papers found that, for any fixed n , when g becomes large

$$V_{g,n} = (4\pi^2)^{2g+n-3} (2g-3+n)! \frac{1}{\sqrt{g\pi}} \left(1 + \mathcal{O}\left(\frac{1}{g}\right) \right). \tag{4.7}$$

We do not need the subleading terms and are only interested in the leading asymptotic behavior, which is given by

$$\lim_{g \rightarrow \infty} \frac{\log(V_{g,n})}{(2g+n) \log(2g+n)} = 1, \tag{4.8}$$

and holds when *either* n or g become large.

Note that putting $n = 0$ in (4.8) we have

$$\lim_{g \rightarrow \infty} V_{g,0} \propto (2g)!. \tag{4.9}$$

Here, as in the rest of this paper, when we use the symbol \propto , we mean that we have captured the leading growth of the physical quantity. For example, in (4.9), we have dropped factors that may even grow exponentially with g since these factors are subleading compared to the factorial growth.

This large- g growth in the volume of moduli space was independently obtained in the physics literature by using matrix model techniques in [27]. By combining these results with the analysis of [28], it is possible to show that this growth implies that the vacuum amplitude in the closed bosonic string theory also grows as $(2g)!$ for large g . It is well-known that this growth implies that nonperturbative effects in string theory appear with a strength of $\mathcal{O}(e^{-1/g_s})$. At weak coupling, this is larger than the size of non-perturbative effects in ordinary quantum field theory, which is expected to be $\mathcal{O}(e^{-1/g_s^2})$. These stringy nonperturbative effects are related to the existence of d -branes in string theory.

In our case, we are interested in the growth of the volume at large n with $g = 0$, which is given by

$$\lim_{n \rightarrow \infty} \frac{\log(V_{0,n})}{n \log(n)} = 1 + \mathcal{O}\left(\frac{1}{\log(n)}\right),$$

or in simpler notation

$$\lim_{n \rightarrow \infty} V_{0,n} \propto n!, \tag{4.10}$$

where we have kept only the leading part of the growth and dropped other factors, including those that are merely exponential in n .

We will now argue that the full scattering amplitude is dominated by the factorial growth (4.10) since the other terms in the string scattering amplitude have subfactorial growth.

C. Bounds on functional determinants

The functional determinants that appear in (4.4) can be related to special values of the Selberg zeta function and its derivatives. The Selberg zeta function is defined as

$$Z(s) = \prod_{\Theta} \prod_{k=1}^{\infty} (1 - e^{-(s+k)\ell_{\Theta}}), \quad \text{Re}(s) > 1.$$

Here the product labeled by Θ is over all simple closed geodesics on the Riemann surface. The length of Θ, ℓ_{Θ} , is measured with respect to the hyperbolic metric on the Riemann surface. In terms of $Z(s)$, we have [29]

$$\begin{aligned} \det' \Delta &= e^{-c_0 \chi} \left(\frac{dZ}{ds} \right)_{s=1}, \\ \det'(P_1^\dagger P_1) &= e^{-c_1 \chi} Z(2), \end{aligned} \tag{4.11}$$

where χ is the Euler characteristic of the Riemann surface under consideration. The constants c_n are $\mathcal{O}(1)$ numbers given by

$$\begin{aligned}
 c_\alpha &= \sum_{0 \leq m \leq \alpha - \frac{1}{2}} [(2\alpha - 2m - 1) \log(2\alpha - m)] \\
 &\quad - \left(\alpha + \frac{1}{2}\right)^2 + 2(\alpha - [\alpha]) \left(\alpha + \frac{1}{2}\right) \log(2\pi) \\
 &\quad + 2\zeta'(-1). \tag{4.12}
 \end{aligned}$$

In particular, we have $c_0 \approx -0.58$ and $c_1 \approx -1.89$.

In most of the moduli space, we expect the Selberg zeta function to be well behaved. However, it is important to bound this term near the boundaries of moduli space and ensure that it cannot affect the factorial growth of the amplitude.

The behavior of the Selberg zeta functions near the boundaries of moduli space, where the Riemann surface degenerates was studied in [28–30]. The basic method of estimating the behavior of the Selberg zeta function near the boundaries of the moduli space is as follows. Let Θ be a simple closed geodesic which gets pinched when two given punctures approach each other on the world sheet and consider the collar region around Θ , which is defined as

$$C(\Theta, r) = \{z : \rho(z, \Theta) < r\},$$

where $\rho(z, \Theta)$ is the hyperbolic distance of a point z from Θ , and r is referred to as the radius of the collar.

It turns out that as the length of this geodesic, $\ell \rightarrow 0$, the geometry of the Riemann surface excluding a collar of radius $r = \log(\ell^{-1})$ remains uniformly bounded. The geodesics, which only remain in this part of the Riemann surface and do not intersect Θ , are not affected by the degeneration of Θ , and their contribution to the Selberg zeta function is simply a constant in the limit where $\ell \rightarrow 0$. However, the lengths of the geodesics, which happen to intersect Θ , tend to infinity since these now have to cross the collar region. As a result, their contribution to $Z(s)$ is simply 1 in the limit where $\ell \rightarrow 0$.

This leaves behind the degenerating geodesic itself, and its inverse, which has the same length. Their contribution to the infinite product can be calculated explicitly. This analysis allows one to estimate the behavior of the Selberg zeta function and its derivatives as the moduli space degenerates [29], and one finds that

$$Z(s) \xrightarrow{\ell \rightarrow 0} \ell^{-2s+1} e^{-\frac{\pi^2}{3\ell}}.$$

In our case, we recall that we have bounded the length of the smallest simple geodesic on the Riemann surface by an n -independent constant ℓ_0 . This means that near the boundaries of this cutoff moduli space

$$Z(2) \propto \ell_0^{-3} e^{-\frac{\pi^2}{3\ell_0}},$$

up to an ℓ_0 independent constant. From the formulas above, we can also derive the behavior of $Z'(1)$ near the boundary of the cutoff moduli space, which is given by

$$Z'(1) \propto \ell_0^{-1} e^{-\pi^2/3\ell_0}.$$

Our interest in these formulas is restricted to the fact that, in the cutoff moduli space, the Selberg zeta functions that appear in (4.11) are bounded away from 0 and ∞ by finite constants.

Now note that the only other terms that appear in (4.11) are $e^{-c_0\chi}$ and $e^{-c_1\chi}$. While these vary exponentially with n , since $\chi = 2 - n$, this behavior is subleading compared to the factorial growth of the volume of moduli space. In fact, since $c_0 < 0$ and $c_1 < 0$, both determinants in (4.11) decay with n .

As far as the ghost determinant is concerned, the main result that we are interested in is that it is bounded below and does not decay as rapidly as a factorial. However, we see that it also does not grow as a factorial and that

$$\frac{\log\{\det'(P_1^\dagger P_1)\}}{n \log(n)} = 0 + \mathcal{O}\left(\frac{1}{\log(n)}\right),$$

in the cutoff moduli space. On the other hand, to bound the amplitude from below we only need the result that the determinant of the scalar Laplacian is bounded above so that its inverse does not decay as rapidly as a factorial. However, we have the stronger result that

$$\frac{\log\{\det'(\Delta)\}}{n \log(n)} = 0 + \mathcal{O}\left(\frac{1}{\log(n)}\right),$$

in the cutoff moduli space.

Therefore, the functional determinants in (4.11) do not alter the factorial growth shown in (4.10).

D. Bounds on Green's functions

We now argue that the term Q_n also does not affect the factorial growth of the amplitude.

Once again, we start by ensuring that its contribution from the boundaries of moduli space is bounded. The region on the Riemann surface near the degeneration locus, where two punctures on the world sheet collide, is conformally equivalent to a twice punctured disc. In [31], it was shown that on a twice punctured disc $\mathcal{D} \setminus \{-a, a\}$, the length, ℓ of the smallest closed geodesic separating the punctures $\{-a, a\}$ from the boundary of the disc $\partial\mathcal{D}$ is given by

$$\ell = \frac{2\pi^2}{\log(1/a)},$$

as the punctures coalesce, i.e., as $a \rightarrow 0$. (See also [32].)

Using this result, and the formula for the world sheet Green's functions (4.6), we see that on the cutoff moduli space, G_{ij} can then be bounded as

$$|G_{ij}\ell_0| \leq C,$$

near the region where the punctures i, j come close and where C is a constant, which is independent of the number of punctures, n .

It is also of interest to determine the contribution of the Green's function between two generic punctures on the punctured Riemann surface. We can set up holomorphic coordinates on the punctured Riemann surface by setting the hyperbolic metric to be $ds^2 = e^\rho dz d\bar{z}$. The function ρ must have the appropriate logarithmic singularities at the positions of the punctures where the metric behaves like (4.3). Furthermore, we can fix the isometries of this metric by demanding that three of the punctures be at $z = 0$, $z = 1$ and $z = i$.

Then, noting that the area of the Riemann surface is $2\pi(n-2)$, at a generic point in moduli space, we expect that we will have $|\log|z_i - z_j|| = O(\log(n))$ for two generic punctures.

Next, note the energy per particle scales as $E \propto \frac{1}{n^\gamma}$. Therefore, the factor of $k_i \cdot k_j$ in the exponent of Q_n has a magnitude of order $\frac{1}{n^{2\gamma}}$. Second, although there are $O(n^2)$ terms in the exponent of Q_n , the factor of $k_i \cdot k_j$ does not have a definite sign. More precisely, correlations between these terms come only from the momentum conserving delta function, but this single constraint tends to become unimportant in the large n limit. The Green's function G_{ij} also has varying signs on the moduli space, although it is bounded and has typical magnitude $\log(n)$ as explained above.

Hence, we expect that the sum of $O(n^2)$ terms, each of the size $O(\log(n)n^{-2\gamma})$, will only contribute an $O(\log(n)n^{1-2\gamma})$ factor to the magnitude of the exponent, at least at most points in the moduli space.

These arguments suggest that

$$\frac{\log|\log Q_n|}{(1-2\gamma)\log(n)} \leq 1 + \mathcal{O}\left(\frac{\log(\log(n))}{\log(n)}\right),$$

on a significant fraction of the moduli space. This can be rewritten as

$$|\log(Q_n)| \leq q \log(n)n^{1-2\gamma}, \quad (4.13)$$

where q is some $O(1)$ factor on a significant part of moduli space.

From (4.13) we see that, for any value of $\gamma > 0$, the possible suppression due to Q_n is subleading compared to the factorial growth of the volume of the moduli space. Note that, in this section, we have been somewhat heuristic. We have also not bounded Q_n everywhere in the cutoff moduli space, as we were able to do for the term involving the functional determinants. In particular, we have not ruled out the possibility that Q_n might become very large on some parts of moduli space, which might enhance the factorial growth of the amplitude.

Nevertheless, insofar as the issue of proving a *lower bound* on the growth of the moduli-space integral in the cutoff moduli space is concerned, we believe that these arguments are correct. Our conclusions above are also verified very nicely by the numerical calculations of Sec. V. In particular, we direct the reader to Fig. 4, which shows that (4.13) provides an excellent fit to our numerical results when $\log(Q_n)$ is evaluated at the saddle points of the moduli-space integral.

This analysis suggests that the factor of Q_n does not modify the factorial growth of the amplitude that comes from the volume of moduli space.

E. Prefactors

We are left with the prefactors P_n and \bar{P}_n . These prefactors contain, within themselves, both a factorial number of terms and also a product of n Green's functions. Thus, in principle, this prefactor could either grow factorially or suppress the overall factorial growth of the volume.

A heuristic, and indirect, argument that this prefactor does not alter the factorial growth of the amplitude is as follows. By unitarity, we know that massless amplitudes appear in the residues of the poles of tachyon amplitudes. If we can show that tachyon amplitudes grow factorially, then it is likely that these residues—and, hence, the massless amplitudes—also grow factorially and that the factors of P_n do not suppress this growth.

If we had been considering tachyon amplitudes, then the analysis of the volume of moduli space and functional determinants would have been just as in the previous sections. The analysis of the factor Q_n would also have been similar to Sec. IV D except that the typical components of the tachyon momenta would scale with $\frac{1}{\sqrt{\alpha}}$. This corresponds to $\gamma = 0$ in the notation above.

The analysis of this case is somewhat delicate since, on a large fraction of moduli space, we see from (4.13) that the suppression due to Q_n may itself be as strong as $\frac{1}{n!}$. However, we would still expect that, at least on an exponentially small fraction of moduli space, the inequality $|\log(Q_n)| < n \log(n)$ would hold. Since the volume of this fraction already grows factorially with n , this suggests that the final answer for tachyon amplitudes also grows factorially with n .

By the argument above, this suggests that the factors of P_n and \bar{P}_n do not suppress the factorial growth of massless amplitudes.

However, this argument is only suggestive and not entirely satisfying. Therefore, here, we will borrow a result from Sec. V. The numerical analysis of Sec. V shows that P_n does not decay as a factorial. The numerical analysis also suggests that P_n does not grow factorially but we are unable to entirely rule out this latter possibility due to subtleties in the numerics explained in Sec. V C.

Nevertheless, this result tells us that the factorial growth shown in (4.10) provides at least a *lower bound* on the rate of growth of string amplitudes.

F. Growth of tree amplitudes

Combining the results above, we have

$$\frac{\log(M^{\text{tr}}(k_1 \dots k_n))}{(n-2)\log(g_s) + n\log(n)} \rightarrow 1,$$

in the limit where $g_s \rightarrow 0$ and $n \rightarrow \infty$. Said differently, we see that

$$M^{\text{tr}}(k_1, \dots, k_n) \propto g_s^{n-2} n!, \quad (4.14)$$

where the \propto sign indicates that we have dropped other terms that do not grow as fast as $n!$. This is the result that we wanted to prove.

G. Superstring amplitudes

We now describe how this analysis can be extended to the superstring. The analysis is largely parallel to the bosonic string, so we will be brief and not repeat all of our steps. We will also confine ourselves to tree-level amplitudes to avoid subtleties with the superstring moduli space at higher genus.

We start with the formulation of the superstring scattering amplitude as

$$M^{\text{tr}}(k_1 \dots k_n) = \mathcal{N} g_s^{n-2} \int d\mu_{\text{WP}} (\det' P_{\frac{1}{2}}^\dagger P_{\frac{1}{2}})_{\frac{1}{2}}^{-1} (\det' P_1^\dagger P_1)^{\frac{1}{2}} \times (\det' \mathcal{D})^{\frac{d}{2}} (\det' \Delta)^{-\frac{d}{2}} Q_n P_n \bar{P}_n, \quad (4.15)$$

which is the natural supersymmetric generalization [33] of (4.1). The determinant of the world sheet Dirac operator arises from integrating out the world sheet fermions, and in addition, we also obtain a determinant by integrating out the superghosts. In the expression above, we have performed the integral over the odd moduli, which leaves behind the expression $Q_n P_n \bar{P}_n$. The only integral that remains is over the positions of the punctures. We have used notation to use the same symbols P_n, \bar{P}_n for the world sheet correlators as in the bosonic string but in the case of the superstring, the values of these correlators are different. This should not cause any confusion and which expression for P_n needs to be used should be clear from the context.

We now describe this world sheet correlator in more detail. We define

$$V(\theta, \chi, z) = \exp(i\theta\chi\epsilon \cdot \partial X + \theta k \cdot \psi - \chi\epsilon \cdot \psi), \quad (4.16)$$

which depends on the momentum k^μ , a polarization vector ϵ^μ , and two auxiliary Grassmann variables θ, χ [34]. Then,

we recall that the vertex operators for NS-NS sector operators in the type II superstring with polarization tensor $\epsilon_{\mu\nu} = \epsilon_\mu \epsilon_\nu$ ⁵ are given in the $(-1, -1)$ -picture and the $(0, 0)$ picture by

$$\mathcal{V}^{(-1,-1)}(z, \bar{z}) = e^{-\phi-\bar{\phi}} e^{ik \cdot X} \int d\bar{\chi} d\chi V(\theta=0, \chi, z) \times \tilde{V}(\bar{\theta}=0, \bar{\chi}, \bar{z}),$$

$$\mathcal{V}^{(0,0)}(z, \bar{z}) = e^{ik \cdot X} \int d\bar{\chi} d\bar{\theta} d\chi d\theta V(\theta, \chi, z) \tilde{V}(\bar{\theta}, \bar{\chi}, \bar{z}),$$

where the operator $e^{-\phi-\bar{\phi}}$ arises from the bosonized superconformal ghost insertions and \tilde{V} is defined in analogy to (4.16) with all the left-moving fields replaced with right-moving ones.

