A self-consistency check for unitary propagation of Hawking quanta

Daniel Baker,[1,2,*] Darsh Kodwani,[1,2,†] Ue-Li Pen,[1,3,4,5,‡] and I-Sheng Yang[1,5,§]

[1] *Canadian Institute of Theoretical Astrophysics, 60 St George St, Toronto, ON M5S 3H8, Canada.*
[2] *University of Toronto, Department of Physics, 60 St George St, Toronto, ON M5S 3H8, Canada.*
[3] *Canadian Institute for Advanced Research, CIFAR program in Gravitation and Cosmology.*
[4] *Dunlap Institute for Astronomy & Astrophysics, University of Toronto, AB 120-50 St. George Street, Toronto, ON M5S 3H4, Canada.*
[5] *Perimeter Institute of Theoretical Physics, 31 Caroline Street North, Waterloo, ON N2L 2Y5, Canada.*

The black hole information paradox presumes that quantum field theory in curved spacetime can provide unitary propagations from near-horizon modes to asymptotic Hawking quanta. Instead of invoking conjectural quantum-gravity effects to modify such an assumption, we propose a self-consistency check. We establish an analogy to Feynman's analysis of a double-slit experiment. Feynman showed that a unitary propagation of the interfering particles, namely ignoring the entanglement with the double-slit, becomes an arbitrarily reliable assumption when the screen to project the interference pattern goes to infinitely far away. We argue for an analogous self-consistency check for quantum field theory in curved spacetime. We apply it to the propagation of Hawking quanta, testing whether ignoring the entanglement with the geometry also becomes arbitrarily reliable in the limit of a large black hole. We present curious results to suggest a negative answer, and we discuss how this loss of naïve unitarity in QFT is related to the soft-hair-memory effect.

* dbaker@cita.utoronto.ca
† dkodwani@physics.utoronto.ca
‡ pen@cita.utoronto.ca
§ isheng.yang@gmail.com

I. INTRODUCTION AND SUMMARY

A. Black Hole Information Paradox

The black hole information paradox [1] was sharpened to show a self-inconsistency among the following three widely-believed statements [2][1]:

- **Unitary evaporation:** A black hole will totally evaporate; the formation and evaporation of a black hole is described by an asymptotic observer by a unitary S-matrix, whose size is the exponential of the black hole's Bekenstein entropy.
- **General relativity:** The collapse-Schwarzschid geometry given by general relativity is a valid description of the spacetime everywhere except for near the singularity. Therefore, there is no drama while crossing the horizon.
- **Local quantum field theory:** Away from the singularity, we can apply local quantum field theory (QFT) which describes microscopic steps of evaporation. For large black holes, it manifests itself as unitary propagation of every near-horizon mode into every asymptotic Hawking quanta.

Basically, Statement Three demands that a near horizon mode is related to an asymptotic Hawking quanta by a unitary transformation, therefore carrying the same qubit of information. However, Statement One and Two demand this unique qubit to be maximally entangled with two distinct objects, violating monogamy. This conflict has inspired many different proposals to modify one of the above three statements. For example, information loss or remnant [3] modifies Statement One; firewall [2, 4] or ER=EPR [5] modifies Statement Two; various proposals of nonlocal effects near the horizon [6–8], causal patch complementarity [9–11], or computational complementarity [**?**] modify Statement Three.[2]

In this paper, we will try to provide a stronger motivation to modify Statement Three. In fact, we will argue that it is already problematic on its own. In other words, **the application of local QFT to unitary propagation of Hawking quanta is not self-consistent.**

Before presenting our argument, we will first explain its relation to some existing ideas. The recent paper by Osuga and Page [8] provides an abstract framework of how modifying Statement Three resolves the paradox. Basically, the near horizon mode is thermal as demanded by Statement Two. But it propagates out in a non-unitary process, gaining the appropriate information to restore a unitary S-matrix as demanded by Statement One. For such abstract model to work in practice, this non-unitary propagation should be understood as interaction with a hidden system. Here hidden simply means something that a naïve application of local QFT is ignorant to. The obvious candidate of such hidden system is the geometry.

One proposal along this line of thoughts was discussed in [16]. The idea was that the black hole acquires a large uncertainty during the evaporation process. If we treat the superposition of different classical geometries as a quantum system, it may qualify as the hidden system that is responsible for the non-unitary propagation of Hawking quanta. Unfortunately, the early version of such proposal can be defeated by counting entropy. In order for such non-unitary process to restore the asymptotic S-matrix, the hidden system must store the required information, which is comparable to the amount of microscopic information carried by the black hole. However, the black-hole-no-hair theorem states that two black holes with identical macroscopic parameters (mass, momentum, angular momentum, charge) are isometric to each other. This limits the number of different classical geometries to be parametrized by macroscopic parameters only. As a result, the Hilbert space of superposed classical geometries has far fewer dimensions than what is necessary to carry the required amount of information.[3] Thus, even if a propagating Hawking quantum does interact with this hidden system, it cannot reproduce a unitary S-matrix; therefore such modification of Statement Three alone is insufficient to resolve the paradox.

The recent development on black-hole-soft-hairs [12] hinted a way to revive this idea. Basically, it was realized that two isometric (regions of) geometries should not always be considered as the same state. If one maps the geometries with congruences of geodesics, two isometric regions with different maps are actually different states. That is because changes in the mapping geodesics are measurable memory effects, and a consistent physical theory must be able to distinguish them. This realization may dramatically increase the number of different states encoded in classical geometries. A tentative counting of state on a black hole horizon yields exactly the required number such that a modification to Statement Three can indeed resolve the paradox [12].

---

[1] Different physicists may believe in some of those more strongly than the others, but very few would have argued against any single statement without the explicit conflict in the information paradox.

[2] Various versions of complementarity basically claim that QFT breaks down if one applies it compare quantities which are in-principle not measurable (or difficult to measure) by the same observer.

[3] We thank Raphael Bousso for private communication.

In this paper, we will take another step that is complementary to the above progress. It is exciting to know that modifying Statement Three can in-principle resolve the paradox. However, it remains a great mystery of why should local QFT break down in low curvature regions. If we just demand that it breaks down in order to resolve the information paradox, then there is no reason why this solution is preferred over modifying Statement One or Two. We would like to argue that there is a self-consistency check that we should apply to QFT in curved spacetime, for a reason that is totally independent from the information paradox. Applying such self-consistency check, we can show that a unitary propagation from a near horizon mode to an asymptotic Hawking quantum is already self-inconsistent. Thus it is only natural to modify Statement Three.

B. Quantum Field Theory in Curved Spacetime Requires a Self-Consistency Check

The starting point of our argument involves no new physics at all. We simply recognize that local QFT implies a standard assumption of **subsystem unitarity**. Quantum mechanical evolution of a full system is by-definition unitary. However in practice, we are not able to describe everything by quantum mechanics all together, so we describe only subsystems. Namely, we assume that the full system can be separated into a "classical background" and a "quantum subsystem". After such artificial separation, since we only use quantum mechanics to describe the quantum subsystem, it always appears to be unitary. However, such unitarity should not be taken as a physical fact unless one specifically checks the nature of interactions with the background.

