The tortoise and the hare: a causality puzzle in AdS/CFT

David Berenstein∗ and David Grabovsky

Department of Physics, University of California, Santa Barbara, CA 93106, United States of America

E-mail: dberens@physics.ucsb.edu

Received 16 December 2020, revised 11 March 2021
Accepted for publication 24 March 2021
Published 20 April 2021

Abstract

We pose and resolve a holographic puzzle regarding an apparent violation of causality in anti-de Sitter (AdS)/conformal field theory. If a point in the bulk of AdS moves at the speed of light, the boundary subregion that encodes it may need to move superluminally to keep up. With AdS3 as our main example, we prove that the finite extent of the encoding regions prevents a paradox. We show that the length of the minimal-size encoding interval gives rise to a tortoise coordinate on AdS that measures the nonlocality of the encoding. We use this coordinate to explore circular and radial motion in the bulk before passing to the analysis of bulk null geodesics. For these null geodesics, there is always a critical encoding where the possible violation of causality is barely avoided. We show that in any other encoding, the possible violation is subcritical.

Keywords: anti-de Sitter/conformal field theory, entanglement wedges, causality violations

(Some figures may appear in colour only in the online journal)

1. Introduction

Invitation. Imagine that you are wandering a circular room illuminated only by a lamp in its center. Your shadow dances on the walls as you move, by turns larger as you near the center and smaller as you near the edge. If you move at the speed of light, your shadow will have to move even faster to keep up with you1. Ordinarily this is not a problem, since shadows carry no physical information. But this is a metaphor for holographic duality, where both you and your shadows are real. The room is anti-de Sitter (AdS) space, the walls are home to a

∗Author to whom any correspondence should be addressed.

1Apparent faster-than-light travel is common in the observation of quasar jets in astrophysics [1, 2]. A detailed understanding of the kinematics of light emission by matter—when, from where, and in which direction they are emitted and observed—resolves the apparent paradox here.

1361-6382/21/105008+22$33.00 © 2021 IOP Publishing Ltd Printed in the UK
conformal field theory (CFT) on its boundary, and you are a local bulk field configuration—a particle—encoded nonlocally on the walls. An apparent paradox emerges: lightspeed motion in the bulk seems to violate causality on the boundary!

**AdS/CFT and entanglement wedges.** The AdS/CFT correspondence states that a semiclassical theory of quantum gravity in an asymptotically AdS spacetime is equivalent to a CFT on the conformal boundary of that spacetime [3–5]. Even if one only has access to the quantum information contained in a subregion of the boundary, one can completely reconstruct a corresponding part of the bulk theory. This is usually exemplified by the HKLL construction [6] (see also [7] for earlier work on this). More precisely, suppose that all of the local and nonlocal CFT operators supported on some subregion of the boundary are given. Then there is a region of AdS, called the entanglement wedge of the boundary subregion, within which one can realize any bulk field as a nonlocal CFT operator smeared over the boundary subregion.

For static bulk backgrounds, there are canonical constant-time slices orthogonal to a time-like Killing field. In this case, the entanglement wedge is found using the Ryu–Takayanagi prescription [8, 9]: one studies the minimal-area surface in the bulk, anchored on the boundary of the region, which is homologous to it and lives in the given constant-time slice. The entanglement wedge is then the region bounded by this surface and the boundary. For pure states, the boundary subregion complementary to a given one ends up encoding the corresponding complementary region of the bulk [11]. There are some additional restrictions, in that the operators supported on the boundary region cannot cause a large backreaction in the bulk: this is part of the idea that the reconstruction occurs only in a code subspace [12].

This has a particularly simple realization in pure AdS$_3$, which can be visualized as a solid cylinder whose constant-time slices are Poincaré disks $D^2$. The geodesics of $D^2$, shown in figure 1, are circles orthogonal to its boundary. Each geodesic subtends an interval of the boundary circle, which we call the *encoding interval*, and is uniquely labeled by its ‘deepest’ bulk point, which we call its *gem*. As we enlarge the encoding interval, its gem moves further toward the center of the disk. The constant-time slice of the entanglement wedge of a given encoding interval, shaded in figure 1, is the region of $D^2$ enclosed by the boundary of $D^2$ and the subtending geodesic. We regard the geodesic itself as part of the entanglement wedge: if we enlarge the encoding interval infinitesimally, the points on the geodesic will lie in the entanglement wedges of both the encoding interval and its complement (this follows the ‘all-or-nothing’ point of view of [13]). Each boundary interval encodes all bulk points $p \in D^2$ caught in its entanglement wedge, and conversely, $p$ is encoded on any boundary interval that has $p$ in its entanglement wedge. Among these intervals is a unique one of minimal length. Its gem is $p$ itself, and it provides the ‘most efficient’, or minimal, encoding of $p$.

**A causal puzzle.** Our goal in this paper is to better understand causality from the perspective of encodings. This turns out to be very nontrivial, as we explain presently. The null motion of a local bulk field in AdS may drag (the midpoint of) its encoding interval around the boundary faster than light. This is easiest to visualize in another geometry: AdS black holes. There one can show that circular timelike geodesics in the bulk have angular velocities greater than one [14], the speed of light on the boundary. Finding paradoxical configurations takes slightly more effort in global AdS. Naively, null motion in AdS could cause the initial and final encoding regions to become spacelike-separated, causing the operators supported on each boundary region to mutually commute. The bulk point would be unable to carry information from one interval to the other. To carry information, an action must have a nontrivial response. In linear
response theory, one needs a non-vanishing retarded Green’s function, computed by a commutator between operators in the initial and final boundary regions. If all of the commutators vanish by boundary causality, but the motion of the bulk point is our means of conveying a message, then we have a contradiction. We can make matters worse by using an encoding where the bulk point starts near the edge of its encoding interval in its direction of motion, but ends up near the opposite edge when it arrives. Making the intervals asymmetric (i.e. non-minimal) and having them ‘swing around’ can, in principle, exacerbate the problem.

We can only resolve this puzzle if the encoding intervals are ‘big enough’. As shown in figure 1, the instant that $p$ departs from its starting point in the bulk, we send a light signal $q$ along the boundary, starting in the initial encoding region and moving at speed $c = 1$ (by convention). Like a photophobic ant hiding in the shadows on the wall, the signal must always stay inside the (moving) encoding interval if causality is to be preserved. The signal must travel from the ‘leading’ edge $a$ of the initial interval to the ‘trailing’ edge $b$ of the final interval before the bulk point arrives; the tightest bounds are obtained if $p$ lies on the geodesic defining its encoding interval instead of strictly inside. Due to the possible asymmetry of the encoding interval about the bulk point, the ‘forward’ signal ($a \rightarrow b$) might fail to preserve causality. Instead, the encoding intervals must bow out far enough to let a ‘backward’ signal ($a' \rightarrow b'$) preserve causality whenever the forward signal does not.

It is worth noting that our investigation bears some similarity to the problem of gravitational time delay studied by Gao and Wald [15, 16]. A special case of the Gao–Wald theorem guarantees that for any causal bulk path that starts and ends on the boundary of empty AdS, a light signal traveling entirely on the boundary will always beat the bulk point. Our purpose here is to slightly extend this instance of the Gao–Wald result to motion deep in the bulk. The discussion above provides a prescription for which boundary paths to compare to the bulk path, and this paper seeks to prove that such a statement holds in empty AdS. More broadly, we expect that this generalized causality will be preserved if the null energy condition is satisfied in the bulk, even for dynamical gravity [17]. In this paper we elaborate on the example illustrated
by figure 1 of [17], which shows that the RT formula marginally satisfies causality on a plane
generated by null geodesics. Here we choose different entanglement wedge encodings for the
bulk point, and then check how close we come to a critical situation where the bulk and the
boundary travel times coincide. The main point is that we can be very explicit: the computa-
tions are done directly using elementary geometry. This should, in principle, also inform us
better as to how the different code subspaces of the bulk theory talk to each other.

Outline. In this paper, we resolve the paradox in several special cases before giving a general
proof. In section 2, we bring the paradox into stark relief by studying circular and radial bulk
motion; along the way, we describe the geometry of encoding intervals and prove that their
length acts as an infalling tortoise coordinate on AdS3. In section 3, we prepare the ground
for the general case: we first parametrize null geodesics, the fastest possible bulk paths, using
the tortoise coordinate; then, we discuss the minimality of encoding intervals in more detail.
In section 4 we prove, in several settings of increasing generality, that the bulk point’s travel
time always exceeds the shortest of the forward and backward signals’ travel times. We also
explain how our main result and proof is valid in AdS3 with small modifications, as long as
we use spherical regions in the boundary as a substitute for encoding intervals. We conclude in
section 5 with a discussion of how tight the bounds on causality are, and discuss generalizations
to higher dimensions and more general spacetimes.