The n -point world sheet correlation function that we need is then obtained by inserting 2 $(-1, -1)$ picture operators and $(n-2)$ - $(0, 0)$ picture operators. This is given by

$$Q_n P_n \bar{P}_n = \langle \mathcal{V}^{(-1,-1)}(z_1, \bar{z}_1) \mathcal{V}^{(-1,-1)}(z_2, \bar{z}_2) \mathcal{V}^{(0,0)}(z_3, \bar{z}_3) \dots \times \mathcal{V}^{(0,0)}(z_n, \bar{z}_n) \rangle \\ = \frac{1}{|z_{12}|^2} \int \prod_{i=3}^n d\theta_i d\bar{\theta}_i \prod_{j=1}^n d\chi_j d\bar{\chi}_j \exp I \exp \bar{I},$$

where

$$I = \sum_{i \neq j} \frac{1}{2} k_i \cdot k_j \ln(z_i - z_j) - \frac{\theta_i \theta_j k_i \cdot k_j}{z_i - z_j} \\ + \frac{\theta_i - \theta_j}{z_i - z_j} (\chi_i \epsilon_i \cdot k_j + \chi_j \epsilon_j \cdot k_i) - \frac{\chi_i \chi_j \epsilon_i \cdot \epsilon_j}{z_i - z_j} \\ - \frac{\theta_i \theta_j \chi_i \chi_j \epsilon_i \cdot \epsilon_j}{(z_i - z_j)^2},$$

and it is understood that $\theta_1 = \theta_2 = 0$ in the expression for I .

Now the integral over the Grassmann variables may be done as follows. We separate the last term in I , which is quartic in the Grassmann variables, and then pull it down using a power series expansion of the exponent. Each power of this term soaks up some of the Grassmann integrals. The integral over the rest of the Grassmann variables is Gaussian and so can be performed in terms of a Pfaffian. The result, after this manipulation, can be seen to be $Q_n P_n \bar{P}_n$, where

⁵For this choice of polarization tensors, which corresponds to the scattering of linear combinations of the graviton and the dilaton, tree-level amplitudes in the type I theory are the same as tree-level amplitudes in the type II theories.

$$Q_n = \prod_{i \neq j} |z_i - z_j|^{k_i \cdot k_j},$$

$$P_n = \frac{1}{z_{12}} \left[\text{Pf}(M_{12}) + \sum_{\{k\}, \Pi_k} \left(\prod_{p=1}^k \frac{(\epsilon_{\pi_{2p-1}} \cdot \epsilon_{\pi_{2p}})}{(z_{\pi_{2p-1}} - z_{\pi_{2p}})^2} \right) \times \text{Pf}(M_{12\pi_1(\pi_1+n)\dots\pi_{2k}(\pi_{2k}+n)}) \right]. \quad (4.17)$$

Here, the matrix M is defined as

$$M = \begin{pmatrix} A & -C^T \\ C & B \end{pmatrix}, \quad (4.18)$$

and the matrices A , B , C are defined through

$$A_{ij} = \begin{cases} \frac{k_i \cdot k_j}{z_i - z_j} & i \neq j \\ 0 & i = j \end{cases}; \quad B_{ij} = \begin{cases} \frac{\epsilon_i \cdot \epsilon_j}{z_i - z_j} & i \neq j \\ 0 & i = j \end{cases};$$

$$C_{ij} = \begin{cases} \frac{\epsilon_i \cdot k_j}{z_i - z_j} & i \neq j \\ -\sum_p \frac{\epsilon_i \cdot k_p}{z_i - z_p} & i = j \end{cases},$$

and the notation $M_{i_1 \dots i_n}$ means that the rows i_1, \dots, i_n and the columns $i_1 \dots i_n$ are removed before taking the Pfaffian. The sum over $\{k\}$, π_k is over all distinct choice of k pairs from the range $(3 \dots n)$, where k runs from $1 \dots (n-2)/2$. The pairs themselves are specified by $\{(\pi_1, \pi_2) \dots (\pi_{2k-1}, \pi_{2k})\}$. For future reference, we note that the Pfaffian in the last term in the sum above, with $k = (n-2)/2$, can be simplified using

$$\text{Pf}(M_{1,2,3,3+n,\dots,n,2n}) = \frac{\epsilon_1 \cdot \epsilon_2}{z_{12}}, \quad (4.19)$$

since only the rows and columns $(n+1)$ and $(n+2)$ are left after the deletion above.

The rest of the analysis for the type II superstring amplitude is entirely parallel to the analysis for the bosonic string that we displayed above. In particular, the functional determinants that appear in the path integral above can again be related to special values of the Selberg zeta function [35] through

$$\det' P_{\frac{1}{2}}^\dagger P_{\frac{1}{2}} = \exp(-c_{\frac{1}{2}} \chi) Z\left(\frac{3}{2}\right),$$

$$\det' \mathcal{D} \mathcal{D} = \exp(-c_{-\frac{1}{2}} \chi) \frac{Z^{(2N)}(\frac{1}{2})}{(2N)!}.$$

Here, N is the number of zero modes of the Dirac operator, and the constants $c_{\pm\frac{1}{2}}$ are also given by (4.12). It is not difficult to see, by a simple extension of the analysis above, that these functional determinants are also bounded away

from factorial growth, and therefore do not affect the leading factorial behavior of the amplitude.

Since Q_n for the superstring has the same form as Q_n for the bosonic string, our bounds on this term that we analyzed for the bosonic string also apply here. To analyze the prefactors P_n and \bar{P}_n , the heuristic relation to tachyon amplitudes that was given for the bosonic string can also be utilized here. This is because the processes we are considering, in principle, also make sense in the type 0 string theories. The massless scattering amplitudes we are considering here can, therefore, be obtained by factorizing tachyon amplitudes in the type 0 theories. By using the bounds on functional determinants and the bound on Q_n above, we can argue that tachyon amplitudes in the type 0 theories grow factorially. This suggests that at the poles of these amplitudes, the residues, which include the massless amplitudes, also grow factorially.

However, this argument is not watertight and, with just these arguments, we cannot entirely rule out the possibility that the expansion of the Pfaffians leads to an additional term that also grows like $n!$ or alternately, becomes very small. With the help of numerical analysis, in the next section, we will be able to rule out the possibility that $P_n \bar{P}_n$ decays as rapidly as $\frac{1}{n!}$. The numerical analysis also suggests that these terms *do not* grow as rapidly as $n!$ but this conclusion is less robust for reasons that we detail below.

Assuming this property of P_n , we find that (4.14) holds for the superstring amplitude as well.

V. NUMERICAL ESTIMATES OF THE GROWTH OF STRING AMPLITUDES

In this section, we will turn to a numerical analysis of string scattering amplitudes. In our analysis above, we provided some evidence that string amplitudes grow at least as fast as $n!$ at large n . However, we were unable to deal precisely with the effect of the factor of $P_n \bar{P}_n$ that appears in (4.4) and (4.15). By numerically estimating the growth of string amplitudes in this section, we will verify the factorial growth in an entirely independent manner and also check that the factors of P_n and \bar{P}_n do not seem to change this behavior. Our conclusion that the factors of $P_n \bar{P}_n$ do not suppress the factorial growth is robust. However, our result that they do not enhance this growth is subject to some caveats as we describe in Sec. V C.

In this section, it will be convenient to go back to the choice of a flat world sheet metric. Therefore, our expressions for the bosonic and type II superstring amplitudes can both be written as

$$M^{\text{tr}}(k_1 \dots k_n) = 4\pi g_s^{n-2} \int \prod_{i=4}^n d^2 z_i Q_n P_n \bar{P}_n |G|, \quad (5.1)$$

where P_n and Q_n are given in (4.5) for the bosonic string and (4.17) for the type II superstring and G is the ghost contribution given by

$$\mathcal{G} = z_{12}^2 z_{23}^2 z_{13}^2. \quad (5.2)$$

Note that although the superstring also receives a contribution from the superghost insertions, we have already included them in (4.17). We have now also fixed the overall normalization of the scattering amplitude, which can be determined by unitarity [17].

We note that this integral runs over $(n - 3)$ -complex dimensions. Moreover, as we mentioned above, it suffers from divergences when $z_i \rightarrow z_j$. As we explained above, these divergences can be regulated systematically. One procedure was outlined in [21]—which suggested a suitable $i\epsilon$ prescription—and another procedure was described in [22]—which described how the region of moduli space that led to the divergences could be isolated and dealt with using field theoretic techniques. However, in practice, both of these prescriptions are nontrivial to implement on a computer. Therefore, we will not directly attempt to perform the integral in (5.1).

Instead, we are able to proceed further as follows. Upon considering the string integral, we find that, even though individual energies are small, $\log(Q_n)$ becomes large at large n . The exponent in Q_n has $O(n^2)$ terms, and while we expect these terms to cancel among each other, we still expect that the saddle is of order $n^{1-2\gamma}$ as shown in (4.13). Therefore, even in the situation where the individual energies are small, we can approximate the amplitude by localizing the moduli-space integral onto the points where $\log(Q_n)$ is maximized. This procedure is not only numerically efficient, it also has the advantage that it sidesteps the issue of divergences in the moduli-space integral.

Extremizing the exponent in Q_n leads us to focus on the values of z_i that satisfy

$$E_i = \sum_{j \neq i} \frac{k_i \cdot k_j}{z_i - z_j} = 0, \quad \forall i. \quad (5.3)$$

These equations were first discovered in [36] in the study of high-energy string scattering. However, they have recently turned out to be useful in the study of scattering in ordinary quantum field theories [37].

Note that solutions to the scattering equations are invariant under simultaneous $SL(2, C)$ transformations of the variables $z_i \rightarrow \frac{az_i + b}{cz_i + d}$ with $ad - bc = 1$. This invariance also appears in the string amplitude and can be gauge fixed by setting z_1, z_2, z_3 to definite values. Modulo this gauge invariance, it is easy to see that there are actually $(n - 3)!$ solutions of the scattering equations. This was proved in [38]. The proof is not difficult and proceeds by induction.

Assume that the number of distinct solutions to the scattering equations with $n - 1$ particles is $(n - 4)!$. Now consider the scattering equations with n particles. We can use the $SL(2, C)$ freedom to set $z_1 = \infty$. Then z_1 and k_1 drop out of the Eq. (5.3). We now deform the first and the

last momenta through $k_n \rightarrow \alpha k_n$, while simultaneously deforming $k_1 \rightarrow k_1 + (1 - \alpha)k_n$. The deformation of k_1 has no effect since we have set $z_1 = \infty$. As we take α from 1 to 0, we see that the αk_n drops out of the scattering equations, $(E_1 \dots E_{n-1})$, and this set of equations becomes an independent set of scattering equations for $n - 1$ particles. By assumption, this set has $(n - 4)!$ solutions. On the other hand, E_n is independent of α . This gives us a polynomial equation of order $(n - 3)$ for z_n that should have $(n - 3)$ roots. Therefore, for each of the $(n - 4)!$ solutions to the equations (E_1, \dots, E_{n-1}) , we now have $(n - 3)$ solutions for z_n leading to a total of $(n - 3)!$ solutions for the full system (E_1, \dots, E_n) . Now, as we deform α away from 0, and back towards 1, we can assume that these $(n - 3)!$ solutions move continuously in the complex plane leading to $(n - 3)!$ solutions for the original undeformed scattering equations.

It is rather remarkable that this is exactly the same number as the estimate of the volume of moduli space in (4.7) at $g = 0$ including even the subleading terms in n . Therefore, there is *one solution of the scattering equations per unit volume of moduli space*. We do not understand the reason for this phenomenon.

Localizing on the solutions to the scattering equations, and after performing the integral about the Gaussian fluctuations about these saddle points, our prescription for numerical evaluation of the amplitude becomes

$$M^{\text{tr}}(k_1 \dots k_n) = (4\pi g_s)^{n-2} \sum_{\{z_i\}} |\mathcal{G}| |\mathcal{J}|^{-1} Q_n P_n \bar{P}_n. \quad (5.4)$$

The sum goes over *all* inequivalent solutions of the scattering equations. Here, $|\mathcal{J}|^{-1}$ is the Jacobian factor that results from integrating fluctuations about the saddle points and is given by $\mathcal{J} = \det(\partial_j E_i)$. More explicitly, this matrix is

$$\partial_j E_i = \begin{cases} \frac{k_i \cdot k_j}{(z_i - z_j)^2}, & i \neq j \\ -\sum_{q \neq i} \frac{k_i \cdot k_q}{(z_i - z_q)^2}, & i = j, \end{cases} \quad (5.5)$$

and $i, j \in (4 \dots n)$. The overall factor multiplying the amplitude arises because the integral over Gaussian fluctuations yields $(4\pi)^{n-3} |\mathcal{J}|^{-1}$, and this factor combines with the $4\pi g_s^{n-2}$ in (5.1) to give $(4\pi g_s)^{n-2}$.

We see immediately that we can also write our estimate as

$$M^{\text{tr}}(k_1 \dots k_n) = (n - 3)! (4\pi g_s)^{n-2} |\mathcal{G}| \langle |\mathcal{J}|^{-1} Q_n P_n \bar{P}_n \rangle, \quad (5.6)$$

where by $\langle |\mathcal{J}|^{-1} Q_n P_n \bar{P}_n \rangle$, we mean the mean of this quantity across the set of solutions of the scattering equations. This mean can be estimated statistically by taking

a *sample* of the set of saddle points. Thus, by sampling over a large set of solutions to the scattering equations, we obtain an estimate for the full amplitude without having to find all $(n-3)!$ solutions to the scattering equations.

A. Brief description of algorithm

We now briefly describe the algorithms that we use here to solve the scattering equations and evaluate the integrand. We provide some additional details in Appendix A.

We choose a set of random external kinematics using a uniform measure in phase space. This can be done using standard algorithms [39] used to generate events in phenomenological calculations, as explained further in Appendix A.

Given a set of external momenta, the first task in evaluating (5.6) is to obtain solutions of the scattering equations. Our algorithm simply starts at a random point in $C^{(n-3)}$ (with some cutoffs that disallow very large initial values of z) and then uses a variant of the multidimensional Newton's method to seek the nearest solution. Multidimensional root-finding algorithms are not guaranteed to converge, so if the algorithm does not converge, it simply picks another starting point and flows to a nearby root. We caution the reader that although we pick our initial starting point using as uniform a distribution as possible, this does not mean that we are sampling the roots uniformly. This is because we do not have any prior estimates for the sizes and shapes of the basins of attraction of different roots.

After having obtained a root, our next task is to evaluate the summand in (5.6). The terms \mathcal{J}^{-1} and Q_n are straightforward to evaluate. Q_n can be obtained by summing over the $\frac{n(n-1)}{2}$ terms in its exponent. The Jacobian factor, \mathcal{J}^{-1} is a determinant of a $(n-3) \times (n-3)$ matrix, which can also be evaluated efficiently using LU decomposition [40]. However, note that P_n contains a large number of terms. Several of these terms are numerically expensive to evaluate, as we describe below. Therefore, in order to evaluate this term efficiently, we truncate the prefactor for the superstring as

$$P_n \bar{P}_n \approx \left| \frac{\text{Pf}(M_{12})}{z_{12}} \right|^2 + \sum_{\pi} \left| \frac{\epsilon_1 \cdot \epsilon_2}{z_{12}^2} \prod_{i=1}^{(n-2)/2} \frac{\epsilon_{\pi_{2i-1}} \cdot \epsilon_{\pi_{2i}}}{(z_{\pi_{2i-1}} - z_{\pi_{2i}})^2} \right|^2. \quad (5.7)$$

Here, we note that the last term comes from (4.19). Here, the sum over permutations runs over distinct pairings $(3, \dots, n) \rightarrow \{(\pi_1, \pi_2) \dots (\pi_{n-3}, \pi_{n-2})\}$.