The most direct way to justify the classical-background assumption is to also describe the background in quantum mechanics, and verify that the quantum interaction between the background and the quantum subsystem indeed does not entangle them. This is clearly difficult in practice. The very reason why we would like to apply the classical-background approximation is to avoid describing the far-too-complicated background by quantum mechanics. A consistency check which requires us to do so defeats the purpose. In particular, there are situations in which we in-principle do not know how to describe the classical background in quantum mechanics. Our problem at hand, QFT in curved spacetime, is exactly this annoying situation. Since the classical background is the geometry, one in-principle needs to know quantum gravity in order to model the quantum interaction with the background.

Fortunately, Feynman, in his famous lectures, presented an interesting trick to circumvent this obstacle. This trick, to a certain extent, enables us to perform a **self-consistency check** of the classical-background assumption **without** knowing the quantum nature of the background. In Sec.II, we will review Feynman's analysis of the double-slit experiment. In his case, the double-slit is the classical background, the particle passing through it is the quantum subsystem, and subsystem unitarity is checked by whether an interference pattern can be observed on the final projection screen. He showed that we can ignore the quantum details of the double-slit and summarize that as an uncertainty in its position, which is a classical quantity. When such uncertainty is fixed, a classical calculation can show that the interference pattern is always visible as long as we put the final projection screen to be infinitely far away. It is in this sense that a unitary evolution of the subsystem is an arbitrarily good approximation.

The fundamental principle behind Feynman's trick is that "classical wave coherence" should be taken as the pre-requisite of "quantum unitarity". In Sec.III, we argue that this principle enables a general self-consistency check of the classical background assumption. We provide a natural generalization to local QFT in curved spacetime, and we discuss why such self-consistency check is different from perturbative calculations of gravitons.

In Sec.IV, we apply the self-consistency check to the propagation of Hawking quanta in a close analogy to Feynman's analysis. The classical geometry plays the role of the double-slit, a background that can potentially entangle with the Hawking quanta wavefunction and ruin its subsystem unitarity. We then assume that the unknown quantum nature of geometry can be parametrize by a classical uncertainty with an unknown but fixed, gauge-invariant size. A classical wave plays the role of the interference pattern. When even a classical wave decoheres due to the uncertainty of geometry, there is no reason to believed that the quantized version of such wave has unitarity. The common expectation is that in the limit of a large black hole, gravitational effects outside the horizon are arbitrarily weak, thus local QFT should be arbitrarily trustworthy. By analogy to Feynman's analysis, we should expect that with a fixed uncertainty in the geometry, any effect on the classical wave should drop to zero when we take the large black hole limit. Interestingly, what we found seems to be the opposite. We show that with a fixed geometric uncertainty, a classical wave suffers an *arbitrarily large* correction in the large black hole limit. By analogy to Feynman's analysis, this is strong evidence that the apparent unitarity of QFT in this case is not a valid assumption.

In Sec.V, we discuss various implications of our finding. First of all, our self-consistency check should be generally applicable to any geometry, and it does not invalidate local QFT entirely. Even when unitarity is lost in local QFT, the values of many observables do not have to change. Most existing applications of QFT in curved spacetime actually only care about the expectation value of particle numbers, which can be unaffected by the loss of subsystem unitarity. Secondly, our finding resolves the information paradox by selecting one culprit among the three statements, but it does not yet resolve the information transfer puzzle. We argue that our finding naturally implies an information

exchange between a QFT quantum and the local geometry [8]. The reason why local geometry carries the appropriate information to give in the first place may be related to the soft-hair proposal [12].

II. FEYMAN'S ANALYSIS ON DOUBLE-SLIT EXPERIMENTS

Double-slit interference is the signature experiment to demonstrate the nature of quantum mechanics. Figure 1 shows a simple example of such experiment. When a group of particles of the same momentum pass through slit 1 only, they reach the final screen as some probability distribution $P_1(y)$. When they pass through slit 2 only, they reach the final screen as a probability distribution $P_2(y)$. When both slits are open, the probability distribution we find on the final screen is not a simple sum; $P_{12}(y) \neq P_1(y) + P_2(y)$, instead, the final screen shows an interference pattern which can only be explained by wavefunctions. $P_i = \langle\phi_i|\phi_i\rangle$ and $P_{12} = (\langle\phi_1| + \langle\phi_2|)(|\phi_1\rangle + |\phi_2\rangle) = P_1 + P_2 + 2\mathrm{Re}\langle\phi_1|\phi_2\rangle$.

