PHYSICAL REVIEW D **89,** 086010 (2014)

State-dependent bulk-boundary maps and black hole complementarity

Kyriakos Papadodimas[1,2,*] and Suvrat Raju[3,†]

[1]*Centre for Theoretical Physics, University of Groningen, Nijenborgh 4, 9747 AG, The Netherlands*
[2]*Theory Group, Physics Department, CERN, CH-1211 Geneva 23, Switzerland*
[3]*International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, IISc Campus, Bengaluru 560012, India*

(Received 9 November 2013; published 29 April 2014)

We provide a simple and explicit construction of local bulk operators that describe the interior of a black hole in the AdS/CFT correspondence. The existence of these operators is predicated on the assumption that the mapping of CFT operators to local bulk operators depends on the state of the CFT. We show that our construction leads to an exactly local effective field theory in the bulk. Barring the fact that their charge and energy can be measured at infinity, we show that the commutator of local operators inside and outside the black hole vanishes exactly, when evaluated within correlation functions of the CFT. Our construction leads to a natural resolution of the strong subadditivity paradox of Mathur and Almheiri *et al.* Furthermore, we show how, using these operators, it is possible to reconcile small corrections to effective field theory correlators with the unitarity of black hole evaporation. We address and resolve all other arguments, advanced in A. Almheiri *et al.* J. High Energy Phys. 09 (**2013**) 018 and D. Marolf and J. Polchinski, Phys. Rev. Lett. **111**, 171301 (2013), in favor of structure at the black hole horizon. We extend our construction to states that are near equilibrium, and thereby also address the "frozen vacuum" objections of R. Bousso, Phys. Rev. Lett. **112**, 041102 (2014). Finally, we explore an intriguing link between our construction of interior operators and Tomita-Takesaki theory.

DOI: 10.1103/PhysRevD.89.086010 PACS numbers: 11.25.Tq, 04.70.Dy, 89.70.Cf

I. INTRODUCTION

In a previous paper [1], we proposed a holographic description of the interior of black holes in anti-de Sitter space (AdS). In this paper we expand on several aspects of our proposal and address the information paradox for black holes in AdS in the light of the extensive recent discussion on the firewall proposal [2–9].

The central point that we wish to make in this paper is that the assumption that gravity can be described in a unitary quantum mechanical framework is consistent with the existence of operators $\phi_{\rm CFT}(x)$ labeled by a point x that can be interpreted as a spacetime point, and low-point correlation functions $\langle\Psi|\phi_{\rm CFT}(x_1)\ldots\phi_{\rm CFT}(x_n)|\Psi\rangle$, in the black hole state $|\Psi\rangle$ that can be understood as coming from effective field theory. These low-point correlators are the natural observables for a low-energy observer. However, if we take the number of points n to scale with the central charge of the boundary CFT, $\mathcal{N}$, or take two points to be very close (comparable to $l_{\rm pl}$), then this effective spacetime description may break down. Nevertheless, this breakdown is not consequential for a low-energy observer, and does not imply the existence of firewalls or fuzzballs, or require any other construction that radically violates semiclassical intuition.

A key feature of our description of local operators in this paper is that mapping between CFT operators to the bulk-local operator $\phi_{\rm CFT}(x)$ depends on the state of the CFT. This is not a violation of quantum mechanics: the operator $\phi_{\rm CFT}(x)$ is an ordinary operator that maps states to states in the Hilbert space. However, it has a useful physical interpretation as a local operator only in a given state. Said another way, the analysis in this paper relies on the assumption that to obtain a convenient description of the physics, in terms of a local spacetime, we need to use different operators in different states. This issue is related to the issue of whether it is possible to have "background independent" local operators in quantum gravity. If one gives up the idea of "background independence," one is naturally led to the "state-dependent" constructions that we discuss here.

Nevertheless, granting this assumption, we show that our construction resolves all the arguments that have been advanced to suggest that the black hole horizon has structure, or that AdS/CFT does not describe the interior of the black hole.

In our previous paper [1], we had proposed a construction of interior operators by positing a decomposition of the CFT Hilbert space into "coarse" and "fine" parts. In this paper, we present a refinement of our proposal that does *not* rely on any such explicit decomposition, although it reduces to our previous proposal in simple cases. The feature of state dependence of the interior operators carries over from [1]. But our refined construction removes some of the ambiguity inherently present in our previous proposal, and allows us to write down an explicit formula for

*kyriakos.papadodimas@cern.ch
†suvrat@icts.res.in

interior operators in the CFT, without necessarily understanding the detailed structure of its Hilbert space at strong coupling.

The thrust of our paper is rather simple to summarize. First, we point out that the issue of whether there is structure at the horizon of the black hole, and the related issue of whether the black hole interior is visible in the CFT, can be translated to a simple question about CFT operators. It is well known that local operators outside the black hole horizon can approximately be mapped to modes of single-trace operators on the boundary, which we call $\mathcal{O}^i_{\omega_n,\boldsymbol{m}}$, where i labels the conformal primary and $\omega_n, \boldsymbol{m}$ are its modes in frequency space and the angular momentum on the spatial sphere. To describe a smooth interior, we need to effectively "double" these modes and find another set of operators $\tilde{\mathcal{O}}^i_{\omega_n,\boldsymbol{m}}$, which not only commute with the original operators but are also entangled with them in the state of the CFT. So, within low-point correlators, where the number of insertions of single-trace operators does not scale with the central charge $\mathcal{N}$ [$\mathcal{N} \propto N^2$ in $\mathcal{N} = 4, SU(N)$ theory], we require

$$[\tilde{\mathcal{O}}^{i_1}_{\omega_1,\boldsymbol{m}_1}, \mathcal{O}^{i_2}_{\omega_2,\boldsymbol{m}_2}]\mathcal{O}^{i_3}_{\omega_3,\boldsymbol{m}_3}\ldots\mathcal{O}^{i_K}_{\omega_K,\boldsymbol{m}_K}|\Psi\rangle = 0,$$
$$\tilde{\mathcal{O}}^{i_1}_{\omega_1,\boldsymbol{m}_1}|\Psi\rangle = e^{-\frac{\beta\omega_1}{2}}\mathcal{O}^{i_1}_{-\omega_1,-\boldsymbol{m}_1}|\Psi\rangle.$$

Several authors have pointed out that the CFT does not seem to have enough "space" for the existence of the $\tilde{\mathcal{O}}$ operators. However, our punch line is as follows. In a *given* state $|\Psi\rangle$, the equations above must hold provided we do not have too many operator insertions and $K \ll \mathcal{N}$. The set of all possible such insertions is finite, and loosely speaking, scales like $\mathcal{N}^K$. So, demanding that $\tilde{\mathcal{O}}$ has the correct behavior within low-point correlators computed in a given state simply leads to a set of linear equations for the $\tilde{\mathcal{O}}$ operators, which can be solved in the large Hilbert space of the CFT, which has a size that scales like $e^{\mathcal{N}}$ for energies below $\mathcal{N}$. Moreover, as we discuss in detail, these equations are consistent precisely when $|\Psi\rangle$ is close to being a thermal state.[1]

This analysis leads to our conclusion that it is possible to find state-dependent local operators in the bulk that commute with the local observables outside the horizon. We then proceed to show that this construction resolves *all* the recent paradoxes associated with black hole information.