For the bosonic string, we truncate the prefactor as

$$P_n \bar{P}_n \approx \left| \prod_i \sum_{j \neq i} \frac{\epsilon_i \cdot k_j}{z_i - z_j} \right|^2 + \sum_{\pi} \left| \prod_{l=1}^{\frac{n}{2}} \frac{\epsilon_{\pi_{2l-1}} \cdot \epsilon_{\pi_{2l}}}{(z_{\pi_{2l-1}} - z_{\pi_{2l}})^2} \right|^2. \quad (5.8)$$

Here, the sum over permutations runs over all distinct pairings of all n particles: $(1, \dots, n) \rightarrow \{(\pi_1, \pi_2) \dots (\pi_{n-1}, \pi_n)\}$.

There are two truncations involved here. First, we see that the truncation reduces the $O(n^2)$ terms that appear in (4.17) and (4.5) to two terms each. We do not expect this to make any difference to the factorial growth, which is our main interest. Moreover, we expect that these two terms are the most important terms in the amplitude. Naively, we might also expect that the term involving the polarization vectors is important at lower energies since it involves the fewest possible factors of the external momenta. On the other hand, we expect that the term involving the Pfaffian will be important at relatively higher energies for the superstring. Similarly, we expect that the term involving dot products of the momenta with the polarizations will be important at higher energies for the bosonic string.

However, more significantly, we have placed the permutation outside the absolute value sign in (5.7) and (5.8) even though the original expressions involved the absolute-value squared of the sum. This is because we expect that the other terms in the sum will suffer cancellations, and therefore this part—which involves the sum of absolute-value squares in which all terms are positive—is likely to dominate. We cannot verify this assumption directly in our numerics since, in order to do so, we would have to check cancellation in the remaining terms to $O(\frac{1}{n!})$, which is not feasible. It would be nice to understand this approximation better.

Empirically, we find for the bosonic string for $\gamma = 0$ (relatively higher energies), the first term in (5.8) starts dominating the second term after about $n \sim 60$. For $\gamma = \frac{1}{24}$ (relatively lower energies), the first term against starts dominating after about $n \sim 90$. From this analysis, it appears that, at large n , the string amplitude is well approximated by only the first term in (5.8). However, we direct the reader to the paragraph on “prefactor estimation” in Appendix A for a discussion of whether this is a genuine result or a numerical artifact.

For the superstring the second term in (5.7) becomes unimportant even earlier—after about $n \sim 30$ in the case where $\gamma = 0$ and after about $n \sim 70$ in the case where $\gamma = \frac{1}{8}$.

Although the prefactor P_n contains only $O(n^2)$ terms, the reason these terms are expensive to evaluate is as follows. Consider the term that involves the product of polarization vectors either in (5.7) or (5.8). We see that this involves a sum over all possible pairings of the polarization vectors. There are $\frac{n!}{2^{\frac{n}{2}}(n/2)!}$ distinct pairings, and so it cannot be evaluated directly. Therefore, to estimate this term, we sum over a set of random pairings and then multiply by the total number of possible pairings. This sampling process is expensive, and it would be very numerically expensive to carry it out for all the other terms in (4.17) and (4.5).

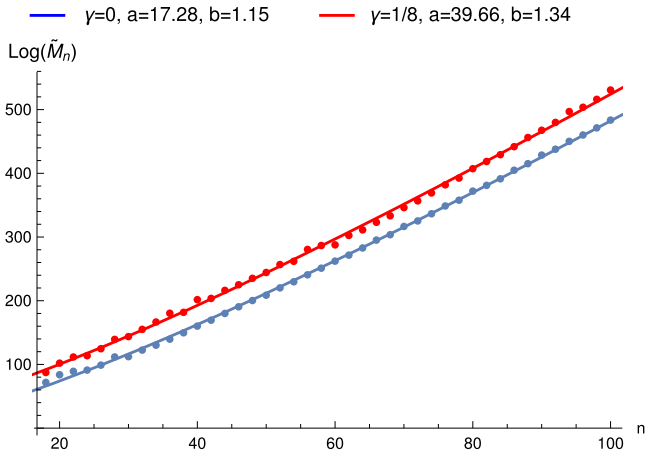


FIG. 2. Scattering amplitudes in the superstring for extreme values of γ . The solid lines show $a + bn + \log((n-3)!)$.

B. Results

We now describe the results of our numerical calculations. For the superstring, we evaluated 1000 solutions of the scattering equations for 500 different sets of random kinematics for each even value of $n \in [4, 100]$. This constitutes 2.25×10^7 distinct solutions of the scattering equations. For the bosonic string, we evaluated 500 solutions of the scattering equations for 500 different sets of random kinematics for each even value of $n \in [4, 100]$.⁶ This constitutes 1.13×10^7 distinct solutions of the scattering equations. Both computations together took about 8000 hours of CPU time.

The results of our calculations for the type II superstring are displayed in Fig. 2. In our calculation, we are interested in values of n that are as large as possible. Accordingly, in the figure, we only show data for $n \in [18, 100]$ and discard the data for smaller values of n . This choice of starting point is simply based on empirical considerations since we find that the data does not obey the same nice trend for smaller values of n . On the y axis, we plot

$$\log(\tilde{M}_n) = \log(M^{\text{tr}}(k_1 \dots k_n)) - (n-2) \log(4\pi g_s) + n \log(d-2), \quad (5.9)$$

where $d = 10$ is the critical dimension for the superstring. The reason that the last factor $n \log(d-2)$ appears is to compensate for the fact that when we dot a set of random polarization tensors with the amplitude, we expect to obtain a factor of $\frac{1}{(d-2)^n}$. Adding $n \log(d-2)$ to the logarithm of the amplitude strips off this irrelevant kinematic factor and gives us a better sense of the magnitude of the amplitude.

⁶Obviously, for $n = 4, 6, 8$, there are only 1, 6, and 120 distinct solutions.

We can fit the amplitude to the following expression:

$$\log(\tilde{M}_n) = a + bn + \log((n-3)!), \quad (5.10)$$

and the values of a, b for two extreme possible choices of γ are shown in Fig. 2. These values are not of any direct relevance to our calculation. Our emphasis is simply on the leading $\log((n-3)!)$ term, and the graphs show that it is an excellent fit to the amplitude. In particular, what Fig. 2 shows is that the factor of P_n does not suppress the amplitude or contribute another factorial term to the amplitude.

These results may be compared to the answer for the bosonic string, which we had previously discussed in [8]. The graphs for the bosonic string are reproduced below in Fig. 3. Once again, we plot $\log(\tilde{M}_n)$ as defined in (5.9) except that now we use $d = 26$, which is the critical dimension for the bosonic string. The best fit values for the fit (5.10), are also displayed in Fig. 3. Note that the values of a, b that appear here are slightly different from [8] because, in order to retain consistency with the superstring case, we have discarded data for $n < 18$ here whereas we had retained data for $n = 12, 14, 16$ in [8]. Once again, we see that the factorial growth, which can be understood as coming from the volume of moduli space or from the number of solutions to the scattering equations, provides an excellent approximation to the amplitude.

Note the various values of a, b are not comparable between the superstring and the bosonic string since the bosonic string amplitudes are evaluated in $d = 26$ whereas the superstring amplitudes are evaluated in $d = 10$. The different dimensionality leads to several subtle effects. For example, it alters the average value of kinematic invariants $k_i \cdot k_j$ even if we choose $|k_i|$ to scale in the same manner.

We can also use our numerical results to check our expectations about the growth of Q_n outlined in Sec. IV D. Note that in Sec. IV D, we were concerned with a generic

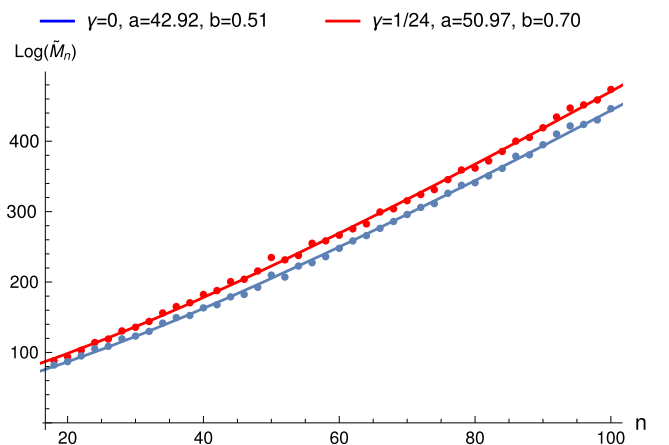


FIG. 3. Scattering amplitudes in the bosonic string for extreme values of γ . The solid lines show $a + bn + \log((n-3)!)$.

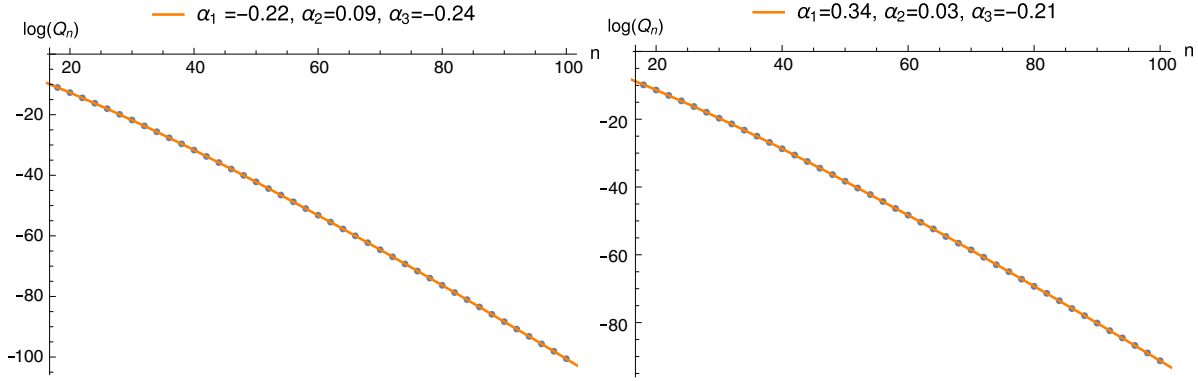


FIG. 4. Plot of $\langle \log Q_n \rangle$ vs n for the bosonic string (left) and the superstring (right). The best fit values $\alpha_1, \alpha_2, \alpha_3$ defined in (5.11) are also shown.

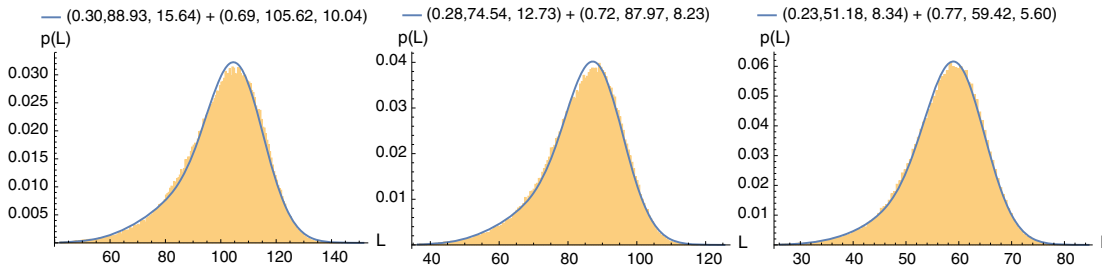


FIG. 5. Histograms for the distribution of amplitudes for $n = 100, 80, 50$, respectively, for the superstring. The legend shows the best fit sum of normal distributions as explained in the text.

point in moduli space whereas, by construction, our numerical results pertain to those points in moduli space where $\log(Q_n)$ is extremized. Nevertheless, in Fig. 4, we show $\langle \log(Q_n) \rangle$, averaged over our numerical samples for $\gamma = 0$. Choosing a nonzero value for γ would just change $\log(Q_n) \rightarrow \frac{1}{n^{2\gamma}} \log(Q_n)$. For $\gamma = 0$, we find that we can fit

$$\langle \log(Q_n) \rangle = \alpha_1 + \alpha_2 n + \alpha_3 n \log(n). \quad (5.11)$$

We see from Fig. 4 that, with the appropriate choice of parameters, this is a perfect fit both for the bosonic string and for the superstring. This *precisely* justifies our analytic expectations from Sec. IV D

It is a fact that, for every solution of the scattering amplitudes $\log(Q_n) < 0$. Physically, this is a sign of the good high-energy properties of string theory since it indicates that as we scale up the momentum, $Q_n \rightarrow 0$. On the other hand, from the point of view of the scattering equations, viewed as a set of polynomial equations, it is somewhat miraculous that the roots always lead to a negative value for $\log(Q_n)$. This property is certainly not true everywhere on the moduli space. It would be nice to understand this property directly from the equations.

We now turn to a discussion of the possible errors in this result.

C. Errors

The main possible source of error in our answers arises as follows. We find, empirically, that our numerical samples obey a log-normal distribution. This is shown in Fig. 5, where we have plotted a histogram of the values obtained for the superstring for

$$L = \log(|\mathcal{G}||\mathcal{J}|^{-1} Q_n P_n \bar{P}_n) + n \log(d - 2).$$

These values cover all the solutions to the scattering equations that we generated for $n = 50, 80, 100$, respectively. As mentioned above, each histogram covers 500,000 distinct solutions of the scattering equations, since we generated 1000 solutions for 500 distinct sets of external momenta. It is convenient to combine all these sets into a single set for the discussion here. Histograms for the same quantity and for the same values of n are shown for the bosonic string in Fig. 6 although, in this case, each histogram covers 250,000 samples for each n .

In each case, we find that the distribution of the amplitude is well-approximated by a sum of log-normal distributions. More precisely, the values of L obtained in our numerical results are distributed according to the probability distribution

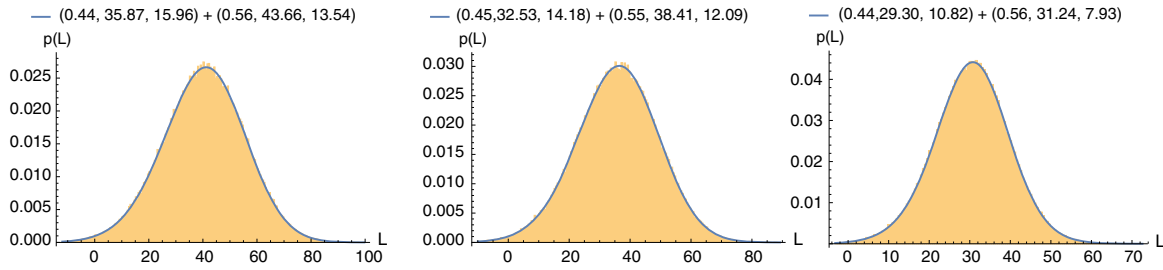


FIG. 6. Histograms for the distribution of amplitudes for $n = 100, 80, 50$, respectively, for the bosonic string. The legend shows the best fit sum of normal distributions as explained in the text.

$$p(L)dL = \frac{\alpha_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(L-\mu_1)^2}{2\sigma_1^2}} + \frac{\alpha_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{(L-\mu_2)^2}{2\sigma_2^2}}. \quad (5.12)$$

In Figs. 5 and 6, the parameters that appear in the sum of distributions indicated above are displayed using the notation $(\alpha_1, \mu_1, \sigma_1) + (\alpha_2, \mu_2, \sigma_2)$.