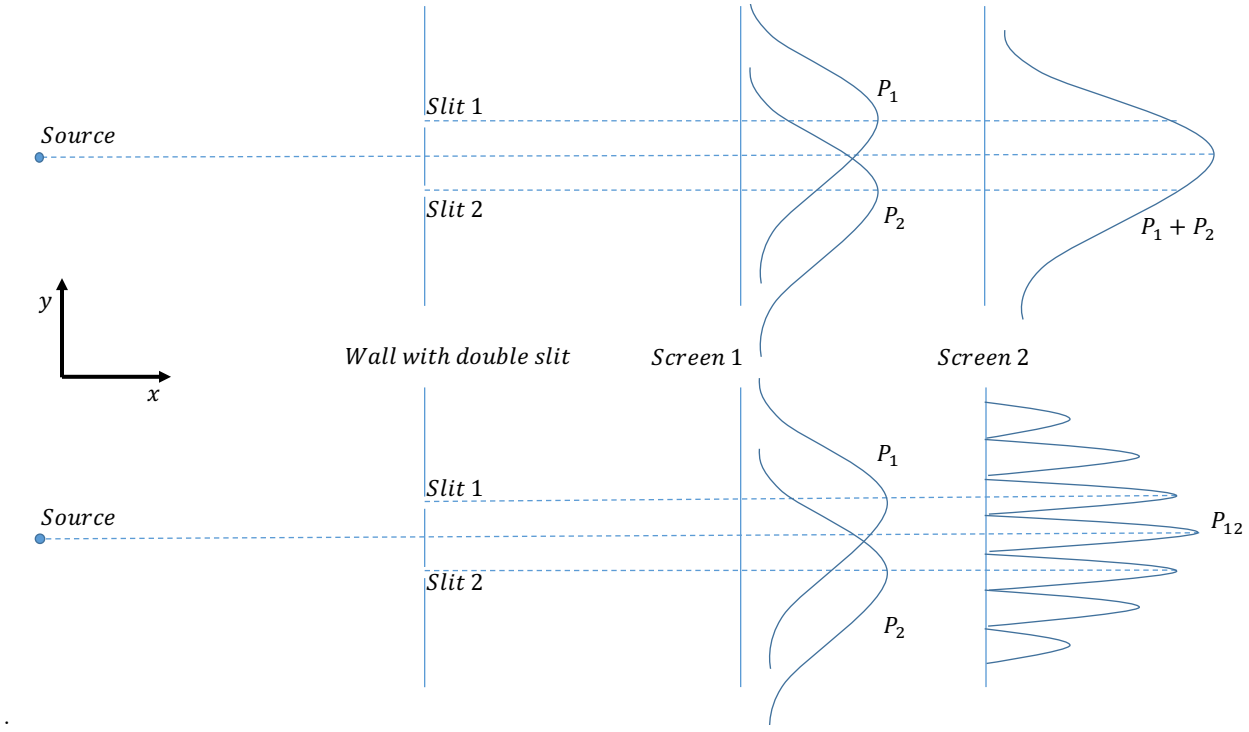


FIG. 1. Schematic of the double slit experiment. The diagram shows two different scenarios. The top figure shows the behaviour one would expect of classical particles going through one slit or the other. The bottom figure shows the behaviour of classical waves creating an interference pattern. Screen one shows the signal that one would see in each scenario (i.e classical particles or waves) if only one slit was open. If slit 1 was open we would see P_1 and so on. Screen 2 shows the signal we see if both the silts are open in each scenario.

This standard explanation works because we assume the particle is the only quantum-mechanical system here, and it can be describe by a pure state, $|\phi\rangle_{\text{particle}} = |\phi_1\rangle + |\phi_2\rangle$. But if quantum mechanics is correct, this particle is not the only thing to be described by a wavefunction. For example, we can also describe the double-slit by its wavefunction $|\psi\rangle_{\text{ds}}$. The total wavefunction of the combined system can be in a pure state,

$$|\Psi\rangle_{\text{combined}} = |\psi_1\rangle_{\text{ds}}|\phi_1\rangle_{\text{particle}} + |\psi_2\rangle_{\text{ds}}|\phi_2\rangle_{\text{particle}} \ , \tag{1}$$

but either subsystem does not have to be pure.

When the two double-slit states are almost indistinguishable,

$$\langle\psi_1|\psi_2\rangle\langle\psi_2|\psi_1\rangle \approx 1 \ , \tag{2}$$

then the combined system factorizes,

$$|\Psi\rangle_{\text{combined}} \approx |\psi_1\rangle_{\text{ds}} \left(|\phi_1\rangle_{\text{particle}} + |\phi_2\rangle_{\text{particle}}\right) \ . \tag{3}$$

The particle subsystem indeed stays as a pure state and there will be an interference pattern.

On the other hand, if the two double-slit states are distinguishable,

$$\langle\psi_1|\psi_2\rangle\langle\psi_2|\psi_1\rangle \ll 1 \ , \tag{4}$$

that means the interaction entangled the two systems. The particle subsystem becomes a mixed state,

$$\rho_{\rm particle} = {\rm Tr}_{\rm ds}|\Psi\rangle\langle\Psi| \approx |\phi_1\rangle\langle\phi_1| + |\phi_2\rangle\langle\phi_2| \ , \tag{5}$$

and there will be no interference pattern.

Whether the double-slit states are given by Eq. (2) or (4) seems to require the knowledge about the actual quantum-mechanical interacton between the particles and the double-slit, which is unavailable in practice. Feynman pushed the above analysis further to overcome such problem. He pointed out that even if we try to keep the interaction minimal, there will be one inevitable interaction that makes $|\psi_1\rangle$ and $|\psi_2\rangle$ different. That is because when a particle reaches some place on the screen, its y-momentum must be different depending on which slit it passed through.

$$\Delta p_y^{\rm particle} \equiv |\langle\phi_1|p_y|\phi_1\rangle - \langle\phi_2|p_y|\phi_2\rangle| \approx p_x \frac{s}{L} \ , \tag{6}$$

where p_x is the x momentum of the particles, s is the separation between the two slits, and L is the distance to the screen. This means two different values of recoil momentum on the double-slit,

$$\Delta p_y^{\rm ds} = \Delta p_y^{\rm particle} \approx p_x \frac{s}{L} \ . \tag{7}$$

The uncertainty principle sets a limit on how well we can measure this difference. Namely, we need a large uncertainty in the position to measure a small change in momentum. Setting $\hbar$ to 1, we can say that when the momentum difference is too small to be measured,

$$\Delta p_y^{\rm ds} < 1/\Delta y^{\rm ds} \ , \tag{8}$$

then Eq. (2) is true and the two states are indistinguishable. Otherwise Eq. (4) is true and the two states are distinguishable.

The key point is that there is an alternative way to appreciate Eq. (8) *without* thinking about the quantum mechanics of the double-slit. First of all, the interference pattern has a fringe width of

$$w = L\frac{\lambda}{s} \ , \tag{9}$$

where $\lambda = p_x^{-1}$ is the de Broglie wavelength of the particle. This is the separation of bright/dark lines in the y direction with a fixed reference point at the y position of the double-slit. Thus the uncertainty in the y position must be smaller than this value, so the interference pattern is not totally blurred.

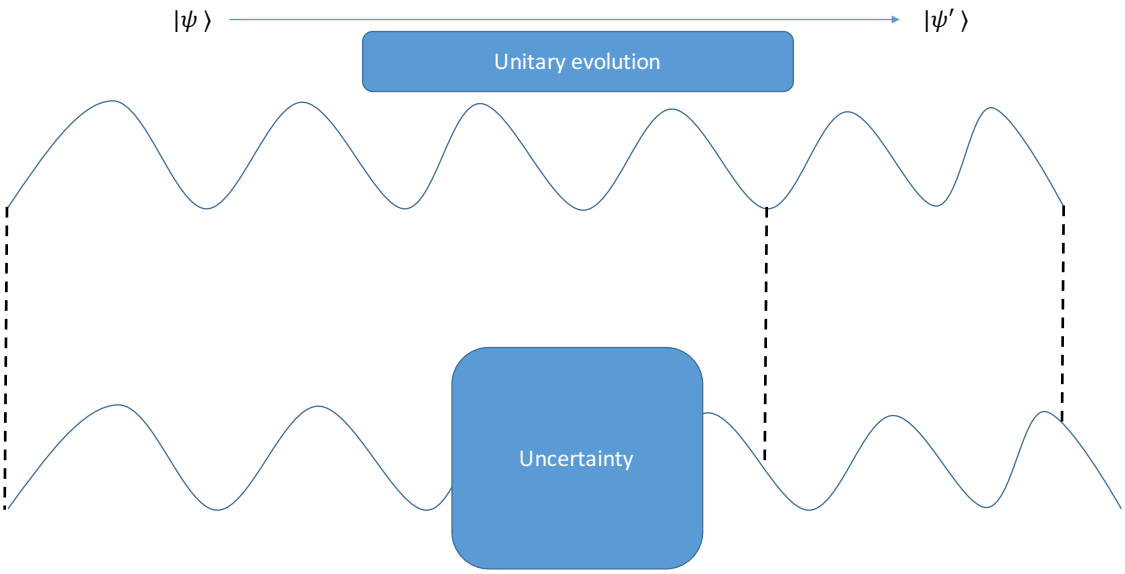
$$\Delta y^{\rm ds} < w = \frac{L}{p_x s} \ . \tag{10}$$

This is exactly the same condition as in Eq. (8).