First, we describe how our construction of interior operators resolves the strong subadditivity paradox. The resolution is simply that the operators inside and outside the black hole are secretly acting on the same degrees of freedom. One of the objections to this idea of black hole complementarity has been that, naively, measurements outside the black hole would not commute with those inside. As we describe in great detail, our construction is tailored to ensure that the commutator of local operators outside and inside the black hole—and all of its powers—vanish *exactly* when inserted within low-point correlators.

We turn our attention to some of the more recent arguments of [7,9], which suggest that the black hole interior cannot be described within the CFT. The authors of [7] pointed out that the $\tilde{\mathcal{O}}$ operators behind the horizon appear to satisfy the usual algebra of creation and annihilation operators, except that "creation" operator maps states in the CFT to those of a lower energy. If this were really the case, it would lead to a contradiction since the creation operator of a simple harmonic algebra always has a left-inverse, and the number of states of the CFT decrease at lower energy.

Our construction resolves this issue, because the operators behind the horizon behave like ordinary creation and annihilation operators, only when inserted within low-point correlators. Since they satisfy the algebra only in this effective sense, and not as an exact operator algebra, there is no contradiction with the "creation" operator having null vectors.

We also address the argument of [9], which we call the $N_a \neq 0$ argument. The authors of this paper pointed out, that assuming that the interior operators were some fixed operators in the CFT, the eigenstates of the number operator for a given mode outside the horizon would not necessarily be correlated with the eigenstates of the number operator for the corresponding mode inside the horizon and so the infalling observer would encounter energetic particles at the horizon. However, this conclusion fails for state-dependent operators. Our interior operators are precisely designed so that, for a generic state in the CFT and its descendants that are relevant for low-point correlators they ensure that the infalling observer sees the vacuum as he passes through the horizon. We describe this in more detail in Sec. IV D.

After having addressed these issues, we then turn to the "theorem" of [10] that small corrections cannot unitarize Hawking radiation. We point out that our construction evades the theorem because of two features: the interior of the black hole is composed of the same degrees of freedom as the exterior, and the operators inside that are correlated with those outside depend on the state of the theory.

This brings us to a final objection that has been articulated against this state-dependent construction: the "frozen vacuum" [8,11]. Although our construction suggests that the infalling observer encounters the vacuum for a generic state, it is true that there are excited states in the CFT, in which we can arrange for the infalling observer to encounter energetic particles. Our equilibrium construction already allows us to analyze such time-dependent processes. For example, we can consider a time-dependent correlation function in an equilibrium state, and our prescription provides an unambiguous answer. However, in Sec. V, we

[1]In this paper, by "thermal state" we mean a typical pure state in the high-temperature phase of the gauge theory.

discuss how to adapt our construction to build the mirror operators directly on nonequilibrium states. This extension takes advantage of the fact that it is always possible to detect deviations from thermal equilibrium by measuring low-point correlators of single-trace operators. To perform our construction on a state that is away from thermal equilibrium, we "strip off" the excitations on top of the thermal state, and then perform our construction in this base state. Low-point correlators in the excited state are now simply equated with slightly higher point correlators in the base state. We describe this construction in Sec. V.

In Sec. VI we discuss a beautiful and intriguing connection of our construction with the Tomita-Takesaki theory of modular isomorphisms of von Neumann algebras. We start this section by reviewing our construction, but from a slightly different physical emphasis. We then show how our construction can be compactly phrased in the language of Tomita-Takesaki theory. In this section, we also clearly show how our construction of the interior in this paper reduces to our previous construction [1] in simplified settings. We hope to revisit this interesting topic again in future work.

This paper is organized as follows. In Sec. II, we show that the issue of whether AdS/CFT describes the interior in an autonomous manner reduces to the issue of finding operators, which we call the "mirror" operators, with certain properties in the CFT. After outlining these constraints, we then explicitly construct operators in III that satisfy them, when inserted within low-point correlators. This central section also contains multiple examples of our construction. We show how our construction works in a general theory, in the CFT, in a toy model of decoupled harmonic oscillators, and also in the spin chain. In Sec. IV, we then apply this construction to the recent discussions of the information paradox, and find that it successfully addresses each of the recent arguments that have been raised in favor of structure at the horizon. In Sec. V, we show how to extend our construction to nonequilibrium scenarios, and thereby also resolve the issue of the "frozen vacuum." In Sec. VI, we explore the link between our construction and Tomita-Takesaki theory. Section VII contains a summary, and some open questions. The appendixes contain several other details, including a discussion of one of the first "measurement" arguments for firewalls articulated in [2].

Appendix E may be particularly interesting to the reader, who wishes to quickly get a hands-on feel for the properties of the mirror operators that we describe. This documents a computer program that numerically constructs these mirror operators in the spin-chain toy model. The essential ideas of this paper are summarized in [12], and the reader may wish to consult that paper first, and then turn here for details.

II. BULK LOCALITY: NEED FOR THE MIRROR OPERATORS

In [1], we discussed how to construct local operators outside and inside the black hole, by using an integral transform of CFT correlators. We review this construction briefly, and explain the need for the mirror operators.

Consider a generalized free-field operator $\mathcal{O}^i(t,\Omega)$ in the conformal field theory at a point t in time and Ω on the sphere S^{d-1}. By definition this is a conformal primary operator of dimension Δ, whose correlators factorize at leading order in the $\frac{1}{\mathcal{N}}$ expansion,

$$\begin{aligned}\langle 0|\mathcal{O}^i(t_1,\Omega_1)\ldots\mathcal{O}^i(t_{2n},\Omega_{2n})|0\rangle &= \frac{1}{2^n}\sum_\pi \langle 0|\mathcal{O}^i(t_{\pi_1},\Omega_{\pi_1})\mathcal{O}^i(t_{\pi_2},\Omega_{\pi_2})|0\rangle\ldots \\ &\quad\times\langle 0|\mathcal{O}^i(t_{\pi_{2n-1}},\Omega_{\pi_{2n-1}})\mathcal{O}^i(t_{\pi_{2n}},\Omega_{\pi_{2n}})|0\rangle + \mathrm{O}\left(\frac{1}{\mathcal{N}}\right),\end{aligned}$$

where π runs over the set of permutations.

In this paper, we will be interested in fields with a dimension that is much smaller than $\mathcal{N}$. We remind the reader that, as in our last paper [1], by $\mathcal{N}$, we are referring to the central charge of the CFT, and if the reader wishes to think about supersymmetric $SU(N)$ theory, then she may take $\mathcal{N}\propto N^2$.