A log-normal distribution (or a sum of such distributions) creates difficulties for numerical sampling. For example, if the distribution in (5.12) had been an exact distribution for L then we would have had

$$(4\pi g_s)^{2-n} \langle \tilde{M}_n \rangle = \int e^L p(L) dL = \alpha_1 e^{\mu_1 + \frac{\sigma_1^2}{2}} + \alpha_2 e^{\mu_2 + \frac{\sigma_2^2}{2}}. \quad (5.13)$$

However, the probability distribution itself takes on an extremely small value at the point where this mean is attained. In fact, the probability that a random variable distributed according to (5.12) will be *above* some given value λ is given by

$$\mathcal{P}(\lambda) = \int_{\lambda}^{\infty} p(L) dL = 1 - \frac{\alpha_1}{2} \operatorname{erfc}\left(\frac{\mu_1 - \lambda}{\sqrt{2}\sigma_1}\right) - \frac{\alpha_2}{2} \operatorname{erfc}\left(\frac{\mu_2 - \lambda}{\sqrt{2}\sigma_2}\right).$$

The numerical values of this probability for λ given by (5.13), with the parameters for the distribution as given in Figs. 5 and 6, are given below.

n	$\mathcal{P}((4\pi g_s)^{2-n} \langle \tilde{M}_n \rangle)$ (superstring)	$\mathcal{P}((4\pi g_s)^{2-n} \langle \tilde{M}_n \rangle)$ (bosonic string)
100	1.44×10^{-15}	4.44×10^{-16}
80	5.26×10^{-11}	4.57×10^{-13}
50	1.03×10^{-5}	2.09×10^{-8}

What this table indicates is that to obtain an accurate value for the mean of the amplitude, using random sampling we would need to obtain about 10^{15} samples for the superstring and 10^{16} samples for bosonic string.

Clearly, this is not feasible using direct sampling. Therefore, it is clear that a better algorithm is required to identify the points where the amplitude is the largest, and sample those accurately rather than simply using random sampling as we have done here. We are not aware of any such algorithm at the moment, and developing such an algorithm remains an important challenge to complete this program of numerically estimating string amplitudes.

An order of magnitude estimate of the number of samples that would be required to accurately sample a log-normal distribution with mean μ and standard deviation σ is given by $e^{\frac{\sigma^2}{2}}$. In Fig. 7, we provide a plot of the standard deviation of our sampled values of L for the bosonic string

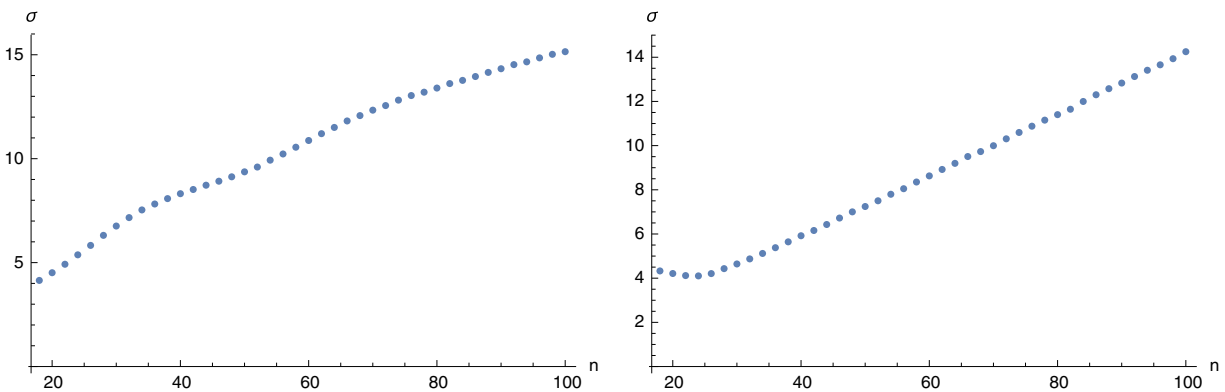


FIG. 7. Standard deviation of the distribution of sampled values of L for the bosonic string (left) and the superstring (right).

and for the superstring, respectively. These graphs are provided for the case $\gamma = 0$.

The fact that this standard deviation does not grow very rapidly, indicates that this sampling error may not affect the leading factorial growth that we have found.

We emphasize that random numerical sampling tends to significantly *underestimate* the mean for log-normal distributions and not overestimate them. Therefore, in any case, our numerical results establish that the amplitude grows *at least* factorially. However, it is clearly important to bound these sampling errors more precisely and check whether the growth is exactly factorial or whether it is larger.

D. Comments on pure gravity

We conclude this section with some comments on scattering amplitudes in pure gravity. It appears to us that, surprisingly, numerically estimating amplitudes in a theory of pure gravity may be *harder* than estimating amplitudes in string theory. This is because although scattering amplitudes in gravity can also be localized to solutions of the scattering equations, pure-gravity amplitudes are given by holomorphic functions of the roots. The holomorphicity of the amplitude creates difficulties for numerical sampling, as opposed to the bosonic string and the type II superstring, where the amplitude can be expressed as a sum of real positive terms, at least for specific choices of the external polarizations.

Note that the pure-gravity amplitude is given by a formula that is very similar to the formula for the superstring (5.4). In particular, we have

$$M^{\text{tr}}(k_1 \dots k_n) = (4\pi g_s)^{n-2} \sum_{\{z_i\}} \mathcal{G} \mathcal{J}^{-1} (P_n^{\text{grav}})^2. \quad (5.14)$$

On the right-hand side, the terms match almost precisely with the superstring. We have $\mathcal{J} = \det(\partial_j E_i)$, which is shown explicitly in (5.5). The ghost term is also as defined in (5.2). Moreover,

$$P_n^{\text{grav}} = \frac{1}{z_{12}} \text{Pf}(M_{12}),$$

where the matrix M is defined in (4.18). We remind the reader that M_{12} means that we remove the first two rows and columns. We note this is precisely the first term that appears in P_n for the superstring in (4.17).

Of course, $P_n^{\text{grav}} \neq P_n$ and also the Koba-Nielsen factor Q_n that appears in (5.4) does not appear in gravity. But an even more important difference between the superstring answer and the gravitational answer is that the superstring answer involves absolute values whereas the expression (5.14) does not have absolute values, either on \mathcal{G} or on \mathcal{J} . Moreover, we simply have the square of the term P_n^{grav} in the pure-gravity answer and not its absolute-value squared as we have in the superstring. This is the reason that the pure-gravity answer must be obtained as a limit of the “ambitwistor string” rather than the usual superstring [41].

The difficulty with parsing the holomorphic answer numerically is that $(n - 3)!$ contributions that come from the distinct solutions to the scattering equations now involve several numerical cancellations since they are not simply the sum over a set of positive terms. This makes the sum difficult to evaluate numerically.

For this reason, we have not carried out our numerical calculations for pure gravity. Nevertheless, we do expect that even in pure gravity the answer for the amplitude will scale like $n!$. In fact, this leading factorial growth may also be seen directly from the Britto-Cachazo-Feng-Witten (BCFW) recursion relations [42].

The BCFW recursion relations state that the n -point tree-level amplitude can be decomposed as a sum over products of lower point tree-level amplitudes. Schematically, we have

$$M_n^{\text{tr}} = \sum_{p=1}^{n-3} \frac{M_{p+2}^{\text{tr}} M_{n-p}^{\text{tr}}}{P_L^2}.$$

This recursion relation arises as follows. We mark two particular legs for performing the BCFW extension. Of the remaining $n - 2$ legs, we must select p legs on one side and $n - p - 2$ legs on the other. On one side, we get a $p + 2$ -point scattering amplitude after adding one of the original marked legs and an intermediate leg. On the other side, we similarly get a $n - p$ point amplitude. This recursion relation is valid for $n \geq 4$.

To estimate the growth of the amplitude, we just want to calculate the number of terms in the BCFW recursion relations. Denoting this number by N_n , we see that it satisfies the recursion relations

$$N_n = \sum_{p=1}^{n-3} \binom{n-2}{p} N_{p+2} N_{n-p}.$$

We define $N_3 = 1$, and we have $N_n = 0$ for $n \leq 3$.

Now, we consider the following generating function:

$$W(g) = \sum_{n=4}^{\infty} \frac{N_n g^{n-2}}{(n-2)!}.$$

With the conventions above, but recalling that the recursion relation is valid only for $n \geq 4$, we then find that

$$W(g) = \sum_{n=4}^{\infty} \sum_{p=1}^{n-3} \frac{N_{p+2} g^p}{p!} \frac{N_{n-p} g^{n-p-2}}{(n-p-2)!}.$$

We write $\tilde{p} = p + 2$, and $\tilde{n} = n - p$, so that the sum above can be written as

$$W(g) = \sum_{\tilde{p}=3}^{\infty} \sum_{\tilde{n}=3}^{\infty} \frac{N_{\tilde{p}} g^{\tilde{p}-2}}{(\tilde{p}-2)!} \frac{N_{\tilde{n}} g^{\tilde{n}-2}}{(\tilde{n}-2)!} = (W(g) + g)^2.$$

We can now solve for

$$W(g) = \frac{1 - 2g - \sqrt{1 - 4g}}{2}.$$

This provides an analytic formula for N_n ,

$$N_n = (-1)^{n-1} \frac{1}{2} 4^{n-2} \binom{\frac{1}{2}}{n-2} (n-2)!.$$

Thus, we see that N_n also grows factorially with n .

Of course, even in the BCFW analysis there are cancellations, and therefore, this estimate cannot be used for direct numerical calculations either. However, this estimate of the number of terms in the BCFW relations reinforces the indications that pure gravity does not behave differently from string theory in any qualitative manner as far as the leading factorial growth of the amplitude is concerned.

However, note that in string theory, it is important to have $\gamma > 0$ to ensure that the rapid falloff of Koba-Nielsen factor at high energies does not overwhelm the factorial growth. In pure gravity, there would be no such constraint.

VI. APPLICATIONS TO THE INFORMATION PARADOX

We now describe how the analysis above can be applied to the information paradox. First, we review the two forms of the paradox that we will discuss here—the cloning paradox and the strong subadditivity paradox. We explain how an important—and often unstated—assumption that goes into formulating these paradoxes is that bulk locality holds even for very high point correlation functions, including $O(S)$ -point correlators, where the typical separation between points is the inverse of the Hawking temperature. We then describe how the loss of locality suggested by the breakdown of the perturbation theory discussed above is precisely sufficient to resolve these paradoxes.

At the end of this section, we return to the toy model of Sec. II B and describe how it can be used to set up some toy models of the information paradox. In this toy-model setting, the resolution to these paradoxes is also very clear. This toy-model analysis supports the idea that these paradoxes can be resolved by recognizing that very complicated correlation functions in gravity may display the effects of a loss of bulk locality.

A. The cloning and strong subadditivity paradoxes

We now review the cloning and strong subadditivity paradoxes that, as we will see, are closely related. We will consider a Schwarzschild black hole in d dimensions with a metric given by

$$ds^2 = -\left(1 - \frac{\mu}{r^{d-3}}\right) dt^2 + \left(1 - \frac{\mu}{r^{d-3}}\right)^{-1} dr^2 + r^2 d\Omega_{d-2}^2, \quad (6.1)$$

where $\mu = \frac{16\pi GM}{(d-2)\Omega_{d-2}}$ and $\Omega_{d-2} = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}$ is the area of a unit $(d-2)$ -sphere. So, the horizon radius is given by $r_h = \left(\frac{16\pi GM}{(d-2)\Omega_{d-2}}\right)^{\frac{1}{d-3}}$.

Strictly speaking, the metric (6.1) represents an eternal black hole, but it is correct at late times for black holes formed from collapse. Second, the metric (6.1) does not account for the effects of the backreaction of Hawking radiation on the geometry. Nevertheless, it gives an excellent approximation to the geometry provided we consider time scales that are short with respect to the evaporation time of the black hole. Whenever we refer to M, r_h below, without qualification, we are referring to the mass and radius of the black hole upon formation.

1. The cloning paradox

The cloning paradox proceeds as follows. Consider a black hole formed from collapse that gradually starts to evaporate. Now, a general argument due to Page [43] tells us that as energy is transferred from the black hole to the Hawking radiation, the von Neumann entropy of the Hawking radiation first increases and then starts to decrease.

We can make this precise by defining a length scale $\delta \gg \ell_{\text{pl}}$ but $\delta \ll r_h$, and using this length scale to define a region outside the black hole, and well separated from the horizon,

$$\text{region A: } r_h + \delta < r < \infty.$$

For convenience below, we will use $r_a = r_h + \delta$. Now to frame the result of Page in this notation we assume that on any spacelike slice drawn through the black hole spacetime, as in other local quantum field theories, the full set of operators in the theory factorizes as

$$\mathcal{A} = \bar{\mathcal{A}}_A \otimes \mathcal{A}_A, \quad (6.2)$$

which emphasizes the tensor product decomposition of the full set of observables, \mathcal{A} , into a factor corresponding to the set of operators localized in A , \mathcal{A}_A and a complementary factor, $\bar{\mathcal{A}}_A$. The assumption (6.2) is the same as assuming that the Hilbert space of the theory factorizes into a factor corresponding to the degrees of freedom on A and a complementary factor.

With this assumption, we can define a density matrix for local regions. The result of Page concerns the von Neumann entropy of the region A

$$S_A = -\text{Tr}(\rho_A \log(\rho_A)).$$

We are interested in the excess entropy over the entropy of the vacuum, so we define

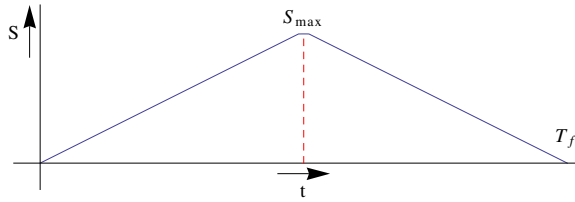


FIG. 8. Variation of the von Neumann entropy of the Hawking radiation with time. The “Page time” is shown by the dashed red line.

$$S_A = S_A^{\text{hawk}} + \frac{\text{Area}(r_a)}{4G},$$

where the second term on the right-hand side is given by $\text{Area}(r_a) = \Omega_{d-2}(r_a)^{d-2}$ and we are assuming that this is the vacuum von Neumann entropy of region A.

Page then argued that, if we consider a black hole formed from the collapse of a pure state that then proceeds to evaporate completely then S_A^{hawk} should start from zero, increase monotonically, turn around after a time called the “Page time”, and then decrease monotonically down to zero. This is shown schematically in Fig. 8.

The fact that the entropy starts to decrease after the Page time implies that the Hawking radiation outside starts to become pure. This can also be understood as the transfer of information from the interior of the black hole to the exterior.

To obtain the cloning paradox, we consider a black hole formed from collapse. Then, in the extended Penrose diagram, it is possible to draw a nice slice that cuts through the infalling matter and captures a significant fraction of the late Hawking radiation. This slice is shown in Fig. 9.

Now, by the argument above, the section of the slice outside the black hole in the region A contains a significant fraction of the information that was present in the infalling matter. However, we see that the infalling matter is itself present on another section of the slice. This seems to suggest that information has been duplicated or “cloned”, in violation of the linearity of quantum mechanics. This is the “cloning paradox” [44]. A closely related paradox was outlined by Hayden and Preskill [45]. They considered a situation where after the black hole has been formed, one throws in some *additional information*. They then argued that this additional information should return to the exterior in a “scrambling time.” This also leads to a cloning paradox since we can again draw a slice that captures the information thrown in and the exterior Hawking radiation that appears to carry the same information.