We should emphasize the key value of this result. We can treat both the position uncertainty $\Delta y^{\rm ds}$ and the interference fringe width w as classical quantities. Eq. (10) is then a purely classical calculation to compare these two quantities, which tells us whether a classical wave pattern (the intereference pattern) loses coherence. If we only take its face value, it seems to be only a practical obstacle of measuring the interference pattern. One might argue that the underlying subsystem unitarity is still valid, just difficult to measure.

What Feynman showed by this example is that Eq. (10) tells us exactly the same thing as Eq. (8). The later **is** directly about quantum mechanics and shows how the subsystem unitarity can fail. Thus an apparently classical calculation can tell us something about the hidden quantum-mechanical nature of the interaction between the double-slit and the particle passing through it. In this example, it tells us whether the particles get entangled with the double-slit and loses its subsystem unitarity.

One might object that the Eq. (10) is not entirely classical, since the value of $\Delta y^{\rm ds}$ has to originate from the unknown quantum mechanics of the double-slit. This is a valid concern, and it actually shows another strength of Eq. (10). Indeed we do not know the value of $\Delta y^{\rm ds}$ just from classical physics, but it is very reasonable to assume that it is intrinsic to the double-slit. Namely, its value should be fixed if we move the screen further away. Eq. (10) shows that with a fixed $\Delta y^{\rm ds}$, we can always make L large enough to satisfy this condition. Thus, at least at the level of gedanken experiment, one can always arrange a situation that the interference pattern is visible, thus the subsystem unitarity is valid.



.

FIG. 2. Schematic showing the rule we advocate; as classical coherence fails, we loose quantum unitarity.

III. ANALOGY IN QUANTUM FIELD THEORY

The double-slit interference is an example that both the classical coherence and quantum unitarity can be calculated explicitly and show that they follow the same condition. Here, we advocate that such a relation is actually a general rule. *Classical coherence, instead of thinking of it as a practicality of measurements, should be treated as a prerequisit to quantum unitarity.* Thus when classical coherence fails, not only have we no practical measurements to confirm the purity of the subsystem wavefunction, such wavefunctions *actually did not evolve unitarily.* Whatever effect responsible for a classical uncertainty large enough to disrupt classical coherence must have also interacted with the subsystem quantum-mechanically and became entangled with it.

For quantum field theory in a curved spacetime, we always assume that the quantum field is a subsystem that never entangles with the geometry, thus remains unitary. Since we do not know quantum gravity/geometry, we cannot directly check this assumption by an actual calculation of entanglement. However, we can always check classical coherence, and by this rule we advocate, the answer directly proves/disproves subsystem unitarity.

If we go back to the basics of quantum field theory, there is a natural description of this new rule. As shown in Fig.2, QFT starts from solving a classical mode function, then quantizing its amplitude into a quantum state. Any method to measure such quantum state assumes the knowledge of the classical mode function. If the geometric uncertainty leads to a significant uncertainty in the classical mode function, then there is no practical way to reliably measure its quantum state. We advocate that this is not only a practicality about measuring the quantum state, but it directly tells us that the quantum state loses its unitarity. Just like in Feynman's example where the uncertainty of the double-slit guarantees its entanglement with the particles, whatever effect that leads to the geometric uncertainty here must entangle with the quantum state of this mode. Without a theory of quantum geometry, we cannot describe how that happens. But through a classical calculation, we can determine whether it has happened or not.

While such generalization to QFT seems straightforward in theory, there are a few subtleties in practice. In the double-slit experiment, the position uncertainty of the double-slit was the only obvious uncertainty of the background that we need to keep track of. It is also the reasonable thing to hold fixed while moving the projection screen away. In QFT, the background is the entire geometry, and there are infinitely many ways a geometry can be uncertain. For a full analogy to Feynman's analysis, not only do we need to parametrize those uncertainties, we also need to choose the appropriate combination to hold fixed as we take a similar limit. At this moment, we do not have a general formalism to setup such self-consistency check that can be applied to any general geometry. In Sec.IV, we propose a specific way to apply such self-consistency check to the propagation of Hawking quanta. In addition to gaining insights into

the information paradox, we hope to learn some lessons to inspire a more general formalism of such self-consistency check.

QFT theorists might think that a more direct self-consistency check is feasible. Instead of parametrizing a classical geometric uncertainty, one can treat deviations from the background geometry as a field and quantize it as any other fields. Then a standard QFT calculation of scattering with this new field (graviton) should capture all interactions with the geometry, and a small scattering amplitude should guarantee that the interaction is suppressed. The concern for such method is that IR issues for these scattering calculations cannot be unambiguously regulated like in global Minkowski space. Since soft gravitons lead to real physical changes in the geometry that can be seen in memory effects [13, 14], an IR ambiguity signals our ignorance on how these geometric changes enter QFT. Thus a direct graviton scattering calculation cannot be the full extent of a self-consistency check, and it is reasonable for our method to give extra constraints on the reliability of local QFT.

IV. HAWKING QUANTA PROPAGATION

The usual treatment of how a near-horizon mode becomes an asymptotic Hawking quantum is also following a fixed-background assumption. The background is the Schwarzschild geometry.

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \frac{dr^2}{1-\frac{2M}{r}} + r^2 d\Omega_2^2 \ . \tag{11}$$

When one applies local quantum field theory to describe a field on this fixed background, the result is by-definition unitary, but that is not necessary a physical fact. Given how little we understand quantum gravity, a classical self-consistency check as we described in the previous section is a reasonable thing to do.

A. Setup and summary

- **Classical coherence by tracking geodesics.** The classical mode of a massless field basically has its peaks and nodes following null geodesics. Thus, instead of actually solving the mode function, we can perform a much simpler calculation to track geodesics. We start with two out-going null geodesics with the separation of one period of such mode near the horizon, and then we calculate the change in their separations at spatial infinity. As this change increases, one should gradually lose faith in the unitary propagation of such mode, since it becomes increasingly difficult to experimentally verify.

- **The background and the uncertainty.** The classical background in this case is the Schwarzschild geometry, which can be written as the metric $g_{\mu\nu}$. A classical uncertainty can be expressed as deviation from such metric, $\Delta g_{\mu\nu}$. There are two major challenges here.

  1. There are many more variables than the double-slit experiment. We consider only one particular form of $\Delta g_{\mu\nu}$ in this paper—back-reactions from the presence of extra matter of zero total energy. It is simple to calculate and reflects the physical intuition of vacuum pair fluctuation.
  2. The value of $\Delta g_{\mu\nu}$ is gauge dependent. By relating it to the presence of extra matter, we can express it as a gauge-invariant local quantity, which is the natural thing to hold fixed in our analysis.