2. An illustration of causality

We begin by putting a particle (what we call a bulk field) in a rocket ship and sending it on
circular and radial trips through AdS3. In both cases we force the rocket to move along a
null path, and we encode the bulk point minimally on the boundary. Even in these simplest
eamples, the encoding region moves superluminally, making the paradox apparent.

2.1. Circular motion

AdS3 has topology \( \mathbb{R} \times D^2 \), where \( \mathbb{R} \) is the time axis with coordinate \( t \) and \( D^2 \) is the Poincaré
disk. Putting (Euclidean) polar coordinates \((r, \theta)\) on \( D^2 \) gives AdS3 the metric

\[
\begin{align*}
\text{d}s_{\text{AdS3}}^2 &= -\left(1 + r^2\right)^2 \text{d}t^2 + \frac{4 \left(2r^2 + r^2 \text{d}\theta^2\right)}{\left(1 - r^2\right)^2} = \text{d}s_{\mathbb{R}}^2 + \text{d}s_{D^2}^2.
\end{align*}
\]

These are sometimes called ‘sausage coordinates’ on AdS [18, 19], and in these coordinates
\( D^2 \) is conformal to a flat disk in polar coordinates. In sausage coordinates, the geodesics of
\( D^2 \) are arcs of circles that meet the boundary at right angles. The coordinate \( r \) should not be
confused with with the standard global AdS coordinate \( \rho \), which measures the distance to the
origin in \( D^2 \). Fixing the \( r \)-coordinate of the bulk point \( p \), let \( \gamma(p) \in [0, \pi/2] \) denote half of the
(Euclidean) arclength of its minimal encoding interval, as shown in figure 2.

Now we send \( p \) along a null, circular worldline, so that \( \text{d}r = 0 \); hence \( \gamma(p) \) remains constant
throughout the trip. We show in section 2.2 that \( \omega \equiv \frac{\text{d}r}{\text{d}t} = \frac{1}{\cos \gamma} > 1 \). Unless \( p \) moves along
the boundary \( (r = 1) \) or sits motionless at the origin \( (r = 0) \), the center point of its encoding
interval always seems to move faster than light. This same construction of a paradoxical
situation can work in spacetimes with spherically symmetric matter or with black holes. The
resolution that we will discuss depends on the precise details of the AdS metric, where we can
explicitly do computations. Issues to consider in these more general cases include the presence
of entanglement shadows and the lack of explicit analytic computations of the entanglement
wedges. Because of these complications, we will consider such setups beyond the scope of the
present work.
To resolve the paradox, we compare the bulk travel time $t_p$ to that of the signals described in section 1. If $p$ moves through angle $\theta$, its transit takes time $t_p = \frac{\theta}{\omega} = \theta \cos \gamma$. Meanwhile, the signals’ travel times are exactly equal to the angles they traverse. Referring to figure 2, the forward signal takes time $t_f = \theta - 2\gamma$, while the backward signal takes time $t_b = 2\pi - \theta - 2\gamma$. (Note that it is no loss of generality to treat the motion in a single constant-time slice, since both the bulk point and the boundary signal move uniformly in $t$.) We now prove that at least one of the signals always beats $p$ to its destination.

**Proposition 2.1.** For all $\theta \in [0, 2\pi)$ and $\gamma \in [0, \frac{\pi}{2}]$, $\Delta t \equiv \max \{\Delta t_f, \Delta t_b\} \geq 0$, where

$\Delta t_f \equiv t_p - t_f = 2\gamma - \theta(1 - \cos \gamma)$,

$\Delta t_b \equiv t_p - t_b = 2\gamma + \theta(1 + \cos \gamma) - 2\pi$. \hspace{1cm} (2)

**Proof.** We begin by noting that both $-\Delta t_f$ and $-\Delta t_b$ are convex functions of $\gamma$. Indeed, consider their second derivatives: $-\Delta t_f''(\gamma) = -\Delta t_b''(\gamma) = \theta \cos \gamma \geq 0$. It remains to ‘anchor down’ $\Delta t_f$ and $\Delta t_b$. When $\theta \in [0, \pi]$, $\Delta t_f(0) = 0$ and $\Delta t_f(\frac{\pi}{2}) = \pi - \theta \geq 0$. Thus $\Delta t_f \geq 0$ for $\theta \in [0, \pi]$ by convexity and the intermediate value theorem. A similar argument applies to $\Delta t_b$: when $\theta \in [\pi, 2\pi]$, $\Delta t_b(0) = 2\theta - 2\pi \geq 0$ and $\Delta t_b(\frac{\pi}{2}) = \theta - \pi \geq 0$; hence $\Delta t_b \geq 0$ for $\theta \in [\pi, 2\pi]$. (For $\theta = \pi$, we have $t_f = \pi - 2\gamma = t_b \leq t_p = \pi \cos \gamma$.)

We have proven that the forward signal always beats the bulk point on short trips, that the backward signal always wins on long trips, and moreover that they trade off at $\theta = \pi$.

Apparent motion faster than light also occurs when one studies the phase velocity of a wave. Stretching the analogy, we might imagine the midpoint of the entanglement wedge as moving with the phase velocity, while the time of flight between the intervals’ endpoints measures the group velocity, in this case the speed of light. This analogy is somewhat strained: as we discuss...
in detail in the rest of the paper, we have a good deal of freedom in choosing entanglement wedges. There is no canonical notion of phase velocity independent of the choices we make, owing essentially to the nonlocality of the bulk encoding.

2.2. The tortoise coordinate

The calculation above demonstrates that the resolution of the paradox relies crucially on the finite extent of $\gamma(p)$. Evidently $\gamma$ is simultaneously a measure of (1) the size of the encoding interval of $p$, (2) the nonlocality of its encoding, (3) the ‘danger’ $p$ poses to causality, and (4) the depth of $p$ in the bulk. To explore this last observation, it will be convenient to study $D^2 \subset \mathbb{R}^2$ in terms of Euclidean plane geometry.

Notice that $\gamma(p)$ depends only on the radial distance $r(p)$ of $p$ from the origin. As $r \to 0$, we have $\gamma \to \frac{\pi}{2}$; as $r \to 1$, we have $\gamma \to 0$. Thus inside $D^2$, $\gamma$ parametrizes the radial position of $p$. To relate $\gamma$ to $r$, we let $\rho$ be the radius of the orthogonal circle defining the minimal geodesic through $p$, as shown in figure 3. The Pythagorean theorem gives

$$\rho^2 + 1 = (\rho + r)^2 \implies \rho = \frac{1 - r^2}{2r}. \quad (3)$$

Since $\gamma$ is half of the angle subtended by this arc, we find

$$\sin \gamma = \frac{\rho}{\rho + r} = \frac{1 - r^2}{1 + r^2}, \quad \cos \gamma = \frac{1}{\rho + r} = \frac{2r}{1 + r^2}, \quad \tan \gamma = \rho = \frac{1 - r^2}{2r}. \quad (4)$$

All three of these relations will prove useful. The last one yields a quadratic equation relating $r$ and $\gamma$, of which we take the positive root to ensure that $r > 0$:

$$1 - r^2 = 2r \tan \gamma \implies r = \sec \gamma - \tan \gamma. \quad (5)$$

Substituting the change of coordinates (5) in the metric (1), we rewrite the AdS$_3$ metric:

$$ds^2_{AdS_3} = -\left(\frac{1 + r^2}{1 - r^2}\right)^2 dr^2 + \frac{4 (dr^2 + r^2 d\theta^2)}{(1 - r^2)^2} = \frac{1}{\sin^2 \gamma} (-dr^2 + d\gamma^2 + \cos^2 \gamma d\theta^2). \quad (6)$$
These new coordinates explain the angular velocity quoted above: if $p$ stays at fixed radius ($d\gamma = 0$) and has null trajectory ($ds^2 = 0$), then we find $dt = \pm \cos \gamma \, d\theta$, so $\omega = \frac{1}{\cos \gamma}$.

2.3. Radial infall

The physical meaning of $\gamma$ is revealed when we consider the metric induced on a radial ($d\theta = 0$) null worldline. Since $\gamma$ and $t$ enter into $ds^2$ with no relative factors,

$$ds^2 = 0 \Rightarrow dt = \pm d\gamma \quad \Rightarrow \quad \gamma - \gamma^* = \pm t. \quad (7)$$

We interpret $\gamma$ as an infalling Eddington–Finkelstein, or tortoise, coordinate [20]. This analysis extends to AdS spacetimes of any dimension, where one needs a higher-dimensional analog of the orthogonal circles in $D^2$. If we use spherical regions of the boundary as substitutes for encoding intervals, then the minimal surfaces anchored on the boundary will be sphere caps orthogonal to the boundary; see for example [21].