2. The strong subadditivity paradox

A closely related paradox is the strong-subadditivity paradox. To set up this paradox, we consider an old black hole, *after* the Page time, but well before the end of evaporation. At this point, the black hole has shrunk so that its horizon radius is r_h^{old} but we still have $r_h^{\text{old}} \gg \ell_{\text{pl}}$. We now define three regions A, B, C as follows:

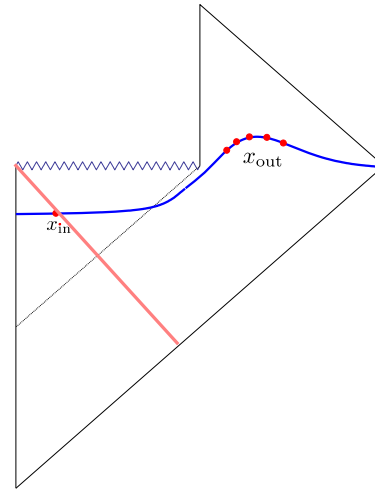


FIG. 9. A nice slice that captures the infalling matter and the late Hawking radiation. The cloning paradox is a measurement at the points x_{in} which seems to contain the same information as a measurement at the points x_{out} .

$$\text{region C: } r_h^{\text{old}} - \delta < r < r_h^{\text{old}};$$

$$\text{region B: } r_h^{\text{old}} < r < r_h^{\text{old}} + \delta;$$

$$\text{region A: } r_h^{\text{old}} + \delta < r < \infty.$$

Note that in this discussion the region A is now defined with respect to r_h^{old} and not r_h . We also use $r_a^{\text{old}} \equiv r_h^{\text{old}} + \delta$ and use $r_c^{\text{old}} \equiv r_h^{\text{old}} - \delta$.

These regions are shown in Fig. 10, where we also show a spacelike slice that cuts through all of them. Only the late-time parts of these regions—the parts above the spacelike slice—are shaded, since we are only interested in these regions at late times.

Now, to frame the strong subadditivity paradox, we assume, analogously to (6.2) that the set of operators in the theory factorizes as

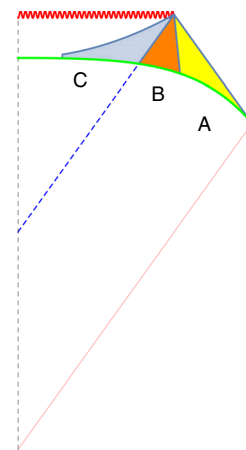


FIG. 10. A representation of the late-time parts of regions A (yellow), B (orange), and C (blue) and their intersection with a spacelike slice (green).

$$\mathcal{A} = \bar{\mathcal{A}} \otimes \mathcal{A}_C \otimes \mathcal{A}_B \otimes \mathcal{A}_A, \quad (6.3)$$

where $\bar{\mathcal{A}}$ is another residual factor and the main point of (6.3) is to emphasize that the set of local operators corresponding to the regions A , B , C appear as tensor factors in the algebra of the full theory.

Subject to the assumption (6.3), we may then consider the entropy of the regions A , B , C . This entropy can be written as a sum of the entropy of the Hawking radiation in this region and the vacuum entanglement entropy. So we have

$$\begin{aligned} S_A &= S_A^{\text{hawk}} + \frac{\text{Area}(r_a^{\text{old}})}{4G}, \\ S_B &= S_B^{\text{hawk}} + \frac{\text{Area}(r_a^{\text{old}})}{4G} + \frac{\text{Area}(r_h^{\text{old}})}{4G}; \\ S_C &= S_C^{\text{hawk}} + \frac{\text{Area}(r_c^{\text{old}})}{4G} + \frac{\text{Area}(r_h^{\text{old}})}{4G}. \end{aligned} \quad (6.4)$$

We can also consider the entropies of pairs of regions, which gives us

$$\begin{aligned} S_{BC} &= S_{BC}^{\text{hawk}} + \frac{\text{Area}(r_a^{\text{old}})}{4G} + \frac{\text{Area}(r_c^{\text{old}})}{4G}; \\ S_{AB} &= S_{AB}^{\text{hawk}} + \frac{\text{Area}(r_h^{\text{old}})}{4G}. \end{aligned} \quad (6.5)$$

Now the process of Hawking radiation involves pair creation in the region BC . The Hawking particles that exit into region B are entangled with their pairs in region C that fall into the black hole. On the other hand, if one views the Hawking radiation only in B or only in C , then this radiation appears almost thermal. This argument leads to the inequality

$$S_{BC}^{\text{hawk}} < S_C^{\text{hawk}}. \quad (6.6)$$

Now we note that the Hawking radiation in B will emerge into region A . If we are in the phase of Hawking evaporation after the Page time, then the radiation in B purifies the radiation in A . This leads to the inequality

$$S_{AB}^{\text{hawk}} < S_A^{\text{hawk}}. \quad (6.7)$$

However, a general theorem concerning von Neumann entropies—called the strong subadditivity of von Neumann entropy [46]—states that

$$S_{AB} + S_{BC} \geq S_A + S_C.$$

When we add the area terms in (6.5) and (6.4), the strong subadditivity inequality becomes

$$\begin{aligned} S_{AB}^{\text{hawk}} + \frac{\text{Area}(r_h^{\text{old}})}{4G} + S_{BC}^{\text{hawk}} + \frac{\text{Area}(r_a^{\text{old}})}{4G} + \frac{\text{Area}(r_c^{\text{old}})}{4G} \\ \geq S_A^{\text{hawk}} + \frac{\text{Area}(r_a^{\text{old}})}{4G} + S_C^{\text{hawk}} + \frac{\text{Area}(r_c^{\text{old}})}{4G} + \frac{\text{Area}(r_h^{\text{old}})}{4G}. \end{aligned}$$

The area terms neatly cancel in this inequality leaving us just with

$$S_{AB}^{\text{hawk}} + S_{BC}^{\text{hawk}} \geq S_A^{\text{hawk}} + S_C^{\text{hawk}}.$$

However, we see that the inequality above is in contradiction with (6.7) and (6.6). This is the “strong subadditivity paradox” [47].

We note that both the cloning and the strong subadditivity paradox involve assumptions of bulk locality. In the cloning paradox, we assume that operators defined on the nice slice, at the point where the nice slice intersects the infalling matter, commute with the operators that measure the same information in the Hawking radiation. In the strong subadditivity paradox, we assume that the Hilbert space factorizes as (6.3). We now analyze the protocol for determining the von Neumann entropy and extracting information from Hawking radiation more closely. We will find that this process necessarily requires the measurement of very high-point correlators, which invalidates the assumption of locality made in framing both paradoxes.

B. Protocols for extracting information

The key to resolving the paradoxes above lies in setting up a precise protocol to extract information from the emitted Hawking radiation. We can consider a family of observers who stay fixed at a very large value of r and measure the outgoing Hawking radiation as it crosses them at large values of t . Such observers may also have injected information into the system in the far past, as in the Hayden-Preskill scenario mentioned above. For simplicity, we will restrict ourselves to s-wave measurements here, since most of the energy emitted in Hawking radiation from a Schwarzschild black hole emerges in s waves. We will also be interested in a limit where the mass of the black hole becomes very large compared to the Planck mass so that $M \gg M_{\text{pl}}$ and the entropy also becomes very large, $S \gg 1$.

In this section, we would like to argue that this family of observers needs to measure S -point correlators of the emitted Hawking quanta to extract information that is relevant for the cloning and the strong subadditivity paradoxes. However, our arguments in this section will not be entirely precise, and we will be forced to rely on some discretizations and plausible assumptions. Improving this analysis of protocols for extracting information from the Hawking radiation remains an important objective for future work.

The correlation functions measured by these observers are

$$\begin{aligned}
C(u_1, \dots, u_s, v_{s+1}, \dots, v_n) \\
= \lim \int \langle T \{ \phi(t_1, \vec{x}_1) \dots \phi(t_s, \vec{x}_s), \\
\phi(t_{s+1}, \vec{x}_{s+1}), \dots, \phi(t_n, \vec{x}_n) \} \rangle \prod d\Omega_p,
\end{aligned}$$

where the limit is taken so that

$$\begin{aligned}
r_m = |\vec{x}_m| \rightarrow \infty, \quad t_1 \dots t_s \rightarrow \infty, \quad t_{s+1} \dots t_n \rightarrow -\infty \\
u_m = t_m - r_m; \quad v_m = t_m + r_m,
\end{aligned}$$

where we keep $u_1 \dots u_s$ finite and $v_{s+1} \dots v_n$ finite. The integrals are over the sphere at infinity to extract s-wave information. We can easily generalize this to consider higher spherical harmonics.

We should also emphasize that we are only interested in the *connected* part of the position space correlator.

$$C(u_i, v_j) = \frac{N^n e^{\frac{i\pi s d}{2}}}{r^{n \frac{d-2}{2}}} \int M(k_1, \dots, k_n) (2\pi)^d \delta^d \left(\sum k_p \right) e^{-i \sum E_i u_i + i \sum E_j v_j} \prod d\hat{k}_p dE_p (E_p)^{d/2-2}, \quad (6.8)$$

where the on shell d momenta $k_i = (-E_i, E_i \hat{k}_i)$ for $i = 1 \dots s$ and $k_j = (E_j, E_j \hat{k}_j)$ for $j = s+1 \dots n$ and $p = 1 \dots n$. Here, the normalization factor, N , and the phase factor $e^{\frac{i\pi s d}{2}}$ are both irrelevant. The leading suppression by a power of $\frac{1}{r}$ indicates the asymptotic falloff of the correlator and will also not be relevant. This formula is proved in Appendix B. In the case under consideration, we are only interested in correlators measured at *future null infinity*, and so we will suppress the v coordinates now.

The task of setting up a protocol to extract information is to determine which set of correlation functions we must measure at future null infinity to pin down the density matrix of the emitted radiation. We will proceed in a simple minded fashion. First, we choose the origin of coordinates so that the Hawking radiation starts to emerge at $u = 0$. Then, we allow the observer at infinity to measure correlators at $u = 0, \frac{\alpha}{T}, \frac{2\alpha}{T} \dots \frac{n_{\max} \alpha}{T}$, where T is the temperature of the Hawking radiation and α is an order 1 number that controls how finely we can make measurements. We can choose $\alpha \ll 1$, but we do not scale α with the entropy of the black hole so that $\frac{\log(\alpha)}{\log(S)} \ll 1$. We assume that the radiation continues till $\frac{n_{\max} \alpha}{T}$, and we calculate n_{\max} below. Here, we are assuming that the measurement of correlators at finer separation is not relevant for extracting information from Hawking radiation, which has a characteristic energy T . However, it would be nice to make this protocol more precise.

In terms of the horizon radius, the Hawking temperature is given by

$$T = \frac{d-3}{4\pi r_h}. \quad (6.9)$$

This is because the part of a high-point correlator that factorizes into a product of lower-point correlators does not contain any fresh information. In momentum space, it is easy to exclude the disconnected part of the scattering amplitude by simply focusing on external kinematics, where momentum is not conserved by any subset of the particles. However, in position space, the restriction to the connected part of the correlator must be imposed by hand by subtracting off the disconnected parts. We will assume that this has been done in the analysis below.

Although, from a physical perspective, these correlators are more natural, they contain *precisely* the same information as S-matrix elements. In fact, a simple application of the Lehmann-Symanzik-Zimmermann (LSZ) formula tells us that they are Fourier transforms of S-matrix elements.

The entropy of the black hole is given by

$$S = \frac{\Omega_{d-2} r_h^{d-2}}{4G}. \quad (6.10)$$

Note that the total time required for the black hole to evaporate entirely is given by

$$t_{\text{evap}} = \frac{KS}{T},$$

where $K = \frac{1}{\sigma \Omega_{d-2}} \left(\frac{4\pi}{d-3} \right)^{d-2}$, and σ is the Stefan-Boltzmann constant. Therefore, the largest value of u that is relevant in this discussion is $\frac{KS}{T}$, and we have $n_{\max} = \frac{KS}{\alpha}$.

Now the task of extracting information from the outgoing Hawking radiation boils down to the task of determining its density matrix, ρ_A . However, we note that each correlation function measured by our family of observers gives them only partial information about the density matrix ρ_A . When appropriately discretized, we expect this density matrix to have dimension $e^S \times e^S$, where S is the entropy of the black hole. If we are given a set of \mathcal{D} observables corresponding to the region A : $\mathcal{A}_A = \{A_1 \dots A_{\mathcal{D}}\}$, then these observables give us information about matrix elements of this density matrix

$$\langle A_i \rangle = \text{Tr}(\rho_A A_i).$$

To extract the full density matrix, we require $\mathcal{D} = e^{2S}$ distinct observables.

In fact, as explained in [48], if we have only a small number of observations, so that $\mathcal{D} \ll e^{2S}$, then our information about properties of the density matrix ρ_A is very limited. For example, a thermal density matrix and a pure

state density matrix may appear almost identical when the number of observations is small. Therefore, we often require accurate information about all components of the density matrix even if, in the end, we are only interested in a single number like its von Neumann entropy.

Now, here, the set of observables comprises correlation functions at null infinity measured at discrete points as outlined above. If we measure up to p -point correlators, then the number of distinct observables is given by

$$\mathcal{D} = \sum_{j=2}^p \binom{n_{\max}}{j}.$$

At large S , setting $\mathcal{D} = e^{2S}$, we see that we must consider at least $p = \eta S$ point correlators, where $\eta \log(\eta) + (1-\eta) \log(1-\eta) = -\frac{2\alpha}{K}$. The precise value of η depends on our choice of allowed spacing in our measurements, which is controlled by α . However, what is important for us is simply that $p = \mathcal{O}(S)$. In other words, this leads to the following conclusion. *To extract information from the outgoing Hawking radiation, we need to measure $\mathcal{O}(S)$ -point correlation functions of the outgoing Hawking quanta.* Furthermore, by (6.8), these correlators are sensitive to S -matrix elements that involve S particles with typical energies $E_i = \mathcal{O}(T)$.⁷

This conclusion is also relevant for the von Neumann entropies that appear in (6.7). The von Neumann entropy, S_A depends on the density matrix ρ_A . If we use an approximation to this density matrix, which corresponds to using a limited set of observables, then we do not expect (6.7) to hold. This is because we expect that the entanglement between the Hawking radiation in B and the old Hawking radiation in A can only be measured by considering very complicated operators in A . If such complicated operators are excluded from the set of allowed observables in A , then we may not find that $S_{AB} < S_A$. Therefore, the arguments in this section tell us that in order for (6.7) to be hold, we must use the density matrix ρ_A obtained by measuring a set of observables, \mathcal{A}_A , that contains S -point correlators of simple local field operators.

C. Resolving the cloning and strong subadditivity paradoxes

Once we have the conclusion above in hand, the resolution of both the cloning and the strong subadditivity

⁷Note that the exponent in the Fourier transform that appears in (6.8) appears to become large when we take energies to be of order the Hawking temperature because the u values can become large in units of $\frac{1}{T}$. However, this is just an overall large phase and does *not* imply that these energies are unimportant in the Fourier transform. We need energies of a size $\frac{1}{T}$ to distinguish between a correlator with an insertion at $\frac{i\alpha}{T}$: $C(u_1, \dots, \frac{i\alpha}{T} \dots u_s, v_j)$ and another correlator with an insertion at $\frac{(i+1)\alpha}{T}$ and all other insertion points unchanged: $C(u_1, \dots, \frac{(i+1)\alpha}{T} \dots u_s, v_j)$.

paradoxes follows naturally. We have already explained that these paradoxes rely on an assumption of bulk locality. However, the measurement protocol that we have described above relies on measuring an $\mathcal{O}(S)$ point correlation function with separations of $\frac{\alpha}{T}$. Using the conversion (6.8), this simply corresponds to an S -matrix element with $\mathcal{O}(S)$ insertions with a typical energy T .

We now see from (6.9) and (6.10) (where we recall that $8\pi G = \ell_{\text{pl}}^{d-2}$) that as we scale the mass of the black hole, $M \gg M_{\text{pl}}$, so that $S \rightarrow \infty$, the energy and number of insertions precisely satisfy

$$\frac{\log(S)}{(2-d) \log(T \ell_{\text{pl}})} \rightarrow 1.$$

Comparing this with (3.7), we see that string perturbation theory breaks down *precisely* for $\mathcal{O}(S)$ measurements, with energy $\mathcal{O}(T)$. By the arguments of Sec. II, we see that this is also the limit where we might expect a loss of locality.