  Note that if we were reporting a positive result, that such uncertainty upholds unitarity, then one should question the validaty of learning a general lesson from a special example. However, we will show that this particular form of $\Delta g_{\mu\nu}$ is already sufficient to question unitarity, and we did not choose this form to specifically do that. Thus our limited analysis is sufficient to raise a reasonable doubt to the classical-background approximation.

- **Renormalization.** It is well-known that naïve applications of QFT often suffer from UV/IR divergences. In our case, we will see that even in Minkowski space, geometric uncertainty already leads to a finite change between the two null geodesics. We will simply "renormalize" that away by saying that if the black hole case leads to a similar change, it should not be taken as a serious problem. Our surprising result is that in the black hole case, we get an infinitely larger change, which is difficult to blame on the usual divergence of QFT.

More concretely, our setup is shown in Fig.3. We start by choosing two points at some $r_0 \gg M$ with a $\delta\bar{t}_0 \sim M$ coordinate time separation between them. This represents one wavelength of the mode function of an asymptotic Hawking quantum. We then back-track two null geodesics from these two points back to two points near the horizon,

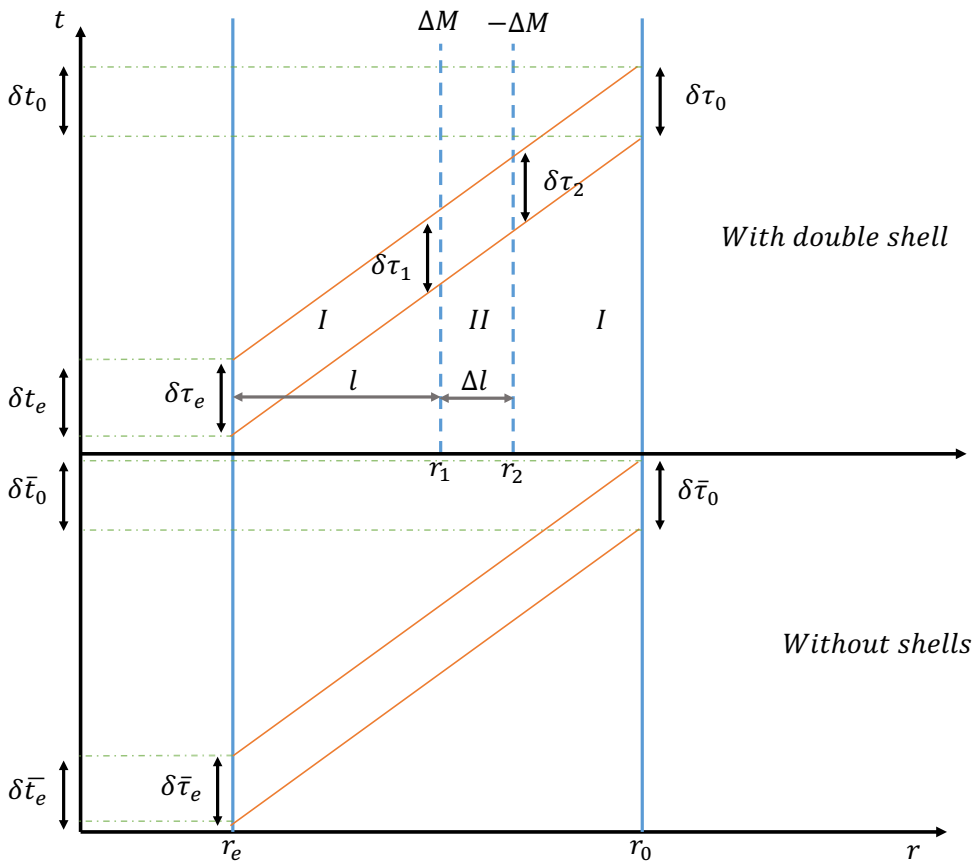


FIG. 3. The two parts of this diagrams describe two different scenarios. The trajectories of the photon are the lines in orange. The trajectory of the observer and the position from which the photons are emitted are represented by solid blue lines. The bottom part is a general case of two Hawking photons coming from a distance r_e from the black hole of mass M and arriving at an observer who is at a distance of r_0. The top part shows two Hawking photons coming from the same distance r_e from a black hole of mass M. Instead of the photons freely propagating through to an observer at r_0, they have to cross two shells, represented by blue dotted line, of equal and opposite mass $\Delta M, -\Delta M$ at r_1, r_2 respectively. The regions I represent a metric with mass M and region II represents a metric with mass $M + \Delta M$.

at r_e with $(r_e - 2M) \ll 2M$. Now their separation represents one wavelength of a near-horizon mode, which would have propagated to the asymptotic mode if there were no geometric uncertainty. We then introduce introduce a geometric uncertainty as two shells of zero total ADM mass. Therefore, both the near horizon and the asymptotic regions are not affected by such uncertainty. Only the process of propagation is affected. With the same two starting points near the horizon we calculate the separation δt_0 of the two null geodesics when they arrive at r_0, and compare it with $\delta\bar{t}_0$.

Note that we have assumed two stationary and spherically symmetric shells to simplify the calculation, but that is not the actual physical picture. There is no reason why vacuum fluctuation suddenly manifests two complete shells together to surround the black hole, and holding them at a fixed location will require unphysically large pressure if they are close to the horizon. As it will become clear later, our result only cares about the local energy density of two small pieces of shells at the "location and time" where the pair of null rays actually pass through. It does not care about whether the whole shell exists or not, and the pressure does not directly affect the answer. Our result is also invariant under radial boosts, thus it does not matter how the shells move before and after the crossing, or whether the shells are actually stationary.

After some formal calculations in Appendix A, we arrive at this simple expression:

$$\frac{\delta t_0 - \delta\bar{t}_0}{\delta\bar{t}_0} = \left(\sqrt{\frac{r_2 - 2M}{r_1 - 2M}} \sqrt{\frac{r_1 - 2(M + \Delta M)}{r_2 - 2(M + \Delta M)}} - 1 \right) , \tag{12}$$

where r_1 is the location of the inner shell, r_2 is the location of the outer shell, and $\pm\Delta M$ are their individual ADM masses.

B. Physical interpretation

In the entire calculation, we will stay in the regime that $r_1, 2M, (r_1 - 2M) \gg \Delta M, \Delta r$, where $\Delta r \equiv (r_2 - r_1)$. This guarantees that the geometric fluctuation is small, as it only affects a small region Δr, and the change in the metric $\Delta g_{\mu\nu}/g_{\mu\nu}$ is small in the affected region. Under this assumption, we can expand Eq (12) to get

$$\frac{\delta t_0 - \delta\bar{t}_0}{\delta\bar{t}_0} = \frac{\Delta M \Delta r}{(r_1 - 2M)^2} = \frac{\Delta M \Delta r}{r_1^2} g_{tt}^{-2} \ . \tag{13}$$

Here $g_{tt} = (1 - 2M/r_1)$ is the time component of the metric in the background geometry, evaluated at the location where the null rays pass through the geometric uncertainty.