Now, we take the CFT to be in a state $|\Psi\rangle$ that is in equilibrium and has an energy $\langle\Psi|H_{\mathrm{CFT}}|\Psi\rangle = \mathrm{O}(\mathcal{N})$. We write $\mathrm{O}(\mathcal{N})$ here, but to be precise, we need to take the energy to be much larger than the central charge so that the theory is unambiguously in the phase corresponding to a big black hole in AdS.

The same generalized free field now factorizes about this energetic state as well. Moreover, at leading order in $\frac{1}{\mathcal{N}}$, we expect that correlators in this state $|\Psi\rangle$ will be the same as thermal correlators

$$\begin{aligned}&\langle\Psi|\mathcal{O}^{i_1}(t_1,\Omega_1)\ldots\mathcal{O}^{i_n}(t_n,\Omega_n)|\Psi\rangle \\ &\quad = Z_\beta^{-1}\mathrm{Tr}(e^{-\beta H}\mathcal{O}^{i_1}(t_1,\Omega_1)\ldots\mathcal{O}^{i_n}(t_n,\Omega_n)),\end{aligned} \tag{2.1}$$

where Z_β is the partition function of the CFT at the temperature β^{-1}.

As we showed in [1] we can use the modes of this operator to construct another CFT operator that behaves like the local field outside the black hole. The formulas of [1] were written for the case of the black brane in AdS, but here we can write down the analogous formulas for the CFT on the sphere to avoid some infrared issues in discussing the information paradox,

$$\phi^i_{\mathrm{CFT}}(t,\Omega,z) = \sum_{\boldsymbol{m}}\int_{\omega>0}\frac{d\omega}{2\pi}[\mathcal{O}^i_{\omega,\boldsymbol{m}}f_{\omega,\boldsymbol{m}}(t,\Omega,z) + \mathrm{H.c.}]. \tag{2.2}$$

Here $\mathcal{O}_{\omega,\boldsymbol{m}}$ are the modes of the boundary operators in frequency space and on the sphere respectively, while the sum over $\boldsymbol{m}$ goes over the spherical harmonics.[2] What this means is that if we consider the *CFT correlators*

$$\langle\Psi|\phi^{i_1}_{\rm CFT}(t_1,\Omega_1,z_1)\ldots\phi^{i_n}_{\rm CFT}(t_n,\Omega_n,z_n)|\Psi\rangle, \tag{2.3}$$

then these CFT correlators behave like those of a perturbative field propagating in the AdS-Schwarzschild geometry. Here, we are assuming that we are in a regime of parameter space where the CFT admits a gravity dual.

The analogue of (2.2) in empty AdS had previously been discussed extensively in the literature [13–16]. However, in writing (2.2), we pointed out that, in momentum space, it was possible to extend this construction in pure states close to the thermal state. This relies on the fact that thermal CFT correlators have specific properties at large spacelike momenta, and this observation allowed us to sidestep some of the complications that were encountered in [14].[3]

Turning now to the region behind the horizon, effective field theory tells us that in the analogue of (2.2), the CFT operator describing the interior must have the form

$$\phi^{i}_{\rm CFT}(t,\Omega,z) = \sum_{\boldsymbol{m}}\int_{\omega>0}\frac{d\omega}{2\pi}[\mathcal{O}^i_{\omega,\boldsymbol{m}}g^{(1)}_{\omega,\boldsymbol{m}}(t,\Omega,z) + \tilde{\mathcal{O}}^i_{\omega,\boldsymbol{m}}g^{(2)}_{\omega,\boldsymbol{m}}(t,\Omega,z)+\text{H.c.}]. \tag{2.4}$$

Here $g^{(1)}_{\omega,\boldsymbol{m}}$ are the analytic continuations of the left-moving modes from outside to inside the black hole, while $g^{(2)}_{\omega,\boldsymbol{m}}$ are right-moving modes inside the black hole.

These right-moving modes can be understood in several ways. In Hawking's original calculation [18], these modes were the very energetic modes in the initial data that can be propagated through the infalling matter using geometric optics. In terms of solving wave equations, the $g^{(2)}_{\omega,\boldsymbol{m}}$ modes can also be obtained by analytically continuing the modes from the "other side" (region III) of the eternal black hole, as we discussed in [1].

However, we should caution the reader that while these physical interpretations are useful as mnemonics, they are both fraught with ultra-Planckian problems. This is clear in Hawking's original interpretation, but we also note that while the analytic continuation from region III is easily performed in the free-field theory, mapping the modes at late times in the black hole, 'back to region III requires us to go through the ultraviolet regime.

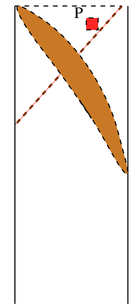

FIG. 1 (color online). A black hole is created in AdS by injecting matter from the boundary. We are interested in the red-colored patch P, behind the horizon, which is far away from both the infalling matter and the singularity.

We emphasize that neither of these ultra-Planckian problems are relevant to our discussion. Here, our statement is simply about effective field theory in the patch P that is shown in Fig. 1. In this patch, we can *locally* expand the field in modes, and we find that to get a local perturbative field, we need both left- and right-moving modes. What is important here, though, is the appearance of the modes $\tilde{\mathcal{O}}^i_{\omega_n,\boldsymbol{m}}$. First, we need these operators to *effectively* commute not only with the ordinary operators of the same species $\mathcal{O}^i_{\omega_n,\boldsymbol{m}}$, but with other "species" of operators $\mathcal{O}^j_{\omega_n,\boldsymbol{m}}$ that enter the fields outside the horizon as well

$$[\mathcal{O}^i_{\omega_1,\boldsymbol{m}_1},\tilde{\mathcal{O}}^j_{\omega_2,\boldsymbol{m}_2}]\doteq 0. \tag{2.5}$$

The $\doteq$ in (2.5) indicates that this equation must hold when this commutator (or a power of this commutator) is inserted within a low-point CFT correlator like (2.3), as we discuss in more detail below. As we have mentioned, and will discuss again below, if we consider a correlator with $\mathcal{N}$ insertions, then we should not expect a semiclassical spacetime, or an equation like (2.5) that expresses locality in such a spacetime to hold.

For the horizon of the black hole to be smooth we require that within a low-point correlator evaluated in a pure state that is close to a thermal state

$$\langle\Psi|\mathcal{O}^{i_1}(t_1,\Omega_1)\ldots\tilde{\mathcal{O}}^{j_1}(t'_1,\Omega'_1)\ldots\tilde{\mathcal{O}}^{j_l}(t'_l,\Omega'_l)\ldots\mathcal{O}^{i_n}(t_n,\Omega_n)|\Psi\rangle = Z_\beta^{-1}\mathrm{Tr}[e^{-\beta H}\mathcal{O}^{i_1}(t_1,\Omega_1)\ldots\mathcal{O}^{i_n}(t_n,\Omega_n)\mathcal{O}^{j_l}(t'_l+i\beta/2,\Omega'_l)\ldots \times\mathcal{O}^{j_1}(t'_1+i\beta/2,\Omega'_1)], \tag{2.6}$$

where Z_β is the partition function of the CFT at temperature β^{-1}. The reader should note that the analytically continued operators, which appear with the index j_p and primed

[2] These need to be suitably regulated in frequency space, and we discuss this carefully in Sec. III B, although this issue is unimportant here.