In particular, this implies that the assumption that goes into the strong subadditivity paradox (6.3) is not justified for such measurements. It also tells us that, in this limit, the assumption of locality required for the cloning paradox fails: measurements of $\mathcal{O}(S)$ -point correlators of the Hawking radiation outside the black hole may not commute with measurements of the infalling matter inside the black hole, even though both of these measurements are made on the same spacelike slice.

Thus, we see that the breakdown of perturbation theory that we have described above appears precisely in time to invalidate the assumption of locality that went into the strong subadditivity and cloning paradoxes.

D. Toy models of the information paradox

It is amusing that we can use our toy model of black hole complementarity to also construct toy models of the information paradox, which are resolved in precisely the manner described above.

Consider a constant-time slice of AdS that we can divide into three regions as shown. More precisely, we define

$$\begin{aligned} \text{region C: } & r_c < r < r_0; & \text{region B: } & r_0 < r < r_a; \\ \text{region A: } & r_a < r < \infty. \end{aligned}$$

The difference with the black hole case is that now the boundaries of all three regions are timelike. So in Fig. 11, we show a constant global-time cross section of empty AdS that displays these regions.

Now in Sec. II B, we showed that local operators in C could be written entirely in terms of a very complicated set of operators in A . This immediately leads to a version of the *cloning paradox*. Consider a qubit in region C that can be manipulated by a local operator ϕ_C . Since this local

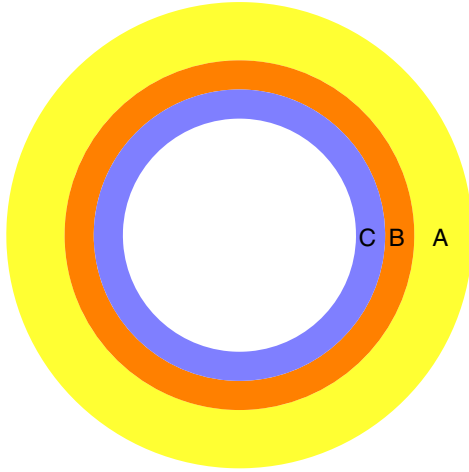


FIG. 11. A schematic description of the intersection of the regions A, B, C with a constant global time slice in empty global AdS.

operator can also be written in terms of operators that are in A, we find that this qubit can also be manipulated by some operator ϕ_A . We thus appear to have a cloning paradox, where the *same information* is present both in C and in A.

In this setup, the resolution to this paradox is trivial. We can make a distinction between operators in C and operators in A only when locality holds approximately. Locality holds provided that we restrict ourselves to “small algebras”, where we do not consider very complicated polynomials in the field operators. If we start considering such polynomials then, as shown in Sec. II B, a local operator in C can be equated to a complicated polynomial of operators in A. Therefore, there is no cloning of information. Rather what we are seeing is a loss of locality when we consider very complicated operators in AdS.

We can also produce a version of a paradox related to strong subadditivity. The strong subadditivity of entropy is closely related to the “monogamy” of entanglement. We see, from the figure, that the region B is entangled with the region C because they are close together in AdS. However, as we explained above, all the information in C is also present in A. Therefore, the region B is also entangled with the region A. Thus, we seem to have a loss of the monogamy of entanglement—with B being entangled both with C and with A.

Once again, the resolution to this paradox is very simple in this case. The regions A, B, C are approximately independent local regions only when we consider low-order polynomials of simple operators localized in these regions. At a fine-grained level, the bulk Hilbert space does not contain tensor factors $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. In fact, we know from AdS/CFT that the entire Hilbert space is contained in \mathcal{H}_A . Therefore, this seeming paradox is also resolved by recognizing that bulk locality is not exact.

VII. DISCUSSION

In the analysis above, we studied S-matrix elements in string theory and argued that they could not be computed perturbatively when the number of external particles became too large, even if each of these particles carried only a small amount of energy in string units.

This is an interesting limit because of the relation between the breakdown of the perturbation theory for S-matrix elements and the loss of bulk locality. As we explained in Sec. II, these two thresholds are related because, in any theory of gravity, the Gauss law leads to small perturbative nonlocal commutators. When perturbation theory breaks down, we must consider the possibility that these commutators become significant at leading order.

As we reviewed in Sec. II B, these arguments which relate the loss of locality to the combination of the Gauss law and the breakdown of the perturbation theory—are known to be correct at least in one entirely controlled setting: empty AdS. In that setting, [5] showed how an operator anywhere in AdS could be rewritten as a sufficiently complicated combination of operators that were all spacelike separated from the original operator.

We showed how these potential nonlocal effects have significant consequences for some versions of the information paradox. In particular, we considered the cloning paradox that appears because the information present in the infalling matter also appears to be present in the outgoing Hawking radiation. We argued that the usual framing of this paradox ignores the fact that to parse this information outside the black hole, we need to measure $O(S)$ -point correlators in the outgoing Hawking radiation, where S is the entropy of the black hole. But the breakdown of gravitational perturbation theory, which we linked to a loss of locality, occurs precisely for $O(S)$ -point correlators. We used this to argue that the complicated observable that is required to perform this measurement is not guaranteed to commute with the operator in the interior that acts on the infalling matter, even though these operators are spacelike separated.

A similar loophole exists in the strong subadditivity paradox. The strong subadditivity paradox is framed in terms of the von Neumann entropies of different parts of the black hole spacetime. In a theory of gravity, we need to be careful about how to define the von Neumann entropy of a spacetime region. This entropy depends on how we truncate the “algebra” of operators localized in this region. We argued that, to obtain a paradox, it was necessary to define the von Neumann entropy in a fine-grained manner, where we expand the algebra of operators to include polynomials that are very high order in the simple local operators from the region. But such high order polynomials may not commute on a spacelike slice. Therefore, with this choice of local algebra, it is not valid to assume that the von Neumann entropies of geometrically disjoint regions will

satisfy the same properties as von Neumann entropies of direct-product Hilbert spaces.

As additional evidence for the resolutions above, we also considered toy models of the strong subadditivity and cloning paradoxes that appear in empty AdS. In this setting, the resolution to these paradoxes is clear and indeed relies on the loss of locality that we have described above.

Nevertheless, while we believe that these arguments are persuasive, we should indicate several points at which they can be made firmer. First, while Gauss-law commutators are clearly also present in flat space, and while our arguments for the breakdown of string perturbation theory are robust, we do not have the same controlled understanding of nonlocal effects in flat space or about black hole backgrounds. It would be nice to make this more precise.

Second, when we applied our analysis of the breakdown of perturbation theory to black hole evaporation, we were forced to use several *ad hoc* discretizations to analyze our measurement protocols. Also, it is possible to set up other protocols for extracting the information in Hawking radiation. For example, one protocol that seemingly avoids the need for measuring high-point correlators is simply to measure a lower-point function to arbitrarily high accuracy. However, then we see that to extract information, we need to measure this lower-point correlation function to exponential accuracy in the entropy of the black hole [48]. This observation is sensitive to *very high* order terms in perturbation theory in $\frac{E}{M_{\text{pl}}}$, where E is the energy of a typical Hawking quantum. Just as perturbation theory breaks down for a very large number of particles, it is well-known that it also breaks down at very high loop order. Therefore, even in a measurement protocol that involves the measurement of low-point correlators to very high accuracy, we would expect nonperturbative and—by the arguments above—nonlocal effects to be important. It would be nice to frame this entire analysis of measurement protocols within a unified framework that makes it clear that extracting information from Hawking radiation necessarily involves some nonperturbative physics.

We would also like to caution the reader on a few points related to our discussion of the information paradox. Our discussion of the information paradox here does not address the question of reconstructing the interior of large black holes in AdS/CFT. This has been the subject of considerable recent discussion [49] and involves several interesting and partially unresolved puzzles [50]. Here, we are only concerned with the information paradox for evaporating black holes. Second, we note that some alternative resolutions to these paradoxes have been proposed in [51], which do not rely on the nonlocal effects described here. In forthcoming work, we hope to provide a comparative analysis of these resolutions.

Finally, we should also mention another important interpretation of the breakdown of perturbation theory that we have described. In this paper, and especially in Sec. VI,

we considered this breakdown in a setting where a black hole already exists in spacetime, and where we are making complicated measurements in the outgoing Hawking radiation.

However, it is also possible to consider high-point S-matrix elements in the absence of a black hole. Simple dimensional analysis then tells us that at the threshold of perturbative breakdown, we expect the *formation* of black holes.

This can be seen as follows. If the momentum components of each particle are of order E then each particle is delocalized over a length of at least $O(1/E)$ and so we expect that the smallest black hole that can be made by many such particles also has a radius $O(1/E)$. The mass of such a black hole in d dimensions is $O(M_{\text{pl}}^{d-2}/E^{d-3})$, which requires $O((M_{\text{pl}}/E)^{d-2})$ particles of energy E . But this is precisely the value of n at which our breakdown occurs. So we see that the breakdown of perturbation theory that we have described is also related to black hole formation. This analysis of the threshold of black hole formation and its relation to the loss of locality was also discussed in [52].

Separate from the issue of the breakdown of perturbation theory, the analysis of string scattering at a large n is of intrinsic interest. It would be very interesting, for example, to examine the limit where n becomes large but we are still below the threshold of perturbative breakdown and ask whether string perturbation theory simplifies in this limit.

Finally, it would be nice to have a better understanding of the nonperturbative effects that become important beyond the perturbative threshold. The recognition that string perturbation theory breaks down at large genus spurred the discovery of d -branes, and we would like to understand if the large- n breakdown is related to similar interesting nonperturbative phenomena.

ACKNOWLEDGMENTS

We are grateful to Nima Arkani-Hamed, Clay Cordova, Steve Giddings, Rajesh Gopakumar, David Gross, Subhojy Gupta, Gautam Mandal, Samir Mathur, Shiraz Minwalla, Yasunori Nomura, Ashoke Sen, Aninda Sinha, Sandip Trivedi, Edward Witten, and especially, R. Loganayagam and Kyriakos Papadodimas for useful discussions. We would also like to acknowledge discussions with the participants of the ICTS discussion meeting, String Theory: Past and Present, ICTS/STTP/2017/01, the ICTS program, Fundamental Problems in Quantum Physics, ICTS/Prog-FPQP/2016/11, and the Bangalore Area Strings Meeting, ICTS/Prog-BASM/2016/07. We would like to acknowledge the hospitality of the Strings 2017 conference (Tel Aviv), the National Strings Meeting (Pune), the program on Black Holes and Emergent Spacetime at Nordita (Stockholm), and the Mandelstam theoretical physics school and workshop (Durban). Our numerical computations were performed on the high-performance computing cluster at ICTS: *Mowgli*.

We would also like to thank the Astrophysical Relativity group at ICTS for giving us access to their computational facilities.

APPENDIX A: ADDITIONAL DETAILS ABOUT THE NUMERICAL ALGORITHM

1. Root finding

Our root-finding algorithm is a generalization of the Newton algorithm. We use the GNU Scientific Library’s “hybridj” function [53]. We are interested in solving the equations $E_i = 0$ specified in (5.3). The basic multi-dimensional Newton’s algorithm starts with an initial guess z_i . Each step updates this to $z'_i = z_i - J_{ij}^{-1} z_j$, where $J_{ij} = \partial_i E_j$. This algorithm was improved by Powell [54] who pointed out that the steps should be restricted to $|z'_i - z_i| \leq \delta$, where δ is a dynamically determined trust region. If the step is outside the trust region, then Powell’s hybrid algorithm uses a combination of the Newton step and gradient flow by taking a step along

$$z'_i - z_i = -\alpha J_{ij}^{-1} E_j - \beta \nabla_i \sum_j |E_j|^2.$$

Here, α, β are fixed by demanding that the step minimize the norm of the function $\sum_j |E_j|^2$, with the constraint that the step also stay inside the trust region. After the step is completed, the algorithm then determines if $|E_i(z'_i)| \leq |E_i(z_i)|$. If this is the case, the step is accepted and δ is increased. Otherwise, the step is rejected, and δ is decreased.

In our case, we choose a starting point by choosing a random value for each z_i satisfying $-50 \leq \text{Re}(z_i) \leq 50$ and $-50 \leq \text{Im}(z_i) \leq 50$. We then use the algorithm detailed above to attempt to find a root.

Even the simple one-dimensional Newton method is not guaranteed to converge. The convergence properties of multidimensional problems are, in general, even worse. However, we were pleasantly surprised to find that the hybrid algorithm converges rather well to a root of the scattering equations, starting with a random initial guess as detailed above. This indicates that most points in our starting region are within the basin of attraction of some root.

Some versions of the hybrid algorithm scale the trust region with a factor that is dynamically determined using the Jacobian but we found, empirically, that the unscaled hybrid method works better than the scaled hybrid method. The vanilla Newton’s method does not converge at all.

Note that the algorithm only solves $(n - 3)$ equations. This may sometimes lead to “false roots”, where some z_i drifts off to ∞ thereby causing the corresponding $S_i \rightarrow 0$ within numerical errors. These false roots can be detected by checking if the roots that have been found also satisfy the other three equations.

A judicious gauge choice is required to ensure that the number of false solutions is kept within limit. We find that the algorithm converges well if we choose $z_1 = 10^6$, $z_2 = 0.0, z_3 = 1.0$. We also find the best performance by then solving the $(n - 3)$ equations $(E_2 \dots E_{n-1}) = 0$.

2. Random number selection

To ensure that we sample the space of roots as uniformly as possible, the choice of random number generator is also important. In fact, hidden correlations in pseudorandom number generators used for Monte Carlo sampling can sometimes lead to significant errors in the final result [55]. We use the random number generator “gsl_rng_ranxs2” provided with the GNU Scientific Library. This uses a variant of the RANLUX algorithm of Luscher [56].

3. Momentum selection

It is also important to correctly sample scattering amplitudes in momentum space. We are interested in generating external kinematics that are uniformly distributed in phase space according to the measure (3.3). For this, we can use the RAMBO algorithm [39], which is used to generate events in particle physics. This algorithm works well for massless particles, where it can generate external kinematics precisely according to the distribution (3.3). The algorithm proceeds as follows.

- (1) First, we pick a set of $n/2$ momenta corresponding to ingoing particles using the following procedure for each momentum. We choose the energy q_0 from the distribution

$$p(q_0) = \frac{1}{\Gamma(d-2)} (q_0)^{d-3} e^{-q_0}.$$

The other components of the momentum are picked by picking a vector on a sphere $\vec{n} \in \Omega_{d-1}$ and then choosing the momentum to be

$$(q_0, q_0 \vec{n}).$$

We will call this set of momenta q_i .

- (2) Let the center of mass-momentum of the momenta so obtained by \vec{P} . Now, we simultaneously boost all the momenta formed so that the new center of mass momentum becomes 0.
- (3) We now rescale the momenta so that the total energy becomes $\frac{nE}{2}$. The transformations corresponding to this boost and rescaling are as follows:

$$\begin{aligned} k_i^0 &= x(gq_i^0 + \vec{b} \cdot \vec{q}_i), \\ \vec{k}_i &= x(\vec{q}_i + \vec{b}q_i^0 + a(\vec{b} \cdot \vec{q}_i)\vec{b}), \end{aligned}$$

where

$$\vec{b} = -\frac{\vec{Q}}{M}, \quad x = \frac{nE}{2M}, \quad g = \frac{Q^0}{M},$$

$$a = \frac{1}{1+g}, \quad Q^\mu = \sum_{i=1}^{\frac{n}{2}} q_i^\mu, \quad M = \sqrt{Q^2}.$$

- (4) The momenta $k_1 \dots k_{\frac{n}{2}}$ obtained above can be taken to be the physical incoming momenta. We now pick another set of $n/2$ momenta corresponding to outgoing particles and use the same procedure to ensure that their center of mass momentum is 0 and total center of mass energy is $\frac{nE}{2}$.