It is more enlightening to write this expression in terms of local physical quantities (see Appendix B for derivations). Let σ be the energy density of the shell in its rest frame, and Δl be the physical distance between the shells in that frame, we get

$$\frac{\delta t_0 - \delta\bar{t}_0}{\delta\bar{t}_0} = 4\pi\sigma\Delta l g_{tt}^{-1} \ . \tag{14}$$

Here we can see that the effect is due to the local energy density which the null rays pass through. Also, the combination $(\sigma\Delta l)$ is invariant under radial boosts, thus Eq. (14) does not care about whether the shells are actually at rest, nor in which frame we calculate it.

In order to appreciate the effect of such geometric uncertainty near the horizon of a large black hole (which is usually considered the ideal situation to formulate the information paradox), we evaluate Eq. (14) in three regimes; 1) The Minkowski limit: $M = 0$; 2) Small black holes / far from horizon: $M \ll r_1$; 3) Large black holes / near horizon: $r_1 \gg M$. Instead of using the coordinate r_1, we will express the answer with the physical distance l between the horizon and the shells. The relation between r and l can also be found in Appendix B. We will hold l, Δl and σ fixed while the black hole mass is being varied.

$$\frac{\delta t_0 - \delta\bar{t}_0}{\delta\bar{t}_0} \begin{cases} = 4\pi\sigma\Delta l & (M = 0) \\ \approx 4\pi\sigma\Delta l \left(1 + \frac{2M}{l}\right) & (M \ll l) \\ \approx 4\pi\sigma\Delta l \frac{16M^2}{l^2} & (M \gg l). \end{cases} \tag{15}$$

- **Minkowski limit** ($M = 0$): The naïve interpretation of the $M = 0$ result is that even in Minkowski space, geometric uncertainty can potentially decohere classical mode functions. By our definition, that poses some threat to QFT in Minkowski space. We are going to assume that unitarity in QFT is fine in Minkowski space, effectively "renormalize away" the effect at $M = 0$. One can imagine a simple subtraction by a counter term. Or alternatively, since the actual value of $(\sigma\Delta r)$ is unknown, we assume that it is small enough to not cause any concern.

- **Far from horizon** ($M \ll l$): If we send signals from the Earth surface to somewhere far in the space, this result tell us how much can we trust a unitary QFT description. It starts to show a small deviation from the Minkowski result, but in the $M \ll r_1$ limit, it is almost indistinguishable.

- **Near horizon** ($M \gg l$): If we consider a region of size l across the horizon of a large black hole $M \gg l$, QFT believes that such region is locally the same as empty Minkowski. As long as l is much larger than the UV cut-off scale of the QFT, one would believe that modes in this region propagates out to become Hawking quanta, and QFT guarantees the unitarity of such process. Here we hold $\sigma\Delta l$ fixed at the same value of empty Minkowski space, we can see that the effect of geometric uncertainty gets arbitrarily larger than in the Minkowski space. Note that the absolute value of the effect cannot become arbitrarily large, since it is actually bounded at order one by the validity constraint of our calculation: $\Delta g_{\mu\nu} \ll g_{\mu\nu}$. However, seeing that it can be arbitrarily larger than the Minkowski result should be a sufficient reason to raise a reasonable doubt about the unitarity of the propagation process.

V. DISCUSSION

A. The self-consistency check of local QFT

In this paper, we proposed a self-consistency check of local QFT in curved spacetime geometries. We have shown that propagation from the near-horizon region of a large black hole seems to fail such check. As explained in Sec.IV,

the consistency requires us to parametrize the uncertainty in the classical geometry. We have chosen a particular form of uncertainty in our explicit calculation. Our result is gauge invariant and makes physical sense, but there might be a more general way to parametrize such uncertainty to make our approach even more general. It is also interesting to apply the same check to other geometries, especially other types of horizons.

We should emphasize that our check does not invalidate QFT all together. We simply suggest that the QFT degrees of freedom lose their subsystem unitarity. In the double-slit experiment, with or without the interference pattern, the total number of particles that passes through the double-slit is not affected, and the expectation value of the particle location remains in the center. The point is that we can always find a pair of pure and mixed density matrices which give identical expectation values to many observables. It is not unexpected that the only affected observable is the one explicitly checking unitarity. Thus we can still rely on QFT to calculate the values of other observables. For example, the particle number and energy spectrum of Hawking radiation can remain the same as the conventional result.

Many existing results of QFT in curved/dynamical geometries are about particle numbers, such as the particle-production calculation following a Bogoliubov transformation. As an example, one can easily see that

$$\rho_{\text{pure}} = \left(\sum_n a_n |n\rangle\right)\left(\sum_n \langle n| a_n^*\right) \tag{16}$$

and

$$\rho_{\text{mixed}} = \sum_n |a_n|^2 |n\rangle\langle n| \tag{17}$$

give the same particle number. Thus there is no general crisis if we admit that local QFT sometimes loses subsystem unitarity.

As far as we know, an observable that explicitly checks the subsystem unitarity of local QFT during particle-production events is the cosmological Bell inequality [15]. One might be able to apply a similar self-consistency check to such a model and update the prediction of whether a violation of Bell inequality is actually expected.

B. The information paradox

While formulating the information paradox, it is customary to take the large black hole limit. Such limit allows us to identify modes with wavelength much shorter than the curvature scale, but much longer than the UV cut-off scale. For these modes, we are confident that they start in local vacuum states of QFT as if they live in Minkowski space. This large black hole limit, which is effectively a near-Minkowski limit of a near horizon region, is the key argument to eliminate many naïve solutions to the information paradox.

The main result in this paper shows that there is a competing effect which stops us from taking this convenient limit. The geometric uncertainty at distance l from the horizon will only affect near-horizon modes of wavelength shorter than l. If we fix the wavelength of the relevant near-horizon modes and increase the black hole size, our result shows that the unitary propagation to asymptotic Hawking radiation becomes less and less trustworthy. Since the dimensionless ratio (l/M) is the only relevant physical parameter that comes into the scaling both effects, there seems to be no middle-ground that can make both unitarity and local-Minkowski-ness arbitrarily reliable. We think such competing nature is a good sign and will take us closer to the solution of information paradox.

Let us assume that indeed the QFT-subsystem unitarity is lost during the propagation of Hawking quanta. This alone does not resolve all the problems. Usually, one pictures a non-unitarity process as losing information. In the context of black hole information, a Hawking quantum actually has to *recover* information. A near-horizon mode starts from a maximally mixed (thermal) state with no information, since its partner is the interior mode that never comes out of the horizon. After the non-unitarity propagation to become an asymptotic Hawking quantum, it has to carry information to purify the rest of the Hawking radiation there.