[3] It is somewhat delicate to write down the position space version of (2.2). This is because the position space "transfer function" must account for the fact that it can only be integrated against valid CFT correlators. So, the transfer function must be understood as a distribution that acts as a linear functional on the restricted domain of multipoint CFT correlators. This leads to subtleties in writing it as a simple integral transform. This observation has led to recent claims that the transfer function does "not exist" in the black hole background or, indeed, in any background with a trapped null geodesic [17]. This statement—which simply refers to the fact explained above—does not have any significant physical implication; the mapping between degrees of freedom between the bulk and the boundary continues to exist.

coordinates, have been moved to the right of all the ordinary operators, and moreover their relative ordering has been reversed.

In momentum space, Eq. (2.6) can be translated to

$$\begin{aligned}\langle\Psi|\mathcal{O}^{i_1}_{\omega_1,\boldsymbol{m}_1}\ldots\tilde{\mathcal{O}}^{j_1}_{\omega'_1,\boldsymbol{m}'_1}\ldots\tilde{\mathcal{O}}^{j_l}_{\omega'_l,\boldsymbol{m}'_l}\ldots\mathcal{O}^{i_n}_{\omega_n,\boldsymbol{m}_n}|\Psi\rangle \\ = e^{-\frac{\beta}{2}(\omega'_1+\ldots\omega'_l)}Z_\beta^{-1}\mathrm{Tr}[e^{-\beta H}\mathcal{O}^{i_1}_{\omega_1,\boldsymbol{m}_1}\ldots\mathcal{O}^{i_n}_{\omega_n,\boldsymbol{m}_n}(\mathcal{O}^{j_l}_{\omega'_l,\boldsymbol{m}'_l})^\dagger\ldots \\ \times(\mathcal{O}^{j_1}_{\omega'_1,\boldsymbol{m}'_1})^\dagger]. \end{aligned} \tag{2.7}$$

In Fourier transforming from (2.6) to (2.7), we should keep in mind that while the modes of $\mathcal{O}^i$ are defined by $\mathcal{O}^i_{\omega,\boldsymbol{m}} = \int\mathcal{O}^i(t,\Omega)e^{i\omega t}Y_m(\Omega)d^{d-1}\Omega dt$, where Y_m is the spherical harmonic on the sphere, the modes of $\tilde{O}^i$ are defined by $\tilde{O}^i_{\omega,\boldsymbol{m}} = \int\mathcal{O}^i(t,\Omega)e^{-i\omega t}Y^*_m(\Omega)d^{d-1}\Omega dt$. This convention simply tells us that the modes $\tilde{O}^i_{\omega,\boldsymbol{m}}$ have the *opposite* energy and angular momentum to the modes $\mathcal{O}^i_{\omega,\boldsymbol{m}}$.

To emphasize again, we require operators that when inserted within a state automatically achieve the ordering within the thermal trace that we have shown here: both in terms of moving to the right of ordinary operators, and in terms of reversing their relative positions.

The reader may wish to consult Sec. 5 of our previous paper [1], where we showed how the condition (2.6) leads to smooth correlators across the horizon.[4] This is clear, because in this case, the calculation of correlators across the horizon reduces to the calculation in the eternal black hole geometry, which is clearly smooth. In fact, assuming the validity of the ansatz (2.4) which we discuss next, the converse also holds: correlators are smooth across the horizon if (2.6) holds, at least at leading order in $\frac{1}{N}$.

Validity of the ansatz for bulk operators.—We briefly discuss the validity of the ansatz for interior operators (2.4), which underpins the reasoning above. Provided that operators that satisfy (2.6) can be found, we conjecture the ansatz (2.4) gives a consistent spacetime description of the interior and exterior of a black hole when the CFT is placed in a state that is expected to be dual to a black hole in the bulk.

This conjecture is a claim about the uniqueness of the functions $g^{(1)}$ and $g^{(2)}$ that appear in (2.4). While we do not have a proof of this claim, we believe that the AdS/CFT conjecture makes it very plausible.

For example, let us say that it was possible to replace the functions $g^{(1)}$ and $g^{(2)}$ with some other functions $g^{(1)'}$ and $g^{(2)'}$, which solve the wave equation in some other background, but continue to preserve the approximate locality of bulk correlators. This would imply that thermal correlation functions on the boundary can be smoothly extended into local bulk correlation functions on a spacetime that is distinct from the AdS-Schwarzschild geometry. In turn, this would imply that a perturbative theory on the AdS-Schwarzschild geometry is dual to a perturbative theory on another spacetime. While such cases can be found at tree-level in the $\frac{1}{N}$ expansion, as for example in the conformal gravity analysis of [19], we believe that it is highly unlikely that such a duality would continue to hold at higher orders in the $\frac{1}{N}$ expansion.

However, the reader should note that the existence of the mirror operators, by itself, is not sufficient to guarantee the presence of a horizon in the bulk. This is not surprising if one thinks of the example of the thermofield doubled state of [20], but at a temperature lower than the Hawking-Page transition temperature. In this case, the dual geometry is not an eternal black hole but two disconnected AdS geometries that are not connected by a macroscopic wormhole. Thus the mere existence of entanglement between the two copies of the CFT is not sufficient to guarantee the presence of a horizon. The mirror operators that we will describe below, are, in a certain sense, like the operators in the second CFT.(However, we will also point out some important differences below.) Therefore while their existence is crucial for the existence of a horizon, it is not sufficient. We expect a horizon to appear only when the CFT is in a state dual to a black hole.

$\frac{1}{N}$ *corrections.*—We should point out that the status of the condition (2.7) [or equivalently (2.6)] is quite different from that of (2.5) with respect to $\frac{1}{\mathcal{N}}$ corrections. When these are included, we would like (2.5) to continue to hold at all orders in the $\frac{1}{\mathcal{N}}$ expansion and its violations, if any, should be suppressed exponentially in $\mathcal{N}$. On the other hand (2.6) can receive corrections at the first subleading order in $\frac{1}{\mathcal{N}}$. We can see that such corrections will come about, purely because of differences between correlators in the state $|\Psi\rangle$ and the thermal state. Another source of $\frac{1}{\mathcal{N}}$ corrections comes from interactions in the CFT which, in the bulk, corresponds to the backreaction of the Hawking radiation on the background geometry.

Charged states.—In writing (2.6) we have tacitly assumed that the state $|\Psi\rangle$ does not have any charge. In fact, the CFT contains several conserved charges, which we will generically call $\hat{Q}$. Just as we can associate a temperature β^{-1} with the state $|\Psi\rangle$ using correlation functions (or the growth in entropy with energy), we can also associate a chemical potential μ with a charged state.