Let us examine radial null paths, shown in figure 4. The null-worldline condition gives

$$ds_p^2 = \frac{1}{\sin^2 \gamma} \left( -dr^2 + d\gamma^2 \right) = 0 \Rightarrow dr = \pm d\gamma \quad \Rightarrow \quad v \equiv \left| \frac{d\gamma}{dt} \right| = 1. \quad (8)$$

The unit speed of light is expected: the purpose of $\gamma$ is to parametrize radial light rays as if they were signals. As $p$ travels into the bulk, $\gamma$ expands until $p$ reaches the origin. As it crosses the origin, the encoding interval—which occupies half of the boundary—discontinuously jumps to occupy the opposite half and begins to shrink. Even though the encoding interval moves with infinite speed when $\gamma(p) = \frac{\pi}{2}$, its size again prevents a paradox. Regardless of whether $p$ crosses the origin, its travel time matches that of both signals:

$$t_p(\gamma_1 \rightarrow \gamma_2) = \gamma_2 - \gamma_1 = t_f = t_b, \quad (p \text{ does not cross the origin}),$$

$$t_p(\gamma_1 \rightarrow O \rightarrow \gamma_2) = \pi - \gamma_2 - \gamma_1 = t_f = t_b, \quad (p \text{ does cross the origin}). \quad (9)$$

Figure 4. Radial motion in the bulk.
In both cases, the time delay vanishes identically. Both the forward and backward signals just barely preserve the causality, independent of the initial and final positions $\gamma_1, \gamma_2$. Thus the encoding of $p$ is critical throughout its trip through the bulk.

The examples above illustrate that roughly speaking, $\gamma$ measures causality in the bulk, while $\theta$ measures causality on the boundary. The example of circular motion introduces trade-offs between the forward and backward signals, while the example of radial infall forces us to extend $\gamma$ beyond the origin, and to confront the discontinuity of encoding there.

3. Coordinate gymnastics

Let us strategize. The examples above fail to be general on two counts: (1) circular and radial paths are not generic, and the circular paths are ‘too slow’ in the bulk; and (2) the bulk encoding need not be minimal. To remedy (1), we should study bulk motion along geodesics; in particular, we will use null geodesics, the ‘fastest’ causal curves connecting two bulk points. This is the path of the hare, who tries to win a race by moving as fast as possible on the shortest path. We want to make sure that a signal—the tortoise—makes its trip along the boundary in the shortest time and beats the hare. To remedy (2), we need to quantify the (non-)minimality of encoding intervals in a tractable way.

3.1. Null geodesics

Explicit parametrization. The details of the first task are relegated to appendix A, where we show that the geodesic equations in AdS$_3$ are solved in $(t, \gamma, \theta)$ coordinates by

$$t(\tau) = \tau, \quad \gamma(\tau) = \tan^{-1}\left(\frac{\sqrt{1 - \ell^2} \tan \tau}{\sqrt{1 + \ell^2 \tan^2 \tau}}\right), \quad \theta(\tau) = \tan^{-1}(\ell \tan \tau). \quad (10)$$

Here $\ell \in [-1, 1]$ plays the role of the angular momentum of $p$. Since $\gamma(\tau)$ and $\theta(\tau)$ are singular at $\tau = \pi/2$, we extend their definitions ‘by hand’, being careful about the branch cut choices for the inverse tangent. The geodesics are parametrized piecewise:

$$\gamma_*(\tau) = \begin{cases} \gamma(\tau), & \tau \in [0, \pi/2], \\ \gamma(\pi - \tau), & \tau \in [\pi/2, \pi] \end{cases}, \quad \theta_*(\tau) = \begin{cases} \theta(\tau), & \tau \in [0, \pi/2], \\ \pi - \theta(\pi - \tau), & \tau \in [\pi/2, \pi] \end{cases}. \quad (11)$$

Conformal structure. We can gain some intuition by looking at the Penrose diagram of AdS$_3$ [22]. If we define a shifted coordinate $\gamma \to \gamma = \gamma + \pi/2$, the AdS metric (6) becomes

$$ds^2_{AdS_3} = \frac{1}{\cos^2 \gamma} \left(-dt^2 + d\gamma^2 + \sin^2 \gamma d\theta^2\right) \sim -dt^2 + d\Omega^2_2 = ds^2_M, \quad (12)$$

where $\sim$ denotes conformal equivalence. This provides a check that AdS is conformal to a patch of the Einstein static Universe $M = \mathbb{R} \times S^2$ [20]. It also shows how the radial direction of AdS$_3$ is folded up into an angular direction of $S^2$ upon conformal rescaling. Finally, it gives a clearer understanding of the null geodesics of AdS$_3$, which must be the same as those of $M$. Since $M$ has a product metric, the $t$ coordinate of its geodesics is independent of its $(\gamma, \theta)$ coordinates. The former is given by $t(\tau) = E\tau + t_1$, while the latter is determined by

$$\cot \gamma = a \cos(\theta - \theta_1) \iff \tan \gamma = -a \cos(\theta - \theta_1). \quad (13)$$

---

4 We mean the term ‘bulk points’ in terms of their coordinates on a constant-time slice.
Here $E, t_1, a$, and $\theta_1$ are integration constants; we set $t_1 = \theta_1 = 0$ and $E = 1$ to match the analysis in appendix A. These geodesics are great circles in $S^2$, so they must meet the equator at diametrically opposite points. Converting back from $\gamma$ to the standard radial coordinate $r$, we find a new parametrization of the null geodesics:

$$r = r(\theta) = \sec \gamma - \tan \gamma = a \cos \theta + \sqrt{1 + a^2 \cos^2 \theta}.$$  

(14)

This curve describes a circular arc intersecting the unit circle at opposite points. Indeed, one may check that $(x - a)^2 + y^2 = 1 + a^2$ for $x = r \cos \theta$ and $y = r \sin \theta$. Here $a$ plays the role of $\ell$ by controlling how close the geodesic comes to the boundary, thereby giving the bulk point angular momentum. The point of closest approach is at $(r_{\text{min}}, \theta_{\text{min}}) = (\sqrt{1 + a^2} - a, \pi)$, when $a > 0$. Thus both spacelike and null geodesics in AdS$_3$ are circular arcs, the former orthogonal to the boundary and the latter passing through diametrically opposite points.

3.2. Non-minimal encoding

Gems and mimic points. Observe that every geodesic of $D^2$ is uniquely labeled by its ‘deepest’ bulk point. We call this point the gem of the geodesic. Suppose that $p_0 = (\gamma_0, \theta_0)$ is the gem of some geodesic in $D^2$, and let $p = (\gamma, \theta)$ be an arbitrary point on the same geodesic. Then $p_0$ and $p$ may both be encoded by the same boundary interval, but $p_0$ is encoded minimally while $p$ may not be: it is as if $p$ is pretending to be at $p_0$.

As shown in figure 5, the coordinates of $p$ and $p_0$ are related by the law of cosines:

$$\rho^2 = r^2 + (\rho + r_0)^2 - 2r(\rho + r_0)\cos(\theta_0 - \theta).$$  

(15)

We divide through by $(\rho + r_0)^2$ and recognize $\sin \gamma_0$ from (4). After some algebra, we obtain

$$\sin^2 \gamma_0 = r^2 \cos^2 \gamma_0 + 1 - 2r \cos \gamma_0 \cos(\theta_0 - \theta) \implies \cos \gamma_0 = \cos(\theta_0 - \theta) \cos \gamma.$$  

(16)

We introduce $\alpha \equiv \theta_0 - \theta \in [-\pi/2, \pi/2]$ to measure the degree of encoding asymmetry. We also change notation $p_0 \to p_\alpha = (\gamma_\alpha, \theta_\alpha)$ to emphasize $p_\alpha$ as a ‘mimic’ point, disheveled from $p$ by angle $\alpha$. This lets us pretend that the bulk encoding is always minimal, as long as we use

![Figure 5. Coordinates of a point lying on a geodesic.](image)
\((\gamma_\alpha, \theta_\alpha)\) instead of \((\gamma, \theta)\). In this notation, our results are
\[
\cos \gamma_\alpha = \cos \alpha \cos \gamma, \quad \theta_\alpha = \alpha + \theta, \quad \alpha \in \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right]. \tag{17}
\]

**Locus of gems.** The discussion above begs a natural question: where can a bulk point pretend to hide? More precisely, what is the locus of gems of geodesics passing through a fixed point \(p\) with coordinates \((\gamma, \theta)\)? It is just the set of points with coordinates in
\[
G_p = \left\{ (\gamma_\alpha, \theta_\alpha) \in \left[ 0, \frac{\pi}{2} \right] \times S^1 \mid \cos \gamma_\alpha = \cos(\theta_\alpha - \theta) \cos \gamma \right\}. \tag{18}
\]

The defining equation of \(G_p\) is one characterization of its shape; another comes from a conversion of \((\gamma_\alpha, \theta_\alpha)\) into polar \((r_\alpha, \theta_\alpha)\) and then into Cartesian \((x, y)\) coordinates. We may use the rotational invariance of \(D^2\) to set \(\theta = 0\) for the moment; then, using (4), we have
\[
\frac{2r_\alpha}{1 + r^2_\alpha} = \cos \theta_\alpha \cos \gamma \implies \kappa r^2_\alpha = (1 + r^2_\alpha) \cos \theta_\alpha
\]
\[
\implies x^3 + xy^2 - \kappa (x^2 + y^2) + x = 0, \quad \kappa \equiv \frac{2}{\cos \gamma} \in [2, \infty).
\tag{19}
\]

So as a curiosity, the set of bulk points encoded by the same boundary interval—the points as which \(p\) can masquerade for the purposes of causality—is a cubic curve in \(D^2\).