It was shown in [39] that this procedure leads to momenta that are uniformly distributed in phase space.

4. Polarization tensor selection

As explained in the text, we choose polarization tensors $\epsilon_{\mu\nu} = \epsilon_\mu \epsilon_\nu$, which are simply outer products of real polarization vectors. For a momentum given by $(k_0, k_0 \vec{n})$, we first perform an orthogonal transformation in the spatial directions so that the components of this momentum become $\vec{n}^1 = 1$ and $\vec{n}^i = 0$ for $i = 2 \dots d-1$. We then choose a random unit vector with components $\epsilon^1 = 0$ and satisfying $\sum_{i=2}^{d-1} \epsilon^i \epsilon^i = 1$. We then perform the inverse orthogonal transformation so that the momentum is rotated back to its original value. By acting with the same transformation on the polarization vector, we generate a transverse polarization vector that has no timelike component. We then simply take the outer product of this polarization vector with itself to obtain a polarization tensor for the spin-2 particles.

5. Prefactor estimation

The primary difficulty in estimating the prefactor P_n originates in evaluating the term that involves the Wick contractions. We find that the most stable solution is obtained by simply selecting a random sample of pairings and by multiplying the mean of this sample with the total number of pairings. Note that trying to identify the Wick contraction that gives the largest contribution is equivalent to pairing the n points so that the product of the distances between pairs is the minimum. This is similar to a traveling salesman problem and therefore, cannot be solved for large n in any computationally efficient manner.

In our computations, we sum over a sample of 1000 samples of different possible pairings; each time, we need to evaluate the term P_n .

The log normal distribution of the samples of amplitudes that we discussed in Sec. VC is also visible in these samples. Therefore, it is possible that a better algorithm that efficiently locates nearest neighbors might increase the significance of this term in the numerical results.

6. Parallelization

Our calculations can be entirely parallelized, since every solution of the scattering equations is independent of every other solution. In principle, it is possible that two random choices of initial points lead to the same solution, and therefore while generating random parallel samples of solutions, it is important to check that we are not generating any duplicate solutions. In practice, we have never found any duplicate solutions for $n > 10$. Therefore, we were able to perform our calculations efficiently by using a number of CPUs running in parallel in a large computational cluster. In advance of our final computations, we also generated some preliminary data sets on a desktop workstation for which we found that GNU parallel [57] was a useful tool.

APPENDIX B: ASYMPTOTIC CORRELATORS AND THE S MATRIX

In this appendix, we review how correlators of fields at null infinity can be rewritten as Fourier transforms of S-matrix elements using the LSZ formula.

Consider a scalar field, $\phi(t, \vec{x})$. As above, in this appendix, we will use \vec{x} —to indicate spatial vectors. Now, as in Sec. VI, we wish to consider the following correlator:

$$C(u_1, \dots, u_s, v_{s+1} \dots v_n)$$

$$= \lim \int \langle T \{ \phi(t_1, \vec{x}_1) \dots \phi(t_s, \vec{x}_s) \phi(t_{s+1}, \vec{x}_{s+1}), \dots$$

$$\times \phi(t_n, \vec{x}_n) \} \rangle \prod d\Omega_p,$$

with

$$r_m = |\vec{x}_m| \rightarrow \infty, \quad t_1 \dots t_s \rightarrow \infty, \quad t_{s+1} \dots t_n \rightarrow -\infty,$$

$$u_m = t_m - r_m; \quad v_m = t_m + r_m,$$

where we keep u_i finite for $i = 1 \dots s$ and v_i finite for $i = s + 1 \dots n$ and $p = 1 \dots n$.

This is a s-wave correlator with some points taken the past null infinity and some others taken to future null infinity. The spherical integrals are taken over the $(d-2)$ -sphere at infinity. We will only consider massless scalar particles here. As we have mentioned, it is not difficult to generalize this to higher spherical harmonics. Now, the time-ordered correlator can be written as the Fourier transform of the momentum space time-ordered Green's function so the correlator of interest becomes

$$C(u_i, v_j) = \lim \int G(\omega_1, \vec{k}_1 \dots \omega_n, \vec{k}_n) e^{i \sum_q \omega_q t + \vec{k}_q \cdot \vec{x}_q}$$

$$\times \prod d\omega_q d^{d-1} \vec{k}_q d\Omega_p.$$

Here, to lighten the notation, we have suppressed the distinct range of indices on u, v although it is understood

that $i = 1 \dots s$ and $j = s + 1 \dots n$. We have also split up the momentum of each insertion into an energy and a spatial component as (ω_i, \vec{k}_i) . Note that these momenta do *not* have to be on shell.

We can do the position space spherical integrals as follows. For each coordinate \vec{x}_m , we can choose coordinates so that $\vec{k}_m \cdot \vec{x}_m = k_m r_m \cos \theta_m$. Here, through a slight abuse of notation, we have used $k_m \equiv |\vec{k}_m|$ although at other times we also use k_m to denote the full on shell d momenta. We hope that this will not lead to confusion. We also have $d\Omega_m = \frac{2\pi^{\frac{d-2}{2}}}{\Gamma(\frac{d-2}{2})} (\sin \theta_m)^{(d-3)} d\theta_m$. We can then do the remaining angular integrals through

$$\int_0^\pi e^{ik_m r_m \cos \theta_m} (\sin \theta_m)^{d-3} d\theta_m = \sqrt{\pi} \Gamma\left(\frac{d-2}{2}\right) {}_0\tilde{F}_1\left(\frac{d-1}{2}; -\frac{k_m^2 r_m^2}{4}\right),$$

where \tilde{F} is the regularized hypergeometric function. Moreover, we are only interested in the limit that $r_m \rightarrow \infty$. In this limit, expanding the hypergeometric function and keeping only the leading terms, we find that

$$\int_0^\pi e^{ikr \cos(\theta)} \sin(\theta)^{d-3} d\theta \rightarrow i 2^{\frac{d-2}{2}} (kr)^{-d/2+1} (e^{ikr-i\pi\frac{d}{4}} - e^{-ikr+i\pi\frac{d}{4}}) \Gamma\left(\frac{d-2}{2}\right).$$

For convenience, we define the constant

$$N = -\frac{1}{2} (2\pi)^{\frac{d}{2}} e^{-i\pi\frac{d}{4}},$$

which comes from combining the various constants that appear above, and we have added an extra factor of πi using some foresight.

We find that our correlator reduces to

$$C(u_i, v_j) = \lim_{r^n \rightarrow \infty} \frac{1}{r^{n(\frac{d-1}{2})}} \int G(\omega_1, \vec{k}_1 \dots \omega_n, \vec{k}_n) \prod_p \left[\frac{d^{d-1} \vec{k}_p d\omega_p}{k_p^{\frac{d}{2}-1}} \times \left(\frac{N}{\pi i} e^{i\frac{\omega_p+k_p}{2}(t_p+r_p)+i\frac{\omega_p-k_p}{2}(t_p-r_p)} - \frac{N^*}{\pi i} e^{i\frac{\omega_p+k_p}{2}(t_p-r_p)+i\frac{\omega_p-k_p}{2}(t_p+r_p)} \right) \right].$$

We change variables to $k_p^+ = \frac{\omega_p+k_p}{2}$; $k_p^- = \frac{\omega_p-k_p}{2}$. We also rewrite the momentum space measure as

$$d\omega d^{d-1} \vec{k} = 2dk^+ dk^- (k^+ - k^-)^{d-2} d\hat{k}.$$

We note that when we do this, we need to be careful about the limits of integration over k^+ and k^- since k must be positive and so $k^+ \geq k^-$. This leads to

$$C(u_i, v_j) = \lim_{r^n \rightarrow \infty} \frac{1}{r^{n(\frac{d-1}{2})}} \int G(\omega_1, \vec{k}_1 \dots \omega_n, \vec{k}_n) \prod_p \left(\frac{2N}{\pi i} e^{ik_p^+ v_p + ik_p^- u_p} - \frac{2N^*}{\pi i} e^{ik_p^+ u_p + ik_p^- v_p} \right) \times (k_p^+ - k_p^-)^{\frac{d-2}{2}} dk_p^+ dk_p^- d\hat{k}_p.$$

Now, we need the following formula:

$$\lim_{k \rightarrow \infty} \int_a^\infty \frac{f(x)}{x - i\epsilon} e^{ikx} dx = (2\pi i) f(0) \theta(-a),$$

where the order of limits is that, after doing the integral, ϵ is taken to zero *before* k is taken to infinity. To prove this, simply consider the contour that goes up in an arc from $x = \infty$ and comes down vertically on a through the complex x plane. This contour encloses the pole at $x = i\epsilon$, if $a < 0$, and not otherwise. Other poles in the function $f(x)$ for $\text{Im}(x) > 0$ do not contribute in the limit where $k \rightarrow \infty$. Also, the part of the contour that is not on the real line does not contribute in the limit where $k \rightarrow \infty$. Repeating this, for other cases, which will be useful below, we get the following additional identities:

$$\lim_{k \rightarrow \infty} \int_a^\infty \frac{f(x)}{x + i\epsilon} e^{ikx} dx = 0,$$

$$\lim_{k \rightarrow -\infty} \int_a^\infty \frac{f(x)}{x - i\epsilon} e^{ikx} dx = 0;$$

$$\lim_{k \rightarrow -\infty} \int_a^\infty \frac{f(x)}{x + i\epsilon} e^{ikx} dx = -2\pi i f(0) \theta(-a).$$

Now, the key point is that the Green's function has a pole when the momenta go on shell, with a residue that is the S-matrix element. In a scheme, where the wavefunction renormalization factors are unity (this is simply a question of the normalization of the two point function), we have that

$$G(\omega_1, \vec{k}_1 \dots \omega_n, \vec{k}_n) = \frac{\hat{M}(k_1 \dots k_n)}{\prod (4k_p^+ k_p^- + i\epsilon)} + F(\omega_p, \vec{k}_p),$$

where F is some function that does not have a pole in the limit where $k_p^+ k_p^- = 0$ and the momenta that enter the scattering amplitude are now on shell. We use \hat{M} to denote the amplitude with the momentum-conserving delta function included

$$\hat{M}(k_1, \dots, k_n) = M(k_1, \dots, k_n) (2\pi)^d \delta\left(\sum_{p=1}^n k_p\right).$$

Now, as we take the points to either future or past null infinity, for each point either $u_p \rightarrow -\infty$ or $v_p \rightarrow \infty$. In each of these limits, the momentum space integrals simplify as follows:

$$\lim_{u_p \rightarrow -\infty} \int_{-\infty}^{\infty} dk_p^- \int_{k_p^-}^{\infty} dk_p^+ \frac{\hat{M}(k_1 \dots k_n) (k_p^+ - k_p^-)^{\frac{d-2}{2}}}{(4k_p^+ k_p^- + i\epsilon)} \times e^{ik_p^+ u_p + ik_p^- v_p} = 0.$$

This is because, to pick up a contribution, we need the pole in k_p^+ to be in the lower half plane, and this requires $k_p^- > 0$.

However, then by the identity above, the integral vanishes. Similarly, we find

$$\begin{aligned} & \lim_{u_p \rightarrow -\infty} \int_{-\infty}^{\infty} dk_p^+ \int_{-\infty}^{k_p^+} dk_p^- \frac{\hat{M}(k_1, \dots, k_n) (k_p^+ - k_p^-)^{\frac{d-2}{2}}}{(4k_p^+ k_p^- + i\epsilon)} e^{ik_p^- u_p + ik_p^+ v_p} \\ &= \int_0^{\infty} dk_p^+ \frac{\pi i}{2} (k_p^+)^{\frac{d}{2}-2} \hat{M}(k_1, \dots, k_n) e^{ik_p^+ v_p}, \end{aligned}$$

and

$$\begin{aligned} & \lim_{v_p \rightarrow \infty} \int_{-\infty}^{\infty} dk_p^- \int_{k_p^-}^{\infty} dk_p^+ \frac{\hat{M}(k_1, \dots, k_n) (k_p^+ - k_p^-)^{\frac{d-2}{2}}}{(4k_p^+ k_p^- + i\epsilon)} e^{ik_p^- u_p + ik_p^+ v_p} \\ &= -e^{\frac{i\pi d}{2}} \int_{-\infty}^0 dk_p^- \frac{\pi i}{2} (k_p^-)^{\frac{d}{2}-2} \hat{M}(k_1 \dots k_n) e^{ik_p^- u_p}, \end{aligned}$$

while

$$\lim_{v_p \rightarrow \infty} \int_{-\infty}^{\infty} dk_p^+ \int_{-\infty}^{k_p^+} dk_p^- \frac{\hat{M}(k_1, \dots, k_n)}{(4k_p^+ k_p^- + i\epsilon)} e^{ik_p^- v_p + ik_p^+ u_p} = 0.$$

Putting all of this together, we find the final formula

$$C(u_i, v_i) = \frac{N^n e^{\frac{i\pi s d}{2}}}{r^{n \frac{d-2}{2}}} \int \prod_{i=1}^n d\hat{k}_p dE_p (E_p)^{d/2-2} \hat{M}(k_1, \dots, k_n) e^{-i \sum_{p=1}^s E_p u_p + i \sum_{p=s+1}^n E_p v_p},$$

where now the integral over the energies E_i goes over $(0, \infty)$ and

$$k_i \equiv (-E_i, E_i \hat{k}_i) \quad \text{for } i = 1 \dots s$$

and

$$k_i \equiv (E_i, E_i \hat{k}_i) \quad \text{for } i = s+1 \dots n.$$

This is the formula that is used in Sec. VI.

APPENDIX C: VOLUME OF PHASE SPACE

Now we compute the volume of phase space using the formulas in Appendix A of [39]. The only difference with their analysis is that we work in d dimensions whereas they were exclusively concerned with $d = 4$. We consider $m = \frac{n}{2}$ particles with total d momentum $P = (Q, 0, \dots)$. Recall that in the text we had $Q = \frac{nE}{2}$ and $m = \frac{n}{2}$ but we will work with general parameters here.