In such a state, we need to modify (2.1) to

$$\begin{aligned}\langle\Psi|\mathcal{O}^{i_1}(t_1,\Omega_1)\ldots\mathcal{O}^{i_n}(t_n,\Omega_n)|\Psi\rangle \\ = Z^{-1}_{\beta,\mu}\mathrm{Tr}(e^{-\beta H-\mu\hat{Q}}\mathcal{O}^{i_1}(t_1,\Omega_1)\ldots\mathcal{O}^{i_n}(t_n,\Omega_n)),\end{aligned}$$

with the same modification in subsequent equations.

In this paper to lighten the notation, we will not write the charge $\hat{Q}$ explicitly. But the reader should note that our

[4] In Sec. VII of our previous paper [1], we also showed that this computation was stable under $\frac{1}{\mathcal{N}}$ corrections. More precisely, $\frac{1}{\mathcal{N}}$ corrections on the boundary do not "blow up" at the horizon.

entire analysis below goes through with the replacement of $\beta H \longrightarrow \beta H + \mu \hat{Q}$.

A. Comparison with flat-space black holes

We briefly mention why these mirror operators are also important in the context of flat-space black holes. The modes in the background of a flat-space black hole have a slightly different structure. Roughly speaking, we can divide the modes into those that are "ingoing" and "outgoing" near the horizon of the black hole, and those that are ingoing and outgoing at infinity.

For the familiar case of a scalar field ϕ propagating in the 4-dimensional Schwarzschild black hole of mass M, we can make this precise by introducing tortoise coordinates $r_* = r + 2M \ln \frac{r-2M}{2M}$ outside the horizon, and by introducing a second Schwarzschild patch just behind the horizon. Effective field theory tells us that, in the free-field limit, near the horizon, and at infinity, we can write

$$\phi(r_*, t) = \sum_{l,m} \int \frac{d\omega}{2\pi\sqrt{\omega}} (a_{\omega,l,m} e^{i\omega(r_*-t)} + b_{\omega,l,m} e^{-i\omega(r_*+t)}) \times Y_{l,m}(\theta,\phi) + \text{H.c.}, \quad \text{just outside}$$

$$\phi(r_*, t) = \sum_{l,m} \int \frac{d\omega}{2\pi\sqrt{\omega}} (a_{\omega,l,m} e^{i\omega(r_*-t)} + \tilde{a}_{\omega,l,m} e^{i\omega(r_*+t)}) \times Y_{l,m}(\theta,\phi) + \text{H.c.}, \quad \text{just inside}$$

$$\phi(r, t) = \sum_{l,m} \int \frac{d\omega}{2\,\pi r\sqrt{\omega}} (c_{\omega,l,m} e^{i\omega(r-t)} + d_{\omega,l,m} e^{-i\omega(r+t)}) \times Y_{l,m}(\theta,\phi) + \text{H.c.}, \quad \text{at } r \to \infty,$$

where "just inside" and "just outside" refers to just inside or outside the horizon. We have taken the field to be massless, which allows both ingoing and outgoing modes to exist at infinity for all frequencies. Note the presence of the potential barrier between $r = \infty$ and $r = 2M$ implies that the oscillators d and a commute whereas the pairs a, b, and c, d have nontrivial commutators. Starting with the Schwarzschild vacuum, which is defined by

$$a_\omega |S\rangle = d_\omega |S\rangle = \tilde{a}_\omega |S\rangle = 0,$$

the Unruh vacuum is defined by allowing the ingoing modes at infinity to remain in their ground state and by entangling the outgoing modes at the horizon with their corresponding tilde partners in a thermofield doubled state

$$|U\rangle = e^{\int e^{-\frac{\beta\omega}{2}} a_\omega^\dagger \tilde{a}_\omega^\dagger d\omega} |S\rangle, \tag{2.8}$$

which leads to $\langle U| a_{\omega'}^\dagger a_\omega |U\rangle = \frac{e^{-\beta\omega}}{1-e^{-\beta\omega}} \delta(\omega - \omega')$.

It is in the Unruh vacuum, that the horizon is smooth. Any significant deviations from this vacuum will generically lead to a firewall. It is clear from (2.8) that

$$\tilde{a}_\omega |U\rangle = e^{-\frac{\beta\omega}{2}} a_\omega^\dagger |U\rangle; \qquad \tilde{a}_\omega^\dagger |U\rangle = e^{\frac{\beta\omega}{2}} a_\omega |U\rangle. \tag{2.9}$$

instead of starting with the definition (2.8) and deriving (2.9), one could also reverse the logic. One can show [21] by a consideration of the expectation value of the stress tensor that the state corresponding to a smooth horizon satisfies (2.9). From here, one can derive (2.8).

Now, we do not have a precise understanding of the Hilbert space of quantum gravity in flat space. However, we believe that it should have a large number of microstates, all of which have the same macroscopic properties as the Unruh vacuum. In particular, this implies that in a microstate corresponding to the black hole we should be able to find operators that meet (2.9). Hence, we see that we require operators that satisfy the properties (2.5) and (2.7) for flat-space black holes as well to obtain a smooth horizon.

B. Summary

In this section, we have tried to argue that the issue of whether the horizon of the black hole is smooth or not has to do with the issue of whether we can find operators in the CFT that satisfy (2.5) and (2.7). *All* the recent discussions of the information paradox can, essentially, be phrased as questions about whether such operators exist. We will make this more clear when we discuss these arguments below. In the next section, we describe how to find operators that satisfy these properties.

We should mention that, in the argument above, we have pointed out the necessity of the mirror operators for generalized free fields in the CFT that enter the modes of perturbative bulk fields. However, we will actually succeed in finding mirror operators, for observables in a large class of statistical-mechanics systems. In the case of the CFT, we will succeed in "doubling" not only the generalized free fields but a much larger class of operators.

We should point out that there are powerful (although, in our opinion, not conclusive) arguments that suggest that one cannot find *fixed* (i.e. state-independent) operators that have the correct behavior specified by (2.5) or (2.6) [or (2.7)] for an arbitrary given state $|\Psi\rangle$. However, if we allow the mapping between CFT operators and local bulk operators to depend on the state itself, then one can indeed find such operators as we show explicitly below.

Moreover, these operators then resolve *all* the recent paradoxes that have been formulated to suggest the presence of a structure at the horizon.

III. CONSTRUCTING THE OPERATORS BEHIND THE HORIZON

In this section, we will explicitly construct operators behind the horizon. We will perform this construction in three steps so as to make this section maximally pedagogical. We start with a description of our idea in a general setting. It is well known that given a limited set of observables, almost any pure state drawn from a large Hilbert space looks "thermal" or equivalently looks *as if* it is entangled with some environment. In the first part of this section, we show how, in this single Hilbert space, it is possible to construct operators that behave as if they were acting on the environmental degrees of freedom.