### 4. Analysis of causality

**Physical setup.** We aim to prove that for motion along null geodesics, and hence along any causal curve, the tortoise on the boundary always beats the hare in the bulk. The latter, a freely falling massless observer \(p \in \text{AdS}_3\), moves from \(p(\tau_1)\) to \(p(\tau_2)\) along a curve parametrized by \(p(\tau) = (\tau, \gamma(\tau), \theta(\tau))\) as in (10) and (11). Because \(\ell(\tau) = \tau\), its trip takes time \(t_p = \tau_2 - \tau_1\).

Meanwhile, the travel time of the signals depends on the encoding of \(p\) on a boundary region subtended by a spacelike geodesic passing through it. This region may encode \(p(\tau)\) non-minimally; as shown in figure 6, one way to describe the encoding is to let \(\alpha(\tau)\) be the angle by which \(p(\tau)\) is disheveled from the gem of its encoding geodesic.

**Basic strategy.** We endeavor to show that at least one boundary signal beats the bulk point as it travels along null geodesics. In other words, we want to prove\(^5\) that
\[
\Delta t^\prime(\tau_1, \tau_2) \equiv \max\{\Delta t^\prime_f, \Delta t^\prime_b\} \equiv \max\{t_p - t^\prime_f, t^\prime_b - t_p\}
\]
\[
= (\tau_2 - \tau_1) - \min\{t^\prime_f, t^\prime_b\} \geq 0, \tag{20}
\]
where \(\Delta t^\prime_f \equiv t_p - t^\prime_f\) and \(\Delta t^\prime_b \equiv t^\prime_b - t_p\) are what we might call the forward and backward holographic time delay functions for \(\text{AdS}_3\).

Referring to figure 6, the forward and backward signals’ travel times are equal to the angles between the neighboring edges of the shaded regions:
\[
t^\prime_f(\tau_1, \tau_2) \equiv (\gamma_{\alpha_f}(\tau_2) + \gamma_{\alpha_f}(\tau_1)) - (\theta_{\alpha_f}(\tau_2) - \theta_{\alpha_f}(\tau_1)),
\]
\[
t^\prime_b(\tau_1, \tau_2) \equiv (\gamma_{\alpha_b}(\tau_2) + \gamma_{\alpha_b}(\tau_1)) + (\theta_{\alpha_b}(\tau_2) - \theta_{\alpha_b}(\tau_1)) - 2\pi.
\tag{21}
\]

\(^5\) Here and below, (e.g.) \(\gamma_{\alpha_i}(\tau)\) stands for \(\gamma_{\alpha_i}(\tau_i)\). In particular, for minimal encoding, \(\gamma_{\alpha_i}(\tau) = \gamma_i(\tau)\).
Figure 6. Null geodesic motion: minimal (left, $\alpha(\tau) \equiv 0$) and non-minimal (right, $\alpha_1, \alpha_2 \neq 0$) encodings. Minimal encodings (thin curves) and gems of $p$ (open circles) are also shown.

The case of minimal encoding, $\alpha(\tau) \equiv 0$, yields in subsection 4.1 to direct attack by a convexity argument much like the one in the analysis of circular motion. For $\alpha(\tau) \neq 0$, proving directly that $\Delta t \geq 0$ for $\alpha(\tau) \neq 0$ is still possible, but is neither elegant nor illuminating. Instead, in subsection 4.2, we observe that for any path through the bulk, there is a unique critical encoding $\alpha_*(\tau)$ for which both the forward and backward boundary signals arrive simultaneously with the bulk point: $t_f = t_b = t_p \Rightarrow \Delta t \equiv 0$. We then study deviations from this critical encoding, as measured by a new angle $\beta(\tau)$. In subsection 4.3, we re-coordinatize AdS$_3$ using this critical encoding. This converts the claim of causality preservation into a statement of spherical trigonometry, which completes the proof.

**Miscellaneous comments.** We begin by restricting the ranges of $\ell$ and $\tau_1$. First, while $\ell \in [-1, 1]$ labels all null geodesics, we consider only $\ell \in [0, 1]$ because the paths labeled by $\ell$ and $-\ell$ have identical time delays. The time delay preserves the rotational U(1) symmetry of $D^2$, as well as a $\mathbb{Z}_2$ reflection symmetry: fixing the boundary point $p(0)$ breaks the U(1), and choosing the sign of $\ell$ breaks the remaining $\mathbb{Z}_2$. Second, although $0 \leq \tau_1 \leq \tau \leq \tau_2 \leq \pi$ labels all points on a null geodesic, we consider only $\tau_1 \in [0, \frac{\pi}{2}]$. Every path from $p(\tau_1)$ to $p(\tau_2)$ with $\frac{\pi}{2} \leq \tau_1 < \tau_2$ starts over halfway into $D^2$ and lies entirely in one quadrant. The $\frac{\pi}{2}$-reflected path $p(\pi - \tau_2) \rightarrow p(\pi - \tau_1)$ lies in the quadrant where $\tau_1 < \tau_2 \leq \frac{\pi}{2}$, and has identical time delay. This choice uses up the inversion symmetry, preserved by $\Delta \tau$, of our geodesics. Next, by way of notation, let us define the function $f(\tau)$ and the quantity $L$ by

$$f^2(\tau) \equiv 1 + \ell^2 \tan^2 \tau \geq 1, \quad L^2 \equiv 1 - \ell^2 \in [0, 1]. \quad (22)$$

Note that the function $f(\tau)$ appears in (10), letting us rewrite it as

$$\gamma(\tau) = \tan^{-1} \left( \frac{L \tan \tau}{f(\tau)} \right), \quad \theta(\tau) = \tan^{-1} (\ell \tan \tau). \quad (23)$$

Finally, it may be checked that $\Delta t_f$ and $\Delta t_b$ are both continuous—moreover, at least $C^1$—in both $\tau_1$ and $\tau_2$. We will make implicit use of this regularity many times.
4.1. Causality I: minimal encoding

Our first serious result is that causality is preserved for minimally encoded bulk points moving along null geodesics. For convenience, we will omit the superscript α and the argument τ₁ when they vanish, writing (e.g.) Δα(0, τ₂) as Δ(τ₂). We begin by additionally enforcing τ₁ = 0, making the bulk point start its journey from the boundary.

Proposition 4.1 (Boundary to bulk). For all ℓ ∈ [0, 1] and all τ₂ ∈ [0, π], we have Δf(τ₂) ≡ 0, while Δb(τ₂) ≤ 0. Thus the forward signal always preserves causality, while the backward signal lags behind the bulk point.

Proof. We construct the graph of Δ(τ₂) in figure 7 by establishing the following claims:

Step 1. −Δf(τ₂) is convex in τ₂ on [0, π/2]. Also, Δf(0) = 0, and Δf(π/2) ≥ 0.

Step 2. Δf(τ₂) is decreasing in τ₂ on [π/2, π]. Also, Δf(π/2) ≥ 0, and Δf(π) = 0.

Step 3. Δb(τ₂) is increasing in τ₂ on [0, π]. Also, Δb(0) ≤ 0, and Δb(π) = 0.

Steps 1 and 2 prove that Δf(τ₂) ≥ 0, while step 3 proves that Δb(τ₂) ≤ 0.

Step 1. To show that Δf(τ₂) ≡ Δf(τ₂) ≡ 0 for τ₂ ∈ [0, π/2] is a direct computation:

\[ \Delta f(\tau_2) = \gamma''(\tau_2) - \theta''(\tau_2) = -\frac{\ell \sec^2 \tau_2}{f^2(\tau_2)} \left[ 2L^2 + \ell L f(\tau_2) \right] \leq 0. \] (24)

It is also readily checked that Δf(0) = 0 and Δf(π/2) = tan⁻¹(\(L/\ell\)) ≥ 0.