Since unitarity is lost to the local geometric uncertainty, what restores information must be the interaction with the local geometry. An abstract realization of this information-recovery process is recently demonstrated in [8]. In order for local geometry to give this information to the Hawking quanta, it must first carry such information. If was first suggested that superpositions of different classical geometries can store such information [16]. In the conventional picture, classical geometries are determined by only a few parameters such as the location and size of the black hole. Thus the major objection to such picture is that the superpositions of classical geometries do not form a large enough Hilbert space to hold the required information. Recently, it was realized that there are actually a much larger number of classical geometries differed by their soft hairs [12]. A simple counting suggested that the size of Hilbert space seems no longer the problem, so one should pursue this line of thoughts even further.

Classically, soft hairs manifest as memory effects that alter the distances between geodesics, which is very similar to the effect of geometric uncertainty as we showed in this paper. Two neighbouring null rays simply "remember" the geometric uncertainty they passed through. Naïvely speaking, a propagating mode of a massless field should remember a similar effect. Based on this picture, one may try to quantize the soft hairs and build a toy model that involves quantum interaction, therefore entanglement, between the soft hairs and the usual QFT modes. That would provide a complete picture that the loss of subsystem unitarity is really just entanglement with another subsystem which was originally (and illegitimately, as we argued) assumed as the classical background.

ACKNOWLEDGMENTS

We thank XXX for discussions. We are supported by XXX.

Appendix A: Calculation of proper time change

1. Background spacetime - no shell

A black hole of mass M defines a Schwarzschild geometry for the spacetime with the following metric

$$ds^2 = -\left(1 - \frac{2M}{r}\right) d\bar{t}^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega_2^2 \tag{A1}$$

where we are working with units in which $G = c = 1$. The bar is used over the time coordinate to differentiate it from the time coordinate we will use in the spacetime with mass M'. We know there will be a flux of particles appearing from the horizon of the black hole [17]. Take some coordinate time interval, $\delta\bar{t_e}$, (which will have some corresponding proper time interval $\delta\bar{\tau_e}$) between two consecutive particles being emitted from the same position above the horizon of the black hole. An observer far away will see these particles arrive with some coordinate time interval $\delta\bar{t_0}$. We can follow the trajectory of the particles (assuming they are relativistic) from the position they are emitted, r_e, to the observer at r_0 by setting $ds = 0$ in Eq (A1). By confining the motion to be only in the radial direction, i.e $d\Omega_2 = 0$, we get the following geodesic equation

$$\bar{t_e}^{(1)} - \bar{t_0}^{(1)} = r_0 - r_e + 2M \ln\left(\frac{r_0 - 2M}{r_e - 2M}\right) \tag{A2}$$

where the superscript represents particle 1. An analogous geodesic equation is given for particle 2.

$$\bar{t_e}^{(2)} - \bar{t_0}^{(2)} = r_0 - r_e + 2M \ln\left(\frac{r_0 - 2M}{r_e - 2M}\right). \tag{A3}$$

Since the right hand side of Eq (A2) and (A3) is the same, we see that

$$\bar{t_0}^{(2)} - \bar{t_0}^{(1)} = \bar{t_e}^{(2)} - \bar{t_e}^{(1)} \Rightarrow \delta\bar{t_0} = \delta\bar{t_e}. \tag{A4}$$

Therefore the coordinate time interval when the particles are emitted remains the same when it is observed at distance r_0 as is shown in the bottom part of figure 3.

2. Perturbed spacetime - double shell

Now we introduce two shells of matter with equal and opposite ADM mass, ΔM and $-\Delta M$ where $|\Delta M| << M$, at a distance of r_1 and r_2 respectively as shown in the top part of figure 3. We model these shells as infinitesimally thin and therefore use the Israel Junction Conditions (IJC) [18] with a delta function shell to analyze the effects of the shells. In this case the particle's move through two different spacetime regions. Region I is defined by the black hole of mass M and region II is defined by a Schwarzschild metric of mass $M' \equiv M + \Delta M$ as shown in the top part of figure 3. The metric in region II is given by

$$ds^2 = -\left(1 - \frac{2M'}{r}\right) dt^2 + \left(1 - \frac{2M'}{r}\right)^{-1} dr^2 + r^2 d\Omega_2^2. \tag{A5}$$

Note that the time coordinate is different from the one in Eq (A1). The radial coordinate is however the same as the shell connecting the two spacetimes is a physical surface with a fixed radius r. As explained in the previous section, the proper time interval between two particles emitted at r_e, $\delta\bar{t_e}$, is the same as the coordinate time interval at r_1, $\delta\bar{t_1}$. Now we need to find the corresponding coordinate time interval in region II, δt_1. To do this we simply note that the corresponding proper time at r_1, $\delta\bar{\tau_1}$ must be the same in both regions of spacetime - this is a simple consequence of the IJC. The coordinate time intervals at r_1, in the two different regions, are therefore related by

$$\delta\bar{t_1} = \delta\bar{t_e} = \left(\frac{1 - \frac{2(M+\Delta M)}{r_1}}{1 - \frac{2M}{r_1}}\right)^{\frac{1}{2}} \delta t_1. \tag{A6}$$

We can propagate the particles through region II from r_1 to r_2 and we know $\delta t_1 = \delta t_2$. From the fact that the proper time corresponding to the coordinate time intervals will be the same, we can relate δt_2 back to $\delta\bar{t_2}$ in region I,

$$\delta t_2 = \delta t_1 = \left(\frac{1 - \frac{2M}{r_2}}{1 - \frac{2(M+\Delta M)}{r_2}}\right) \delta\bar{t_2}. \tag{A7}$$

Since $\delta\bar{t_2}$ is equal to the time interval observed by the observer at r_0, $\delta\bar{t_0}$, we can find the proper time observed by the observer, $\delta\tau_0 = \delta\bar{t_e}\left(\left(\frac{r_2-2M}{r_1-2M}\right)\left(\frac{r_1-2(M+\Delta M)}{r_2-2(M+\Delta M)}\right)\right)^{\frac{1}{2}}$. Where we have dropped the $\left(1-\frac{2M}{r_0}\right)^{\frac{1}{2}}$ term since we assume the observer is very far away. $\delta\bar{t_e} = \delta\bar{\tau_0}$ when the observer is far away therefore we can define a dimensionless quantity that quantifies the change in proper time intervals for consecutive photons in perturbed spacetime case and the unperturbed case,

$$\frac{\delta\tau_0 - \delta\bar{\tau_0}}{\delta\bar{\tau_0}} = \left(\left(\frac{r_2 - 2M}{r_1 - 2M}\right)^{\frac{1}{2}} \left(\frac{r_1 - 2(M+\Delta M)}{r_2 - 2(M+\Delta M)}\right)^{\frac{1}{2}} - 1\right). \tag{A8}$$