In fact, the operators behind the horizon that we have described above are precisely of this form. So, in the second and central part of this section, we go on to describe our construction of these operators in the CFT. This case comes with a few quirks, including the fact that the CFT has conserved charges, and so some properties such as the charge and energy of the mirror operators is still visible outside the horizon.

Finally, we descend from this complicated situation and discuss two toy models in detail. The first is a toy model of decoupled harmonic oscillators. This captures our ideas in a concrete setting, and has many of the essential features of the CFT, without some of the technical complications. The second is a simple spin chain, which is a popular model—and probably the simplest available one—for considering the information paradox. We describe how the mirror operators can be constructed in this setting as well.

The reader may choose to read this section in any order, or even jump directly to the toy models.

A. Defining mirror operators for a general theory

Let us say that we have some system, which is prepared in a pure state $|\Psi\rangle$ drawn from a large, but finite-dimensional, Hilbert space $\mathcal{H}$. We are able to probe the system with a restricted set of operators. Let us call the

$$\text{set of observables:}\ \mathcal{A} = \text{span}\{A_1, \ldots A_{\mathcal{D}_\mathcal{A}}\}. \tag{3.1}$$

As we have written explicitly above, $\mathcal{A}$ is a linear space and we can always take arbitrary linear combinations of operators in $\mathcal{A}$. However, it is important that $\mathcal{A}$ may not quite be an algebra. It may be possible to multiply two elements of $\mathcal{A}$ to obtain another operator that also belongs to $\mathcal{A}$. In fact, we will often discuss such products of operators below. However, we may not be allowed to take arbitrary products of operators in this set. In particular, if we try and take a product of $\mathcal{N}$ operators, it may take us out of the set $\mathcal{A}$.

We wish to consider states $|\Psi\rangle$ that satisfy the following very important property[5]:

$$A_p|\Psi\rangle \neq 0, \quad \forall\ A_p \in \mathcal{A}. \tag{3.2}$$

[5]Later, in the discussion on the CFT, we will consider situations where $|\Psi\rangle$ may be an eigenstate of a conserved charge, in which case (3.2) does not hold for certain operators but, for the current discussion, this is an unimportant technicality.

Note that this statement holds for all elements of $\mathcal{A}$, or equivalently for all possible linear combinations of the basis of observables written in (3.1). An immediate corollary of this statement is that the dimension of $\mathcal{A}$ be smaller than the dimension of the Hilbert space of the theory:

$$\mathcal{D}_\mathcal{A} \ll \dim(\mathcal{H}) \equiv \mathcal{D}_\mathcal{H}.$$

Equation (3.2) also means that if $|\Psi\rangle$ is a state of finite energy, then the energy of our probe operators in the algebra is also limited.

We wish to emphasize that these conditions on the observables we can measure and the state under consideration are physically very well motivated. For example, if the reader likes to think of a spin-chain system, then $\mathcal{A}$ could consist of all local operators—the Pauli spins on each site—bilocal operators—which comprise products of local operators at two sites—all the way up to K-local operators, as long as $K \ll \mathcal{N}$—the length of the spin chain. Generic states in the Hilbert space of the spin chain now satisfy (3.2). We work this spin chain example out explicitly in Sec. III D

However, more generally, as the reader can easily persuade herself, if we place a large system in a state that appears to be thermal, and consider some finite set of "macroscopic observables" (for example, those that obey the so-called "eigenstate thermalization hypothesis"), then the condition (3.2) is easily satisfied. In fact, we can consider a larger class of states, which are excitations of thermal states that are out of equilibrium.

Now, it is very well known that, given such a set of observables $\mathcal{A}$, and a pure state $|\Psi\rangle$, we can construct several density matrices ρ, corresponding to *mixed states*, which are indistinguishable from $|\Psi\rangle$, in the sense that we can arrange for

$$\text{Tr}(\rho A_p) = \langle\Psi|A_p|\Psi\rangle, \quad \forall\ A_p \in \mathcal{A}.$$

Such a density matrix is not unique but the correct way to pick it, assuming that the expectation values of $\langle A_p\rangle$ are all the *information* we have, is to pick the density matrix that maximizes the entropy: $S_{\text{th}} = \max\left(-\text{Tr}(\rho \ln \rho)\right)$ [22]. In fact, this maximum entropy S_{th} is what should correspond to the thermodynamic entropy of the system. For a generic state $|\Psi\rangle$, we expect to find

$$\rho \approx \frac{1}{Z} e^{-\beta H}, \tag{3.3}$$

up to $\frac{1}{\mathcal{N}}$ corrections, where Z is the partition function.[6]

[6]In an equilibrium state, in any case, we expect off-diagonal terms in the energy eigenbasis in the density matrix to be strongly suppressed, although the eigenvalues may be corrected from the canonical ones. For the significance of such corrections, see Appendix A, and for nonequilibrium states, see Sec. V.

It is also well known that the statements above imply that even though the system is in a pure state, it *appears* as if the system is entangled with some other heat bath. This pure state in the fictitious larger system is called the "purification" of ρ. This purification is not unique, even given ρ but given a generic state in which the density matrix is thermal as in (3.3), we will pick it to be the thermofield doubled state [23][7]

$$|\Psi\rangle_{\rm tfd} = \frac{1}{\sqrt{Z}}\sum_{E_i} e^{-\frac{\beta E_i}{2}}|E_i\rangle|\tilde{E}_i\rangle, \tag{3.4}$$

where the sum runs over all energy eigenvalues of the system. Note that the subscript tfd emphasizes that this state is *distinct* from the pure state $|\Psi\rangle$, and lives in a (fictitious) larger Hilbert space.

The new point that we want to make here is as follows. In the pure state $|\Psi\rangle$, we can also *effectively* construct the operators that act on the "other" side of the purification. So, for all practical purposes the thermofield doubled state and the doubled operators may be realized in the same Hilbert space.

More precisely, we want the following. For every operator acting on the Hilbert space of the system

$$A_p|E_i\rangle = (A_p)_{ij}|E_j\rangle, \tag{3.5}$$

we have an analogous operator that acts on the fictitious environment

$$A_p^{\rm tfd}|\tilde{E}_i\rangle = (A_p)^*_{ij}|\tilde{E}_j\rangle.$$

The complex conjugation is necessary to ensure that this map remains invariant if we, for example, decide to rephase the energy eigenstates of the system by $e^{i\phi_i}$ and those of the environment by $e^{-i\phi_i}$ under which the state (3.4) is obviously invariant.