Step 2. When τ ∈ [π/2, π], we use coordinates (γ, θ). We check monotonicity for Δf(τ₂):

\[ \Delta f(\tau_2) = \gamma'\prime(\tau_2) - \theta'\prime(\tau_2) = 1 - \frac{\ell \sec^2 \tau_2}{f^2(\tau_2)} - \frac{L}{f(\tau_2)} \] (25)

To see that this derivative is never positive, we multiply through by \(f^2(\tau_2)\). Then use that \(\ell \geq \ell^2\) and \(f(\tau_2) \geq 1\) to bound the last two terms:

\[ f^2(\tau_2) - \ell \sec^2 \tau_2 - L f(\tau_2) \leq 1 + \ell^2 \tan^2 \tau_2 - \ell^2 \sec^2 \tau_2 - \frac{L^2}{L} \leq 0. \] (26)

It is also direct to check that Δf(π) = 0.

Step 3. We check monotonicity separately for Δb(τ₂) on [0, π/2] and on [π/2, π]:

\[ \Delta b(\tau_2) = \begin{cases} 1 + \frac{\ell \sec^2 \tau_2}{f^2(\tau_2)} + \frac{L}{f(\tau_2)} & \tau_2 \in [0, \pi/2], \\ 1 + \frac{\ell \sec^2 \tau_2}{f^2(\tau_2)} - \frac{L}{f(\tau_2)} & \tau_2 \in [\pi/2, \pi]. \end{cases} \] (27)

For \(\tau_2 \in [0, \pi/2]\), the nonnegativity of Δb(τ₂) is manifest. Meanwhile, for \(\tau_2 \in [\pi/2, \pi]\), we have

\[ \Delta b(\tau_2) = -\Delta f(\tau_2) + 2 - \frac{2L}{f(\tau_2)} \geq 2 - 2 = 0, \] (28)

where we have used Δf(τ₂) ≲ 0 (step 2) and \(\frac{L}{f(\tau_2)} \leq 1\). It is once again straightforward to see that Δb(0) = 2π > 0 and Δb(π) = 0. This completes the proof.

The content of the proof is visualized in figure 7, where monotonicity and convexity are apparent. Many of the inequalities above are saturated at \(\ell = 0\), which describes the radial plunge trajectory of subsection 2.3. In this case, Δf(τ₂) is a straight line for \(\tau_2 \in [0, \pi/2]\), and abruptly goes to zero at \(\tau_2 = \pi/2\). Similarly, Δb(τ₂) is a straight line that reaches zero at \(\tau_2 = 3\pi/2\).
and stays at that value until $\tau_2 = \pi$. And when $\ell = 1$, the ‘bulk’ point moves on the boundary together with the forward signal, so in that case $\Delta t_f \equiv 0$.

We now remove the restriction that $\tau_1 = 0$; the symmetry argument below shows that starting from the bulk poses no danger to causality.

**Proposition 4.2 (Bulk to bulk).** For all $\ell \in [0,1]$, $\tau_1 \in [0,\pi/2]$, and $\tau_2 \in [\tau_1, \pi]$, we have $\Delta t_f(\tau_1, \tau_2) \geq \Delta t_f(0, \tau_2) \geq 0$, while $\Delta t_b(\tau_1, \tau_2) \leq \Delta t_b(0, \tau_2) \leq 0$. Thus moving into the bulk only drives $\Delta t_f$ and $\Delta t_b$ apart, and $\Delta t_f$ continues to preserve causality.

**Proof.** Observe that we can rewrite the time delay $\Delta t(\tau_1, \tau_2)$ of (20) in terms of $\Delta t(0, \tau_2)$:

$$
\Delta t_f(\tau_1, \tau_2) = \Delta t_f(\tau_2) + \delta t_f(\tau_1), \quad \delta t_f(\tau_1) = \gamma_<(\tau_1) + \theta_<(\tau_1) - \tau_1,
$$

$$
\Delta t_b(\tau_1, \tau_2) = \Delta t_b(\tau_2) + \delta t_b(\tau_1), \quad \delta t_b(\tau_1) = \gamma_>(\tau_1) - \theta_>(\tau_1) - \tau_1.
$$

(29)

Consider the parameter reversal $\tau_1 \rightarrow \tilde{\tau}_1 \equiv \pi - \tau_1$, and note that $\tau_1 \in [0,\pi/2] \Rightarrow \tilde{\tau}_1 \in [\pi/2, \pi]$. It follows from the definitions (11) of $\gamma_<$ and $\theta_<$ that $\delta t_f(\tau_1) = \Delta t_f(\tilde{\tau}_1)$ and $\delta t_b(\tau_1) = \Delta t_b(\tilde{\tau}_1)$:

$$
\Delta t_f(\tilde{\tau}_1) = \pi - \tau_1 + \gamma_<(\pi - \tau_1) - \theta_<(\pi - \tau_1)
= -\tau_1 + \gamma(\tau_1) + \theta(\tau_1) = \delta t_f(\tau_1),
$$

$$
\Delta t_b(\tilde{\tau}_1) = \pi - \tau_1 + \gamma_>(\pi - \tau_1) + \theta_>(\pi - \tau_1) - 2\pi
= -\tau_1 + \gamma(\tau_1) - \theta(\tau_1) = \delta t_b(\tau_1).
$$

(30)

So by the previous proposition, the $\delta t$ contributions only strengthen causality preservation:

$$
\Delta t_f(\tau_1, \tau_2) = \Delta t_f(\tau_2) + \Delta t_f(\tilde{\tau}_1) \geq \Delta t_f(\tau_2) \geq 0,
$$

$$
\Delta t_b(\tau_1, \tau_2) = \Delta t_b(\tau_2) + \Delta t_b(\tilde{\tau}_1) \leq \Delta t_b(\tau_2) \leq 0.
$$

(31)

---

**Figure 7.** Time delay for minimally encoded boundary-to-bulk motion.
Figure 8. The critical encoding and the geographic \((\psi, \phi)\) coordinate system it begets, shown by the dotted curves and described in subsection 4.3. The dashed segment, along \(\Gamma\), tracks the gems \(p_\alpha(\tau)\). We denote \(\alpha(\tau_1) \equiv \alpha_1^*\) and \(\alpha(\tau_2) \equiv \alpha_2^*\) for shorthand.

Thus by moving into the bulk \((\tau_1 > 0)\), we gain additional time delay relative to the case where we start on the boundary \((\tau_1 = 0)\), and so causality gets easier to preserve.

What happens if the encoding does not stay minimal? Even if we allow \(\alpha(\tau)\) to wiggle around while the bulk point moves, the time delay \(\Delta t^\alpha\) depends only on the initial and final encoding angles, \(\alpha_1 \equiv \alpha(\tau_1)\) and \(\alpha_2 \equiv \alpha(\tau_2)\). If \(\alpha_2 \leq \alpha_1\), then the arclength of either the forward or the backward signal must get shorter as the other one becomes longer. The boundary trip is now more efficient, and causality becomes yet easier to preserve: \(\Delta t^\alpha \geq \Delta t^0 \geq 0\). Therefore without loss of generality we may take \(\alpha_1 \leq \alpha_2\). But even under this assumption, a direct analysis of \(\Delta t^\alpha(\tau_1, \tau_2)\) quickly becomes very technical.

4.2. Causality II: critical encoding

Consider the encoding symmetric about the null geodesic \(\Gamma\) with \(\ell = 0\), i.e. the line between \(N = p(0)\) and \(S = p(\pi)\). As shown in figure 8, we encode every bulk point \(p \in D^2\) on the boundary interval \(6\) subtended by the geodesic through \(p\) whose gem lies on \(\Gamma\). The symmetry of the encoding about \(\Gamma\) ensures that both the forward and backward boundary signals cover the same arclength, so they arrive simultaneously: \(t_f = t_b\).

We claim that this encoding, whose angle to the minimal encoding we call \(\alpha_*,(\tau)\), is always critical: \(\Delta t^{\alpha_*(\tau_1, \tau_2)} \equiv 0\). To see this, recall from subsection 3.2 that the time delay \(\Delta t^\alpha\) of any path \(p(\tau)\) in the bulk, encoded at angle \(\alpha(\tau)\), is equal to the time delay \(\Delta t^0\) of the minimally encoded path of the mimic point \(p_\alpha\), which follows the gems of the geodesics encoding \(p\). In our case, the gems \(p_\alpha\) lie on \(\Gamma\) and describe radial infall. We studied their the minimal encoding in subsection 2.3 and found that \(\Delta t^0\) vanishes identically for \(p_\alpha\); we will conclude that \(\Delta t^{\alpha_*} \equiv 0\) for \(p\) as well. It is deviations from \(\alpha_*\) to which we turn next.