Then, on grounds of symmetry, it must be true that the volume of phase space is $V_m = v_m Q^{(d-2)m-d}$. We determine v_m by recursion. We have

$$V_m = \int \prod_{j=1}^m \frac{d^d p_j}{(2\pi)^d} (2\pi) \delta(p_j^2) (2\pi)^d \delta^d\left(\sum p_j - P\right).$$

We insert 1 in the integral through

$$1 = \int (2\pi)^d \delta^d\left(\sum_{j=1}^{m-1} p_j - q\right) \frac{d^d q}{(2\pi)^d} (2\pi) \delta(q^2 - w^2) \frac{dw^2}{(2\pi)}.$$

Using Lorentz invariance, and in the presence of the delta functions above, we can write the integral over $m-1$ particles as

$$\begin{aligned} & \int \prod_{j=1}^{m-1} \frac{d^d p_j}{(2\pi)^d} (2\pi) \delta(p_j^2) (2\pi)^d \delta^d\left(\sum p_j - q\right) \\ &= v_{m-1} w^{(d-2)(m-1)-d}. \end{aligned}$$

Inserting this, we find that

$$V_m = v_{m-1} w^{(d-2)(m-1)-d} \int \frac{d^d p_m}{(2\pi)^d} (2\pi) \delta(p_m^2) (2\pi)^d \delta^d(q + p_m - P) (2\pi) \delta(q^2 - w^2) \frac{dw^2}{(2\pi)} \frac{d^d q}{(2\pi)^d}.$$

The last integral over p_m and q can be done as follows. First, we simply do the q integral using the delta functions. The measure over p_m can be written as

$$\frac{d^{d-1} p_m}{(2\pi)^{d-1} 2|p_m|} = \frac{\Omega_{d-2}}{2(2\pi)^{d-1}} |p_m|^{d-3} d|p_m|.$$

The remaining integral becomes

$$\frac{\Omega_{d-2}}{2(2\pi)^{d-1}} \int d|p_m| |p_m|^{d-3} v_{m-1} w^{(d-2)(m-1)-d} \delta\left(|p_m| + \sqrt{|p_m|^2 + w^2} - Q\right) \frac{1}{2\sqrt{|p_m|^2 + w^2}}.$$

The delta function is solved by

$$|p_m| = \frac{Q^2 - w^2}{2Q},$$

at which point, we also have (after including the Jacobian from the delta function) and doing the integral over $|p_m|$

$$\begin{aligned} V_m &= v_{m-1} \frac{\Omega_{d-2}}{4(2\pi)^{d-1}} \int_0^Q (2w) dw 2^{(d-3)} Q^{2-d} 2^{3-d} (Q^2 - w^2)^{(d-3)} w^{(d-2)(m-1)-d} \\ &= v_{m-1} \frac{\Omega_{d-2}}{4(2\pi)^{d-1}} Q^{(d-2)m-d} \frac{2^{3-d} \Gamma(d-2) \Gamma(\frac{1}{2}(d-2)(m-2))}{\Gamma(\frac{1}{2}(d-2)m)}. \end{aligned}$$

This leads to the recursion relation

$$v_m = v_{m-1} \frac{\Omega_{d-2}}{(2\pi)^{d-1}} \frac{2^{1-d} \Gamma(d-2) \Gamma(\frac{1}{2}(d-2)(m-2))}{\Gamma(\frac{1}{2}(d-2)m)}.$$

We can also compute V_2 using

$$V_2 = \int \frac{d^d p_1}{(2\pi)^d} (2\pi) \delta(p_1^2) (2\pi)^d \delta^d(p_1 + p_2 - P) (2\pi) \delta(p_2^2) \frac{d^d p_2}{(2\pi)^d},$$

which just gives

$$V_2 = \frac{\Omega_{d-2}}{(2\pi)^{d-1}} \int |p_1|^{(d-4)} \frac{d|p_1|}{4} (2\pi) \delta(2|p_1| - w) = \frac{\Omega_{d-2}}{8(2\pi)^{d-2}} \left(\frac{w}{2}\right)^{d-4},$$

or

$$v_2 = \frac{\Omega_{d-2}}{(2\pi)^{d-2} 2^{d-1}}.$$

The recursion relation we have is of the form

$$v_m = C \frac{\Gamma(k(m-2))}{\Gamma(km)} v_{m-1},$$

and this is solved by

$$v_m = C^{m-2} \frac{\Gamma(k) \Gamma(2k)}{\Gamma(km) \Gamma(k(m-1))} v_2.$$

Specializing to the constants, we have we find that

$$\begin{aligned}
v_m &= \left(\frac{2^{1-d} \Gamma(d-2) \Omega_{d-2}}{(2\pi)^{d-1}} \right)^{m-2} \frac{\Gamma(\frac{1}{2}(d-2)) \Gamma(d-2)}{\Gamma(\frac{1}{2}(d-2)m) \Gamma(\frac{1}{2}(d-2)(m-1))} v_2 \\
&= \left(\frac{2^{1-d} \Gamma(d-2) \Omega_{d-2}}{(2\pi)^{d-1}} \right)^{m-1} \frac{\Gamma(\frac{1}{2}(d-2))}{2\pi \Gamma(\frac{1}{2}(d-2)m) \Gamma(\frac{1}{2}(d-2)(m-1))}.
\end{aligned}$$

If we now substitute $\Omega_{d-2} = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}$ and use the identity $\Gamma(x)\Gamma(x+\frac{1}{2}) = 2^{(1-2x)}\sqrt{\pi}\Gamma(2x)$ and do some simplifications, we find that

$$v_m = (2\pi)(4\pi)^{\frac{-(m-1)d}{2}} \frac{\Gamma(\frac{d-2}{2})^m}{\Gamma(\frac{1}{2}(d-2)m) \Gamma(\frac{1}{2}(d-2)(m-1))}.$$

The full phase space volume can then be written as

$$V_m = \left(\frac{Q}{m} \right)^{(d-2)m-d} (2\pi)(4\pi)^{\frac{-(m-1)d}{2}} \frac{\Gamma(\frac{d-2}{2})^m m^{(d-2)m-d}}{\Gamma(\frac{1}{2}(d-2)m) \Gamma(\frac{1}{2}(d-2)(m-1))}.$$

Putting $m = \frac{n}{2}$ and $Q = \frac{nE}{2}$, we find the formula (3.5) used in the text.

-
- [1] K. Papadodimas and S. Raju, State-dependent bulk-boundary maps and black hole complementarity, *Phys. Rev. D* **89**, 086010 (2014); Black Hole Interior in the Holographic Correspondence and the Information Paradox, *Phys. Rev. Lett.* **112**, 051301 (2014).
- [2] K. Papadodimas and S. Raju, Local Operators in the Eternal Black Hole, *Phys. Rev. Lett.* **115**, 211601 (2015).
- [3] K. Papadodimas and S. Raju, Comments on the necessity and implications of state-dependence in the black hole interior, *Phys. Rev. D* **93**, 084049 (2016).
- [4] J. M. Maldacena, The large N limit of superconformal field theories and supergravity, *Int. J. Theor. Phys.* **38**, 1113 (1999); S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, Gauge theory correlators from noncritical string theory, *Phys. Lett. B* **428**, 105 (1998); E. Witten, Anti-de Sitter space and holography, *Adv. Theor. Math. Phys.* **2**, 253 (1998).
- [5] S. Banerjee, J.-W. Bryan, K. Papadodimas, and S. Raju, A toy model of black hole complementarity, *J. High Energy Phys.* **05** (2016) 004.
- [6] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Holographic representation of local bulk operators, *Phys. Rev. D* **74**, 066009 (2006); Local bulk operators in AdS/CFT: A boundary view of horizons and locality, *Phys. Rev. D* **73**, 086003 (2006); Local bulk operators in AdS/CFT: A holographic description of the black hole interior, *Phys. Rev. D* **75**, 106001 (2007).
- [7] G. 't Hooft, On the quantum structure of a black hole, *Nucl. Phys. B* **256**, 727 (1985); L. Susskind, L. Thorlacius, and J. Uglum, The stretched horizon and black hole complementarity, *Phys. Rev. D* **48**, 3743 (1993).
- [8] S. Ghosh and S. Raju, Breakdown of String Perturbation Theory for Many External Particles, *Phys. Rev. Lett.* **118**, 131602 (2017).
- [9] G. Dvali, C. Gomez, R. S. Isermann, D. Lüist, and S. Stieberger, Black hole formation and classicalization in ultra-Planckian 2N scattering, *Nucl. Phys. B* **893**, 187 (2015); A. Addazi, M. Bianchi, and G. Veneziano, Glimpses of black hole formation/evaporation in highly inelastic, ultra-planckian string collisions, *J. High Energy Phys.* **02** (2017) 111.
- [10] N. Arkani-Hamed, S. Dubovsky, A. Nicolis, E. Trincherini, and G. Villadoro, A measure of de Sitter entropy and eternal inflation, *J. High Energy Phys.* **05** (2007) 055.
- [11] R. Haag, *Local Quantum Physics: Fields, Particles, Algebras*, 2nd ed. (Springer, New York, 1992).
- [12] H. Osborn and A. Petkou, Implications of conformal invariance in field theories for general dimensions, *Ann. Phys. (N.Y.)* **231**, 311 (1994).
- [13] R. L. Arnowitt, S. Deser, and C. W. Misner, The dynamics of general relativity, *Gen. Relativ. Gravit.* **40**, 1997 (2008); T. Regge and C. Teitelboim, Role of surface integrals in the Hamiltonian formulation of general relativity, *Ann. Phys. (N.Y.)* **88**, 286 (1974).
- [14] W. Donnelly and S. B. Giddings, Diffeomorphism-invariant observables and their nonlocal algebra, *Phys. Rev. D* **93**, 024030 (2016).
- [15] S. Gao and R. M. Wald, Theorems on gravitational time delay and related issues, *Classical Quantum Gravity* **17**, 4999 (2000).
- [16] M. V. Libanov, V. A. Rubakov, and S. V. Troitsky, Multi-particle processes and semiclassical analysis in bosonic field theories, *Phys. Part. Nucl.* **28**, 217 (1997).
- [17] J. Polchinski, *An Introduction to the Bosonic String*, String Theory Vol. 1 (Cambridge University Press, Cambridge, England, 1998).
- [18] E. D'Hoker and S. B. Giddings, Unitarity of the closed bosonic Polyakov string, *Nucl. Phys. B* **291**, 90 (1987).
- [19] H. Sonoda, Functional determinants on punctured Riemann surfaces and their application to string theory, *Nucl. Phys. B* **294**, 157 (1987).

- [20] J. H. Hubbard, *Teichmüller Theory* (Matrix editions, New York, 2006), Vol. 1.
- [21] E. Witten, The Feynman *ic* in string theory, *J. High Energy Phys.* **04** (2015) 055.
- [22] A. Sen, Gauge invariant 1PI effective action for superstring field theory, *J. High Energy Phys.* **06** (2015) 022.
- [23] P. Zograf, The Weil-Petersson volume of the moduli space of punctured spheres, *Contemp. Math.* **150**, 367 (1993).
- [24] P. Zograf, On the large genus asymptotics of Weil-Petersson volumes, [arXiv:0812.0544](https://arxiv.org/abs/0812.0544).
- [25] M. Mirzakhani, Simple geodesics and Weil-Petersson volumes of moduli spaces of bordered Riemann surfaces, *Inventiones Mathematicae* **167**, 179 (2006).
- [26] M. Mirzakhani, Growth of Weil-Petersson volumes and random hyperbolic surface of large genus, *J. Diff. Geom.* **94**, 267 (2013).
- [27] S. H. Shenker, The Strength of nonperturbative effects in String theory, in *Cargese Study Institute: Random Surfaces, Quantum Gravity and Strings* (Plenum Press, New York, 1990), p. 191.
- [28] D. J. Gross and V. Periwal, String Perturbation Theory Diverges, *Phys. Rev. Lett.* **60**, 2105 (1988).
- [29] E. D'Hoker and D. H. Phong, The geometry of string perturbation theory, *Rev. Mod. Phys.* **60**, 917 (1988).
- [30] S. A. Wolpert, Asymptotics of the spectrum and the Selberg zeta function on the space of Riemann surfaces, *Commun. Math. Phys.* **112**, 283 (1987).
- [31] J. A. Hempel and S. J. Smith, Hyperbolic lengths of geodesics surrounding two punctures, *Proc. Am. Math. Soc.* **103**, 513 (1988).
- [32] A. F. Beardon, The uniformisation of a twice-punctured disc, *Comput. Meth. Funct. Th.* **12**, 585 (2012).
- [33] E. D'Hoker and D. H. Phong, Loop amplitudes for the fermionic string, *Nucl. Phys.* **B278**, 225 (1986).
- [34] M. B. Green, J. H. Schwarz, and E. Witten, *Introduction, Superstring Theory* Vol. 1, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 1988).
- [35] P. Sarnak, Determinants of Laplacians, *Commun. Math. Phys.* **110**, 113 (1987); E. D'Hoker and D. Phong, On determinants of Laplacians on Riemann surfaces, *Commun. Math. Phys.* **104**, 537 (1986).
- [36] D. J. Gross and P. F. Mende, String theory beyond the Planck scale, *Nucl. Phys.* **B303**, 407 (1988); The high-energy behavior of string scattering amplitudes, *Phys. Lett. B* **197**, 129 (1987).
- [37] F. Cachazo, S. He, and E. Y. Yuan, Scattering of Massless Particles in Arbitrary Dimensions, *Phys. Rev. Lett.* **113**, 171601 (2014); Scattering of massless particles: Scalars, gluons and gravitons, *J. High Energy Phys.* **07** (2014) 033.
- [38] F. Cachazo, S. He, and E. Y. Yuan, Scattering equations and Kawai-Lewellen-Tye orthogonality, *Phys. Rev. D* **90**, 065001 (2014).
- [39] R. Kleiss, W. Stirling, and S. Ellis, A new monte carlo treatment of multiparticle phase space at high energies, *Comput. Phys. Commun.* **40**, 359 (1986).
- [40] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, England, 1982), Vol. 2.
- [41] L. Mason and D. Skinner, Ambitwistor strings and the scattering equations, *J. High Energy Phys.* **07** (2014) 048.
- [42] R. Britto, F. Cachazo, and B. Feng, New recursion relations for tree amplitudes of gluons, *Nucl. Phys.* **B715**, 499 (2005); R. Britto, F. Cachazo, B. Feng, and E. Witten, Direct Proof of Tree-Level Recursion Relation in Yang-Mills Theory, *Phys. Rev. Lett.* **94**, 181602 (2005).
- [43] D. N. Page, Average Entropy of a Subsystem, *Phys. Rev. Lett.* **71**, 1291 (1993).
- [44] L. Susskind and L. Thorlacius, Gedanken experiments involving black holes, *Phys. Rev. D* **49**, 966 (1994).
- [45] P. Hayden and J. Preskill, Black holes as mirrors: Quantum information in random subsystems, *J. High Energy Phys.* **09** (2007) 120.
- [46] E. Lieb and M. Ruskai, Proof of the strong subadditivity of quantum-mechanical entropy, *J. Math. Phys. (N.Y.)* **14**, 1938 (1973); E. H. Lieb and M. B. Ruskai, A Fundamental Property of Quantum-Mechanical Entropy, *Phys. Rev. Lett.* **30**, 434 (1973).
- [47] S. D. Mathur, The information paradox: A pedagogical introduction, *Classical Quantum Gravity* **26**, 224001 (2009); A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, Black holes: Complementarity or firewalls?, *J. High Energy Phys.* **02** (2013) 062.
- [48] K. Papadodimas and S. Raju, An infalling observer in AdS/CFT, *J. High Energy Phys.* **10** (2013) 212; The unreasonable effectiveness of exponentially suppressed corrections in preserving information, *Int. J. Mod. Phys. D* **22**, 1342030 (2013).
- [49] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, An apologia for firewalls, *J. High Energy Phys.* **09** (2013) 018; D. Marolf and J. Polchinski, Gauge/Gravity Duality and the Black Hole Interior, *Phys. Rev. Lett.* **111**, 171301 (2013).
- [50] S. Raju, Smooth causal patches for AdS black holes, *Phys. Rev. D* **95**, 126002 (2017); The sensitivity of black holes to low energy excitations, [arXiv:1703.10159](https://arxiv.org/abs/1703.10159).
- [51] S. D. Mathur, Tunneling into fuzzball states, *Gen. Relativ. Gravit.* **42**, 113 (2010); I. Bena, D. R. Mayerson, A. Puhm, and B. Vercoe, Tunneling into microstate geometries: Quantum effects stop gravitational collapse, *J. High Energy Phys.* **07** (2016) 031.
- [52] S. B. Giddings, Locality in quantum gravity and string theory, *Phys. Rev. D* **74**, 106006 (2006).
- [53] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth *et al.*, *GNU Scientific Library Reference Manual* (Network Theory Ltd., 2009), <https://www.gnu.org/software/gsl/manual/gsl-ref.pdf>.
- [54] M. J. Powell, *Numerical Methods for Nonlinear Algebraic Equations* (Gordon and Breach Science Publishers, 1970), pp. 87–114.
- [55] A. M. Ferrenberg, D. Landau, and Y. J. Wong, Monte Carlo simulations: Hidden errors from “good” random number generators, *Phys. Rev. Lett.* **69**, 3382 (1992).
- [56] M. Luscher, A portable high quality random number generator for lattice field theory simulations, *Comput. Phys. Commun.* **79**, 100 (1994).
- [57] O. Tange, login: The USENIX Magazine (2011), <https://www.usenix.org/system/files/login/articles/105438-Tange.pdf>.