The operator $A_p^{\rm tfd}$ has two other important properties. First, it clearly commutes with the operators A_m, since these act on different spaces

$$[A_p^{\rm tfd}, A_m]|\Psi\rangle_{\rm tfd} = 0, \quad \forall\ p, m. \tag{3.6}$$

Second, with some simple algebra (see Appendix A) we can see that

$$A_p^{\rm tfd}|\Psi\rangle_{\rm tfd} = e^{-\frac{\beta H}{2}} A_p^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle_{\rm tfd}. \tag{3.7}$$

We now desire the existence of operator $\tilde{A}_p$ that acts in the single Hilbert space $\mathcal{H}$ and *mimics* the action of (3.7) and (3.6) while acting on the state $|\Psi\rangle$. Naively, this may seem impossible. For example, if we consider a spin chain and the set $\mathcal{A}$ comprises the set of Pauli-matrices acting on each site, then there is no operator in the Hilbert space that commutes with all the $A_p \in \mathcal{A}$.

However, as we describe here, given a state $|\Psi\rangle$, there is an elegant and almost unbelievably simple definition of these operators. First, we need to expand the set of observables $\mathcal{A}$ a little so that for each $A_p \in \mathcal{A}$, we adjoin to $\mathcal{A}$ the element $\hat{A}_p = e^{-\frac{\beta H}{2}} A_p^\dagger e^{\frac{\beta H}{2}}$. Next, as we mentioned above, while $\mathcal{A}$ may not be closed under the multiplication of arbitrary pairs, if the product $A_{p_1}A_{p_2} \in \mathcal{A}$, we may also want to include the products $\hat{A}_{p_1}A_{p_2}$ and $A_{p_1}\hat{A}_{p_2}$. We will call this expanded set of observables $\mathcal{A}^{\rm exp}$. If $\mathcal{D}_{\mathcal{A}} \ll \mathcal{N}$, then the elements of this expanded set also satisfy (3.2).

We want to emphasize that the reader should not get lost in the technicalities of this "expanded" set. In fact, in the interesting case of the CFT below, we will see that $\mathcal{A}^{\rm exp}$ coincides with $\mathcal{A}$. This is because in the situation where the A_p have some definite energy ω_p, these factors simply invert the energy, and insert a factor of $e^{-\frac{\beta\omega_p}{2}}$.

Now, we simply define the mirror operators by the following set of linear equations

$$\begin{aligned} \tilde{A}_p|\Psi\rangle &= e^{-\frac{\beta H}{2}} A_p^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle, \\ \tilde{A}_p A_m|\Psi\rangle &= A_m \tilde{A}_p|\Psi\rangle, \end{aligned} \tag{3.8}$$

where $A_p, A_m \in \mathcal{A}^{\rm exp}$. In a given state $|\Psi\rangle$, these two lines together just correspond to $\dim(\mathcal{A}^{\rm exp})$ equations. Note, that we can write these two lines as the single compact equation

$$\tilde{A}_p A_m|\Psi\rangle = A_m e^{-\frac{\beta H}{2}} A_p^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle, \tag{3.9}$$

but we have written them separately because, as will become clear below, the two lines of (3.8) have different physical interpretations.

Note that $\tilde{A}_p$ are linear operators in a Hilbert space of dimension $\mathcal{D}_{\mathcal{H}}$ that we are interested in. Equation (3.9) makes it clear that we are specifying the action of these operators on a linear subspace, $\mathcal{H}_\psi = \mathcal{A}^{\rm exp}|\Psi\rangle$, produced by acting with all elements of the set $\mathcal{A}^{\rm exp}$ on the set $|\Psi\rangle$. Equivalently, we are specifying the action of $\tilde{A}_p$ on $\mathcal{D}_{\mathcal{H}_\psi} = \dim(\mathcal{H}_\psi)$ basis vectors. It is *always* possible to specify the action of an operator on a set of linearly independent vectors that is smaller in size than $\mathcal{D}_{\mathcal{H}}$.

So, the only constraint we have to check is that the vectors $\mathcal{A}_p|\Psi\rangle$ produced by acting on $|\Psi\rangle$ are linearly independent i.e. that we cannot find some coefficients $\sum_p \alpha_p \mathcal{A}_p|\Psi\rangle = 0$. However, (3.2) tells us that there is no such linear combination.

So, we conclude that, provided (3.2) is met, we can *always* find an operator $\tilde{A}_p$ that satisfies (3.8). In fact, it is easy to write down an explicit formula for this operator. Consider the set of vectors

[7] In Appendix A, we discuss other choices of the purification which are, in fact, required at $\frac{1}{N}$ and this issue of the lack of uniqueness.

$$|v_m\rangle = A_m|\Psi\rangle; \qquad |u_m\rangle = A_m e^{\frac{-\beta H}{2}} A_p^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle,$$

where $m = 1 \ldots \dim(\mathcal{A}^{\exp})$ and the operators run over *any* basis of the set $\mathcal{A}^{\exp}$. Now, define the "metric"

$$g_{mn} = \langle v_m|v_n\rangle,$$

and its inverse g^{mn} satisfying $g^{mn}g_{np} = \delta_p^m$. This inverse necessarily exists, because the $|v_m\rangle$ are linearly independent by the conditions above. Now, an operator $\tilde{A}_p$ that satisfies the condition (3.8) above is given by

$$\tilde{A}_p = g^{mn}|u_m\rangle\langle v_n|, \tag{3.10}$$

where the repeated indices are summed, as usual. Of course, the operator $\tilde{A}_p + \tilde{A}^{\rm orth}$, where $\tilde{A}^{\rm orth}$ is any operator that satisfies $\tilde{A}^{\rm orth}|v_m\rangle = 0$, $\forall\, m$ also satisfies (3.8). In (3.10), we have simply taken $\tilde{A}^{\rm orth} = 0$, but this ambiguity is physically irrelevant.

Furthermore, note that the rules (3.8) also allow us to build up the action of products of the mirror operators *recursively*. For example, notice that these rules lead to

$$\tilde{A}_{p_1}\tilde{A}_{p_2}|\Psi\rangle = \tilde{A}_{p_1} e^{-\frac{\beta H}{2}} A_{p_2}^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle = e^{-\frac{\beta H}{2}} A_{p_2}^\dagger e^{\frac{\beta H}{2}} \tilde{A}_{p_1}|\Psi\rangle$$
$$= e^{-\frac{\beta H}{2}} A_{p_2}^\dagger A_{p_1}^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle.$$

Here in the first equality, we use the first rule of (3.8). In the next equality we use the second rule to commute $\tilde{A}_{p_2}$ to the right, and then we use the first rule again to obtain our final expression. Notice in particular that

$$\tilde{A}_{p_1}\tilde{A}_{p_2}|\Psi\rangle = (\widetilde{A_{p_1}A_{p_2}})|\Psi\rangle.$$