\(^6\)Recall that we are working in a fixed time-slice. We will abuse language, saying (e.g.) that bulk points in \(D^2\) are encoded on boundary intervals, that (null and spacelike) geodesics are curves in \(D^2\), and so on.
To that end, we introduce the angle $\beta(\tau)$ by which the encoding fails to be critical. As with $\alpha(\tau)$, the time delay, now labeled $\Delta t'(\tau_1, \tau_2)$, depends only on $\beta_1 \equiv \beta(\tau_1)$ and $\beta_2 \equiv \beta(\tau_2)$, as shown in figure 9. If $\beta_1 = 0$, then $\beta_2 \neq 0$ will lengthen one of the boundary travel times and shorten the other, relative to the critical travel time $t_f = t_b = t_p$. The same is true if $\beta_2 = 0$ and $\beta_1 \neq 0$. Thus if either endpoint of the bulk point’s path is critically encoded, causality is preserved:

$$\Delta t(\beta_1 = 0 \neq \beta_2) \geq \Delta t(\beta_1 = 0 = \beta_2) = 0,$$

$$\Delta t(\beta_1 \neq 0 = \beta_2) \geq \Delta t(\beta_1 = 0 = \beta_2) = 0.$$ (32)

It remains to consider the general case, $\beta_1 \neq 0 \neq \beta_2$. Before we do so, a few comments are in order. First, we can assume without loss of generality that $\beta_1$ and $\beta_2$ have the same sign; otherwise, they work together to increase $\Delta t$ even further above its value when either $\beta_1$ or $\beta_2$ vanishes. Moreover we can take both to be positive, since the symmetry of the critical encoding guarantees that the time delay for an encoding with $\beta_1, \beta_2 < 0$ is equal to the time delay for the encoding with $-\beta_1$ and $-\beta_2$. Second, because a generic encoding geodesic intersects the critical one away from $\Gamma$, we no longer have $\beta_1 \leq \beta_2$ in the same way we had $\alpha_1 \leq \alpha_2$ above. Nevertheless, by definition $\alpha(\tau) = \alpha_*(\tau) + \beta(\tau) \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. This condition affords us the weaker constraint $\alpha_1 + \beta_1 \leq \alpha_2 + \beta_2$.

To summarize, $\alpha$ measures the non-minimality of an encoding, $\beta$ measures its non-criticality, and $\alpha_*$ measures how non-minimal the critical encoding is. This clarifies why radial infall along $\Gamma$ is special: it is both minimal and critical, $\alpha(\tau) = \beta(\tau) = 0$. As we proceed to the general case, it is the $\beta$-perspective that will become ‘critical’ to us.

4.3. Causality III: the general case

New coordinates. We still lack a convenient coordinate system in which to describe deviations $\beta(\tau)$ from criticality. Such a coordinate system will consist of the null and (critical) spacelike

---

More precisely, $\beta(\tau)$ is the signed angle pointing from $\Gamma$ to the gem of the geodesic that encodes $p(\tau)$. We say that $\beta > 0$ if it points counterclockwise. This convention also applies to the angles $\alpha_*(\tau)$ in figure 8 and helps to clarify the result $\alpha_* \sim -\theta$, below.
geodesics, as shown in figure 8. Every bulk point is labeled uniquely by its null geodesic (how far ‘up’ it is) and by its critical encoding (how far ‘in’ it is). This coordinate ruling is orthogonal, and viewed from the side, it resembles the meridians and parallels on a globe. Indeed, we now pass from the faraway imagination of AdS$_3$ to the close reality of our Earth. The idea is to take advantage of the fact that the angular momentum $\ell$ is constant along each null geodesic, while the length $\gamma_{\alpha}$ is constant along each critical spacelike geodesic. For $\tau \in [0, \frac{\pi}{2}]$, null geodesics are described by (10). We may rewrite the equation for $\theta(\tau)$ as $\tan \frac{\theta(\tau)}{2} = \frac{L}{\ell} \tan \theta(\tau)$ and substitute it into the equation for $\gamma(\tau)$; this yields

$$\tan \gamma(\tau) = \frac{L}{\ell} \sin \theta(\tau) \Rightarrow \tan \frac{\gamma(\tau)}{2} = \frac{L}{\ell} \equiv \tan \phi = \text{constant}. \quad (33)$$

Meanwhile, the critical spacelike geodesics are characterized by

$$\cos \gamma_{\alpha}(\tau) = \cos \alpha_{\alpha}(\tau) \cos \gamma(\tau) = \cos \theta \cos \gamma \equiv \cos \psi = \text{constant}, \quad (34)$$

where we have used (17) and the fact that $\alpha_{\alpha} = -\theta$. We have introduced the constants $\phi = \tan^{-1}(L/\ell) = \gamma(\frac{\pi}{2})$ and $\psi = \gamma_{\alpha}$; the former measures how close a null geodesic comes to the boundary, and the latter measures the depth of a critical spacelike geodesic’s gem. One can check using (11) that the results above hold for $\tau \in [\frac{\pi}{2}, \pi]$ as well as for $\tau \in [0, \frac{\pi}{2}]$.

To complete the coordinate change, we solve (33) and (34) for $\gamma$ and $\theta$ in terms of $\psi$ and $\phi$ and rewrite the AdS$_3$ metric (6) in the new coordinates $(t, \psi, \phi)$:

$$ds^2 = \frac{1}{\sin^2 \phi \sin^2 \psi} \left( -dt^2 + d\psi^2 + \sin^2 \psi \, d\phi^2 \right)$$

$$= \frac{1}{\sin^2 \phi \cos^2 \psi} \left( -dt^2 + d\psi^2 + \cos^2 \psi \, d\phi^2 \right). \quad (35)$$

In terms of $(\psi, \phi)$, the spatial part of the metric is conformal to a patch of the round sphere with an adapted north and south pole on the boundary of $D^2$. Meanwhile, shifting to $\bar{\psi} \equiv \psi - \frac{\pi}{2}$, we find that $ds^2$ takes the same form in coordinates $(t, \gamma, \theta)$ as in $(t, \bar{\psi}, \phi)$. In fact, $\bar{\psi}$ is another tortoise coordinate, perhaps more natural for holography than $\gamma$. In these coordinates, null $(d\phi = 0)$ and critical spacelike $(d\bar{\psi} = 0)$ geodesics play the roles of the radial and circular paths from section 2, respectively.$^8$

Since our analysis is insensitive to conformal rescaling, the metric (35) gives us license to work entirely on $S^2$ and affords a fresh interpretation of the time delay $\Delta t$. Null geodesics become meridians or lines on $S^2$, encoding geodesics become parallels or circles on $S^2$, and $t_p$ is the length of a meridian arc between the parallels encoding the initial and final bulk points. An encoding is critical ($t_p = t^\psi_{\beta=0} = t^\beta_{\tau=0}$) if all of the encoding parallels share a center—the north pole—that also lies on the meridian. Meanwhile, non-critical encodings are parallels with an arbitrary center whose location is related to the angle $\beta$ in figure 9.

**Proof of the general case.** To attack the case $\beta_1 \neq 0 \neq \beta_2$, we rotate the globe (and AdS$_3$) by angle $-\beta_1$ to ‘undo’ the non-criticality of the initial encoding. As shown in figure 10, we introduce new coordinates $(\psi', \phi')$ adapted to the initial encoding, which becomes the latitude circle $a d'$ of $p(\tau_1)$, and whose center becomes the new north pole $N'$. This encoding becomes critical for the family of null geodesics taking the form (10) and (11) in coordinates $(\bar{\psi'}, \phi')$.

$^8$ To prove that causality is preserved, it suffices to extend the analysis of radial infall to arbitrary encodings. Equivalently, from figure 9, we must show that the encoding intervals remove more arclength on one side of the boundary circle than they create on the other. But we will not pursue these ideas directly.
Figure 10. $S^2$ and $D^2$, rotated by angle $\beta$. The initial encoding, a parallel with center $N'$, is critical for a new null geodesic $q(\tau)$, a meridian through $N'$, with constant $\phi'$. The initial (a) and final (b) signal locations for paths $p(\circ)$ and $q(\times)$ are indicated on the boundary.

i.e. those with constant $\ell'$ (and hence constant $\phi'$). Among these is the unique null geodesic $q$ passing through $p(\tau_1)$. Let us consider a critically encoded bulk point that moves along $q$: it starts at $p(\tau_1)$, where it shares the initial encoding $aa'$ with the bulk point $p$, and it arrives at the point $q(\tau_0)$ where it orthogonally intersects—and is encoded by—the latitude circle $b_qb'_q$ of $p(\tau_2)$. Meanwhile, $p(\tau_2)$ itself is encoded by the generic circle $b_pb'_p$, which intersects the parallel $b_qb'_q$ exactly once\(^9\) at $p(\tau_2)$.