This shows that when there are no shells, $\Delta M = 0$, there is no change in proper time intervals which is what one would expect. By assuming $r_1 >> \Delta r \equiv r_2 - r_1$ and $r_1 >> \Delta M$ we can expand Eq (12), to leading order in small quantities,

$$\begin{aligned}
\frac{\delta t_0 - \delta\bar{t_0}}{\delta\bar{t_0}} &= \left(\left(1 + \frac{\Delta r}{r_1 - 2M}\right)^{\frac{1}{2}} \left(1 + \frac{2\Delta M}{r_1 - 2M}\right)^{\frac{1}{2}} \left(1 + \frac{\Delta r + 2\Delta M}{r_1 - 2M}\right)^{-\frac{1}{2}} - 1\right) \\
&= -\frac{1}{8}\left(\frac{\Delta r}{r_1 - 2M}\right)^2 + \frac{1}{2}\frac{\Delta r \Delta M}{(r_1 - 2M)^2} - \frac{1}{2}\left(\frac{\Delta M}{r_1 - 2M}\right)^2 - \frac{1}{4}\frac{(\Delta r + 2\Delta M)^2}{(r_1 - 2M)^2} + \frac{3}{8}\frac{(\Delta r + 2\Delta M)^2}{(r_1 - 2M)^2} \\
&= \frac{\Delta M \Delta r}{(r_1 - 2M)^2} = \frac{\Delta M \Delta r}{r_1^2} g_{tt}^{-2}.
\end{aligned} \tag{A9}$$

Appendix B: Local physical quantities

The first step in this derivation is write down the physical, local, quantities needed to replace $\Delta M, r_1$ and Δr. The natural thing to replace ΔM is the surface stress energy, σ, of the shell. This is given by the tt component of the surface stress energy tensor S_{ab} given by the IJC.

$$S_b^a = [K_b^a] - [K]h_b^a, \tag{B1}$$

where K_{ab} is the extrinsic curvature corresponding to the induced metric, h_{ab}, we are using. $K = K_{ab}h^{ab}$ is the trace of the extrinsic curvature. We can define the induced metrics for constant r values; for region I it is

$$ds^2_{3M} = -\left(1 - \frac{2M}{r}\right)^{-1} dt^2 + r^2 d\Omega_2^2. \tag{B2}$$

Analogously we have the metric corresponding to mass $M' = M + \Delta M$ in region II

$$ds^2_{3M'} = -\left(1 - \frac{2M'}{r}\right)^{-1} dt'^2 + r^2 d\Omega_2^2. \tag{B3}$$

We could rewrite Eq (B2) and (B3) in Gaussian normal coordinates (where the coefficient in front of the time component is 1), however the first order junction condition is that the induced metric on both sides of the shell must be the same. Therefore it is easy to find a relation between the two time coordinates

$$dt'^2 = \left(\frac{1 - \frac{2M'}{r_1}}{1 - \frac{2M}{r_1}}\right) dt^2. \tag{B4}$$

Using this, we calculate the extrinsic curvature components

$$\begin{aligned}(K^t_t)^{(M)} &= \frac{M}{r_1^2}\left(1 - \frac{2M}{r_1}\right)^{-\frac{1}{2}} \\ (K^\theta_\theta)^{(M)} &= \frac{1}{r_1}\left(1 - \frac{2M}{r_1}\right)^{\frac{1}{2}} = (K^\phi_\phi)^{(M)}\end{aligned} \tag{B5}$$

where the superscript of M denotes the quantities evaluated in the metric of mass M. Analogous expressions hold for the extrinsic curvature components calculated in the metric of mass M' by just replacing M by M' in Eq (B5). The time component of the surface stress energy tensor S^0_0 is

$$\sigma = S^0_0 = \frac{1}{4\pi r_1}\left(\left(1 - \frac{2M}{r_1}\right)^{\frac{1}{2}} - \left(1 - \frac{2(M + \Delta M)}{r_1}\right)^{\frac{1}{2}}\right). \tag{B6}$$

This expression represents a small change in the mass and thus can be written as

$$\begin{aligned}\sigma &= \frac{1}{4\pi r_1}\left[\partial_X\left(1 - \frac{2X}{r_1}\right)^{\frac{1}{2}}\right]_{X=M} \Delta M. \\ &= \frac{\Delta M}{4\pi r_1^2} g_{tt}^{-\frac{1}{2}}\end{aligned} \tag{B7}$$

Next we need to replace the coordinate distance Δr with the proper distance Δl and the relation between them is given by

$$\Delta l = \left(1 - \frac{2M}{r_1}\right)^{-\frac{1}{2}} \Delta r = g_{tt}^{-\frac{1}{2}} \Delta r. \tag{B8}$$

This can be integrated to give the proper distance, l, between the inner shell and the horizon

$$\begin{aligned}l &= \int_{2M}^{r_1}\left(1 - \frac{2M}{r}\right)^{-\frac{1}{2}} dr \\ &= \sqrt{(r_1 - 2M)r_1} + M \log\left(\frac{r_1 - M + \sqrt{(r_1 - 2M)r_1}}{M}\right)\end{aligned} \tag{B9}$$

We will be interested in three limits of $M = 0$, $M \ll l$ and $M \gg l$ and the value of l in these three limits is

$$l \begin{cases} = r_1 & (M = 0) \\ \approx r_1 - 2M & (M \ll l) \\ \approx \sqrt{8M(r_1 - 2M)} & (M \gg l). \end{cases} \tag{B10}$$

[1] S. W. Hawking, Phys. Rev. D **14**, 2460 (1976).
[2] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, JHEP **1302**, 062 (2013), arXiv:1207.3123 [hep-th].
[3] J. D. Bekenstein, Phys. Rev. D **49**, 1912 (1994), gr-qc/9307035.
[4] S. L. Braunstein, S. Pirandola, and K. Życzkowski, Physical Review Letters 110, **101301** (2013), arXiv:0907.1190 [quant-ph].
[5] J. Maldacena and L. Susskind, (2013), arXiv:1306.0533 [hep-th].
[6] S. B. Giddings, (2012), arXiv:1211.7070 [hep-th].
[7] M. Dodelson and E. Silverstein, (2015), arXiv:1504.05536 [hep-th].
[8] K. Osuga and D. N. Page, (2016), arXiv:1607.04642 [hep-th].
[9] L. Hui and I.-S. Yang, (2013), arXiv:1308.6268 [hep-th].
[10] I. Ilgin and I.-S. Yang, (2013), arXiv:1311.1219 [hep-th].
[11] D. A. Lowe and L. Thorlacius, Phys. Lett. **B737**, 320 (2014), arXiv:1402.4545 [hep-th].
[12] S. W. Hawking, M. J. Perry, and A. Strominger, Phys. Rev. Lett. **116**, 231301 (2016), arXiv:1601.00921 [hep-th].
[13] S. Weinberg, Phys. Rev. **140**, B516 (1965).
[14] T. He, V. Lysov, P. Mitra, and A. Strominger, JHEP **05**, 151 (2015), arXiv:1401.7026 [hep-th].
[15] J. Maldacena, Fortsch. Phys. **64**, 10 (2016), arXiv:1508.01082 [hep-th].
[16] Y. Nomura, J. Varela, and S. J. Weinberg, Phys.Rev. **D87**, 084050 (2013), arXiv:1210.6348 [hep-th].
[17] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
[18] W. Israel, Nuovo Cim. **B44S10**, 1 (1966).