We show presently that the bulk travel time $t_q$ of $q$ exceeds the boundary travel time $\min\{t_f, t_b\}$ of $p$, but is dominated by the bulk travel time $t_p$ of $p$; this demonstrates that $\min\{t_f, t_b\} \leq t_q \leq t_p$ and completes the proof of causality preservation in the general case.

Meanwhile, the claim that $t_q \leq t_p$ can be translated into the situation shown in figure 11: two meridians $p$ and $q$ on (a hemisphere of) $S^2$ intersect at $p_1 = p(\tau_1)$, and a circle perpendicular to $q$ intersects both meridians at $p_2 = p(\tau_2)$ and $q_0 = q(\tau_0)$, respectively. We must prove that the arc $t_q = p_1q_0$ is shorter than the arc $t_p = p_1p_2$. Equivalently, viewing $S^2 \subset \mathbb{R}^3$ as a collection of unit vectors, the angle $\theta(p_1, q_0)$ made by the vectors $p_1$ and $q_0$ must be smaller than the angle $\theta(p_1, p_2)$.

\(^9\)Circles on the sphere actually intersect in two points, but AdS$_3$ is conformally equivalent to a hemisphere of $S^2$ where such intersections are unique. A clever way to see this is to perform an inversion about the center of $D^2$ that preserves the boundary circle. The inversion maps circles to circles, takes the inside of $D^2$ to the outside, and preserves the geodesic circles in $D^2$ because it preserves angles. The intersection point inside $D^2$ is then moved to a point outside $D^2$. 
We begin by fixing the spherical coordinates \((r, \theta, \phi)\) of our points and their ranges:

\[
p_1 = (1, \theta_p, 0), \quad p_2 = (1, \theta_q, \phi), \quad q_0 = (1, \theta_q, 0);
\]

\[
\theta_p, \theta_q \in [0, \pi], \quad \phi \in \left[\frac{-\pi}{2}, \frac{\pi}{2}\right].
\]  

(36)

The angles \(\theta(p_1, q_0)\) and \(\theta(p_1, p_2)\) are computed from their dot products, which are found using the Jacobian of the transformation between Cartesian and spherical coordinates:

\[
(r_1, \theta_1, \phi_1) \cdot (r_2, \theta_2, \phi_2) = r_1 r_2 \sin \theta_1 \sin \theta_2 \cos (\phi_2 - \phi_1) + \cos \theta_1 \cos \theta_2.
\]  

(37)

We compute the difference in the arclengths \(t_q = \theta(p_1, q_0)\) and \(t_p = \theta(p_1, p_2)\) explicitly:

\[
\theta_{pq}(\phi) \equiv t_p - t_q = \theta(p_1, p_2) - \theta(p_1, q_0) = \cos^{-1}(p_1 \cdot q_0) - \cos^{-1}(p_1 \cdot p_2)
\]

\[
= \cos^{-1}\left(\sin \theta_p \sin \theta_q \cos \phi + \cos \theta_p \cos \theta_q \cos \phi\right) - (\theta_p - \theta_q).
\]  

(38)

It is straightforward to show that \(\theta_{pq}(\phi)\) is increasing for \(\phi \geq 0\):

\[
\frac{\partial \theta_{pq}}{\partial \phi} = \frac{\sin \theta_p \sin \theta_q \sin \phi}{\sqrt{1 - (p_1 \cdot p_2)^2}} \geq 0.
\]  

(39)

To wit, \(1 - (p_1 \cdot p_2)^2 \geq 0\) since \(p_1\) and \(p_2\) are unit vectors. Also, \(\sin \theta_p\) and \(\sin \theta_q\) are nonnegative whenever \(\theta_p\) and \(\theta_q\) lie between 0 and \(\pi\). And finally, \(\sin \phi \geq 0\) for \(\phi \in [0, \frac{\pi}{2}]\).

But now we are done: observe that \(\theta_{pq}\) is an even function of \(\phi\) (due to the symmetry of the latitude circle), vanishes at \(\phi = 0\) (where the meridians \(p\) and \(q\) become identical), and increases for \(\phi \geq 0\). By symmetry, \(\theta_{pq}\) is always nonnegative and attains its minimum value at \(\phi = 0\). This shows that the orthogonal meridian \(q\) provides the shortest path from \(p_1\) to the parallel at latitude \(\theta_q\), and that moving further out on the latitude circle (changing \(\phi\), or equivalently varying \(\beta_2 = \beta_2 - \beta_1\)) can only give longer bulk travel times.
**Summary and a generalization.** Let us recapitulate. By adapting the \((\psi, \phi)\) coordinate system to either the initial or final encoding of the bulk point, we discovered coordinates where, fixing one encoding interval, there is a unique encoding of the other point for which all boundary and bulk times of flight are the same. (We fixed the initial encoding purely for convenience.) These coordinates have a north pole at the midpoint of the initial encoding interval. When we toggle the angle of the final interval about the point of arrival, one boundary time becomes shorter and the other gets longer. Unless the bulk point starts and arrives on the same meridian, its time of flight exceeds the travel time on the boundary.

This argument also proves causality in AdS\(_d\), so long as bulk points are encoded in spherical regions of the boundary. Latitude circles are replaced by sphere domes in the Poincaré ball \(D^{d-1}\), generated by the geodesics of \(D^{d-1}\) tangent to any point of the surface in the interior. The time delays for starting and landing on these domes is uniform in these coordinates: the meridians are infalling coordinates relative to the north pole, just as in AdS\(_3\). The domes for different sphere encodings of the arrival point are determined by the plane of tangency at the arrival point. Toggling the plane of arrival against the critical plane yields two transverse hyperplanes that share a common geodesic as long as the dimension \(d - 1\) of the ball is at least 3. (Such geodesics degenerate to a point in \(D^2\).) More spherical angles may be toggled, but the result is the same: part of the dome at the boundary necessarily moves closer to the original encoding along at least one geodesic.

5. Discussion and conclusions

In this paper, we introduced an apparent conflict between the bulk and boundary perspectives on causality: a bulk point traveling through empty AdS\(_3\) seems to drag its entanglement wedge along with it faster than light can move on the boundary. We demonstrated in section 2 how this paradox is resolved when the bulk point travels along circular and radial trajectories. Along the way, we introduced a convenient coordinate \(\gamma\) that measures the size of the point’s encoding interval, its location in the bulk, and the nonlocality of its encoding on the boundary. It also turns out to be a tortoise coordinate on AdS\(_3\), and its construction is perhaps our main technical innovation. We went on in section 3 to parametrize null geodesics in coordinates \((\gamma, \theta)\) and used these coordinates to discuss the minimality of the bulk encoding. These steps prepared us to prove in section 4 that for null geodesics and hence for all causal bulk paths, the bulk point’s travel time always matches or exceeds the travel time of a signal moving along the boundary at the speed of light. The central idea was that every null geodesic has a critical encoding where the bulk point arrives simultaneously with both the forward and backward boundary signals. By re-coordinatizing AdS\(_3\) using these encodings, we converted the study of the time delays of non-critical encodings into simple statements in spherical geometry and completed the proof that causality is preserved.

For any two points on a spatial slice of AdS orthogonal to a timelike Killing field, a unique fastest null geodesic connects them. When the entanglement wedges of the two points are adapted to this null geodesic, causality is critical: the time delay between the bulk point and the boundary entanglement wedges vanishes. This means that in AdS, bulk causality is on the verge of violating boundary causality. If causality on the boundary is not violated, then perturbations of the AdS geometry must lead to additional time delay in the bulk. This idea suggests that an extension of the Gao–Wald theorem into the interior of AdS may follow from holography. Namely, causality in the bulk should be encoded in the bounds imposed on boundary causality by the entanglement wedge.

Given this intuition, one can argue that to ensure this time delay, the metric perturbations must satisfy an energy condition. Such an inequality should take the form of an integral over
the null geodesic to leading nontrivial order in perturbation theory, similar to the formulation of the averaged null energy condition (ANEC). That causality is critical suggests that this is a second-order effect: a linearized perturbation of the metric due to gravity can in principle have either sign, and so should the linearized effect. The vanishing of this linear effect would then constrain the equations of motion for the metric in the bulk. It should be noted that a connection between the ANEC and causality has been proven in quantum field theory \[23, 24\]. Here we do not work with complete null geodesics, but we do have entanglement wedges and Ryu–Takayanagi surfaces playing a role. Our result suggests that there might be a refined form of the ANEC for finite pieces of null geodesics which includes information from the entanglement wedges that connect their endpoints to a boundary. In AdS, these entanglement wedge surfaces are orthogonal to the null geodesic at the points of departure and arrival. The ingredients involved suggest that this is an application of the quantum null energy condition to the AdS perturbations \[25\], which connects it more directly to the ANEC. It would be very interesting to understand this possibility better.

Our results are strongly dependent on the details of the spacelike geodesics in (1), which determine the entanglement wedges. It is interesting to generalize to black hole spacetimes, where already there are timelike geodesics describing particles whose angular frequencies exceed unity (see, for example, \[14\]). Therefore at least for circular motion, black holes seem to make the causality problem more severe. The resolution must be that the entanglement wedges need to be ‘larger’, in some sense, than in pure AdS. There are also additional complications, like the presence of causal shadows and the fact that the geometry of the entanglement wedges may not be easily calculable. Our analysis of the AdS geometry, which assumes that null geodesics start and end on the boundary, will generically fail: a black hole can cause some or all of these geodesics to fall into the singularity. Because our explicit tools cannot address these issues, we consider such analysis beyond the scope of the present paper and leave it to future work.

Acknowledgments

DB would like to thank J Simon for discussions. The work of DB is supported in part by the Department of Energy under Grant DE-SC 0011702.

Data availability statement

No new data were created or analysed in this study.

Appendix A. Null geodesics in AdS

To look for null geodesics, let us start by imposing the null-worldline condition:

\[
d_s^2_{\text{AdS}} = g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu = 0 \implies -\dot{t}^2 + \dot{\gamma}^2 + \cos^2 \gamma \dot{\theta}^2 = 0. \tag{A1}
\]

Here dots are derivatives with respect to an affine parameter \( \tau \). We seek Killing vector fields (KVFs) that give quantities conserved along the geodesics, but since ‘light is conformal’, conformal KVFs (cKVFs) also yield conserved quantities. The metric is independent of \( t \) and \( \theta \), so \( \xi = \partial_t = (1, 0, 0) \) and \( \eta = \partial_\theta = (0, 0, 1) \) are cKVFs for AdS. These give rise to a conserved energy \( E \in \mathbb{R}_+ \) and angular momentum \( \ell \in [-1, 1] \):

\[
-g_{\mu\nu} \xi^\mu \dot{x}^\nu = \dot{t} \equiv E, \quad g_{\mu\nu} \eta^\mu \dot{x}^\nu = \cos^2 \gamma \dot{\theta} \equiv \ell \implies \dot{\theta} = \ell \frac{\dot{\theta}}{\cos^2 \gamma}. \tag{A2}
\]
This immediately solves the geodesic equation for $t$ as $t(\tau) = E\tau + t_0$ and justifies the name $\tau$ for the affine parameter. Substituting these conserved quantities into the null condition above, we convert it into an effective Newtonian problem:

$$-E^2 + \dot{\gamma}^2 + \ell^2 = 0 \implies \frac{1}{2} \frac{d}{d\tau} \gamma^2 + \frac{\ell^2}{2 \cos^2 \gamma} = \frac{1}{2} E^2.$$ (A3)

This ODE is separable, and can be integrated exactly and inverted to find $\gamma(\tau)$:

$$\frac{d\gamma}{d\tau} = \sqrt{E^2 - \frac{\ell^2}{\cos^2 \gamma}} \implies \gamma(\tau) = \tan^{-1} \left( \pm \sqrt{\frac{E^2 - \ell^2}{E^2 + \ell^2 \tan^2(E(c \pm \tau))}} \right).$$ (A4)

We now return to the conservation law for $\ell$ and view it as an ODE for $\theta(\tau)$. Substituting the solution $\gamma(\tau)$, we integrate the differential equation to find

$$\theta(\tau) = \ell \int \frac{d\tau}{\cos^2(\gamma(\tau))} + \theta_0 = \tan^{-1} \left( \frac{\ell}{E} \tan(E(c \pm \tau)) \right) + \theta_0.$$ (A5)

Here $t_0, c \in \mathbb{R}$ and $\theta_0 \in S^1$ are integration constants, while $E \in \mathbb{R}_+$ and $\ell \in [-1, 1]$ are conserved quantities. Without loss of generality, we may take $t_0 = \theta_0 = 0$ to fix the zero-points of $t(\tau)$ and $\theta(\tau)$. We may also set $E = 1$, since $\gamma(\tau)$ and $\theta(\tau)$ depend only on the ratio $\ell/E$. Finally, the constant $c$ parametrizes how far from the boundary $p$ is at $\tau = 0$. We set $c = 0$ to fix $p \in \partial D^2$ at $\tau = 0$; if $p$ starts from within the bulk, we will ‘start its clock’ at $\tau = \tau_1 > 0$. These simplifications immediately yield (10) above and leave $\ell$ as the sole parameter labeling the geodesics. As noted above, we can extend (10) to $\tau \geq \frac{\pi}{2}$ by reversing the direction of $\tau$ and reflecting $\theta$ across $\theta = \frac{\pi}{2}$; this gives (11) directly.

**ORCID iDs**

David Berenstein © [https://orcid.org/0000-0002-9405-7678](https://orcid.org/0000-0002-9405-7678)

**References**

[1] Rees M 1966 Appearance of relativistically expanding radio sources *Nature* **211** 468–70
[2] Zensus J A and Pearson T J 1988 ‘Superluminal Radio Sources’, Symposium - International Astronomical Union (Cambridge: Cambridge University Press)
[3] Maldacena J M 1999 The large N limit of superconformal field theories and supergravity *Int. J. Theor. Phys.* **38** 1113–33
[4] Witten E 1998 Anti-de Sitter space and holography *Adv. Theor. Math. Phys.* **2** 253–91
[5] Gubser S S, Klebanov I R and Polyakov A M 1998 Gauge theory correlators from noncritical string theory *Phys. Lett. B* **428** 105–14
[6] Hamilton A, Kabat D N, Lifschytz G and Lowe D A 2006 Holographic representation of local bulk operators *Phys. Rev. D* **74** 066009
[7] Banks T, Douglas M R, Horowitz G T and Martinec E J 1998 AdS dynamics from conformal field theory (arXiv:hep-th/9808016 [hep-th])
[8] Ryu S and Takayanagi T 2006 Holographic derivation of entanglement entropy from AdS/CFT *Phys. Rev. Lett.* **96** 181602
[9] Ryu S and Takayanagi T 2006 Aspects of holographic entanglement entropy *J. High Energy Phys. JHEP08(2006)045*
[10] Hubeny V E, Rangamani M and Takayanagi T 2007 A Covariant holographic entanglement entropy proposal J. High Energy Phys. JHEP07(2007)062
[11] Dong X, Harlow D and Wall A C 2016 Reconstruction of bulk operators within the entanglement wedge in Gauge-gravity duality Phys. Rev. Lett. 117 021601
[12] Almheiri A, Dong X and Harlow D 2015 Bulk locality and quantum error correction in AdS/CFT J. High Energy Phys. JHEP04(2015)163
[13] Engelhardt N and Fischetti S 2017 Locality from quantum gravity: all or nothing Int. J. Mod. Phys. D 26 1743028
[14] Berenstein D, Li Z and Simon J 2020 ISCOs in AdS/CFT (arXiv:2009.04500 [hep-th])
[15] Gao S and Wald R M 2000 Theorems on gravitational time delay and related issues Class. Quantum Grav. 17 4999–5008
[16] Witten E 2020 Light rays, singularities, and all that Rev. Mod. Phys. 92 045004
[17] Headrick M, Hubeny V E, Lawrence A and Rangamani M 2014 Causality & holographic entanglement entropy J. High Energy Phys. JHEP12(2014)162
[18] Aminneborg S, Bengtsson I, Brill D, Holst S and Peldan P 1998 Class. Quantum Grav. 15 627–44
[19] Bengtsson I, Holst S and Jakobsson E 2014 Classics illustrated: limits of spacetimes Class. Quantum Grav. 31 205008
[20] Wald R M 1984 General Relativity (Chicago: University of Chicago Press)
[21] Berenstein D E, Corrado R, Fischler W and Maldacena J M 1999 The operator product expansion for Wilson loops and surfaces in the large N limit Phys. Rev. D 59 105023
[22] Penrose R 1963 Asymptotic properties of fields and space-times Phys. Rev. Lett. 10 66–8
[23] Kelly W R and Wall A C 2014 Holographic proof of the averaged null energy condition Phys. Rev. D 90 106003
[24] Kelly W R and Wall A C 2015 Phys. Rev. D 91 069902 (Erratum)
[25] Hartman T, Kundu S and Tajdini A 2017 Averaged null energy condition from causality J. High Energy Phys. JHEP07(2017)066
[26] Bousso R, Fisher Z, Koeller J, Leichenauer S and Wall A C 2016 Proof of the quantum null energy condition Phys. Rev. D 93 024017