Next, note that the rules III A lead to the result that *acting on the state* $|\Psi\rangle$, the mirror operators commute with the ordinary operators. For example, consider the commutator of an ordinary and mirror operator within some product of ordinary and mirror operators acting on $|\Psi\rangle$

$$\tilde{A}_{p_1}A_{p_2}\ldots[\tilde{A}_{p_m}, A_{p_{m+1}}]\ldots\tilde{A}_{p_{n-1}}A_{p_n}|\Psi\rangle$$
$$= \tilde{A}_{p_1}A_{p_2}\ldots(\tilde{A}_{p_m}A_{p_{m+1}} - A_{p_{m+1}}\tilde{A}_{p_m})\ldots\tilde{A}_{p_{n-1}}A_{p_n}|\Psi\rangle$$
$$= \tilde{A}_{p_1}A_{p_2}\ldots A_{p_{m+1}}\ldots\tilde{A}_{p_m}\tilde{A}_{p_{n-1}}A_{p_n}|\Psi\rangle -$$
$$\times\, \tilde{A}_{p_1}A_{p_2}\ldots A_{p_{m+1}}\ldots\tilde{A}_{p_m}\tilde{A}_{p_{n-1}}A_{p_n}|\Psi\rangle = 0. \tag{3.11}$$

Here, the key point is that the second line of (3.8) allows us to move $\tilde{A}_{p_m}$ through $A_{p_{m+1}}$ and any other occurrences of A_p operators till the first occurrence of another $\tilde{A}_p$ operator. In writing these equations, we have tacitly assumed that we can take the product of the operators A_{p_i}, while remaining within the set $\mathcal{A}$. This is justified as long as $n \ll \mathcal{D}_{\mathcal{A}}$.

Now, we make a few remarks about correlation functions. First, note that by construction we have ${}_{\rm tfd}\langle\Psi|A_p|\Psi\rangle_{\rm tfd} = \langle\Psi|A_p|\Psi\rangle$, $\quad\forall\, A_p \in \mathcal{A}^{\exp}$. Within mixed correlators involving both A_p and $\tilde{A}_p$, we see that we have the following properties:

$$\langle\Psi|\tilde{A}_{p_1}\ldots\tilde{A}_{p_m}A_{p_{m+1}}\ldots A_{p_n}|\Psi\rangle$$
$$= {}_{\rm tfd}\langle\Psi|A_{p_1}^{\rm tfd}\ldots A_{p_m}^{\rm tfd}A_{p_{m+1}}\ldots A_{p_n}|\Psi\rangle_{\rm tfd}.$$

To show this involves only a small amount of additional work. First, we see that

$$\langle\Psi|\tilde{A}_{p_1}\ldots\tilde{A}_{p_m}A_{p_{m+1}}\ldots A_{p_n}|\Psi\rangle$$
$$= \langle\Psi|\tilde{A}_{p_1}\ldots\tilde{A}_{p_{m-1}}A_{p_{m+1}}\ldots A_{p_n}e^{-\frac{\beta H}{2}}A_{p_m}^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle, \tag{3.12}$$

where we have used the second line of (3.8) to move the $\tilde{A}_{p_m}$ to the right, and then used the first line to substitute its action on $|\Psi\rangle$. Now, given the right-hand side of (3.12), we can use the same procedure to move $\tilde{A}_{p_{m-1}}$ to the extreme right and then substitute for its action. Continuing this, we see that finally

$$\langle\Psi|\tilde{A}_{p_1}\ldots\tilde{A}_{p_m}A_{p_{m+1}}\ldots A_{p_n}|\Psi\rangle$$
$$= \langle\Psi|A_{p_{m+1}}\ldots A_{p_n}e^{-\frac{\beta H}{2}}A_{p_m}^\dagger\ldots A_{p_1}^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle.$$

Now, as we discussed above, correlators of ordinary operators in the set $\mathcal{A}^{\rm ext}$ in the state $|\Psi\rangle$ are the same as those in the thermofield doubled state. So, we find that

$$\langle\Psi|\tilde{A}_{p_1}\ldots\tilde{A}_{p_m}A_{p_{m+1}}\ldots A_{p_n}|\Psi\rangle$$
$$= {}_{\rm tfd}\langle\Psi|A_{p_{m+1}}\ldots A_{p_n}e^{-\frac{\beta H}{2}}A_{p_m}^\dagger\ldots A_{p_1}^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle_{\rm tfd}$$
$$= {}_{\rm tfd}\langle\Psi|A_{p_{m+1}}\ldots A_{p_n}A_{p_1}^{\rm tfd}\ldots A_{p_m}^{\rm tfd}|\Psi\rangle_{\rm tfd},$$

where the reader can easily use the property (3.7) to verify the second equality.

The $\doteq$ *notation.*—This feature, where the properties of the $\tilde{A}_p$ operators hold only within correlation functions evaluated on a particular state is important enough that we will introduce some special notation for it, which we have already used above, and will use extensively later. We will write

$$[\tilde{A}_{p_m}, A_{p_{m+1}}] \doteq 0,$$

to indicate that (3.11) holds, but the operators $\tilde{A}_{p_m}$ and $A_{p_{m+1}}$ may *not commute* as operators. It is just that this commutator annihilates $|\Psi\rangle$ and its descendants produced by acting with elements of the algebra A_p.

The space $\mathcal{H}_\Psi$.— Before we conclude this subsection, let us make a comment about solving the linear equations (3.8). We have carefully argued above that it is possible to find a set of solutions to these equations. In constructing such solutions, we do not even actually need to consider the full vector space $\mathcal{H}$. In fact, it is convenient to consider a slightly smaller vector space

$$\mathcal{H}_\psi = \mathcal{A}^{\text{exp}}|\Psi\rangle,$$

which is just the space formed by the action of the set $\mathcal{A}^{\text{exp}}$ on the state $|\Psi\rangle$. In all cases of interest that we will study below, $\mathcal{A}^{\text{exp}}$ coincides with $\mathcal{A}$, and in these cases we can also write $\mathcal{H}_\Psi = \mathcal{A}|\Psi\rangle$. We see (3.8) is a statement about the action of the operators $\tilde{A}_p$ on the domain $\mathcal{H}_\Psi$ and the action of these operators outside this space is unspecified. In fact, we could even choose $\tilde{A}_p$ to annihilate states in the space of vectors orthogonal to $\mathcal{H}_\Psi$ without affecting low-point correlators. Note that the definition (3.8), and the fact that $\mathcal{A}$ may not be closed under arbitrary pairwise multiplication, implies that the *range* of $\tilde{A}_p$ may differ slightly from $\mathcal{H}_\Psi$, even in this case. These "edge effects" are usually unimportant, and the physically relevant subspace is $\mathcal{H}_\Psi$.

Our construction, as we have presented it here, applies to any statistical-mechanics system. We now specialize to the CFT which, as we will see, has a few new ingredients.

B. Mirror operators in the CFT

We now discuss the construction of the tilde operators in an interacting CFT. Our construction follows the general method that we outlined above, but this section is written so as to be self-contained. We will find two new features in the CFT. One is technical and, in our view, not so important: we have to regularize the modes of the CFT to obtain a finite set of observables $\mathcal{A}$. The second is also somewhat technical, but a little more interesting. The operators that we are constructing are not gauge invariant with respect to bulk diffeomorphisms and conserved charges. So, while they commute exactly with almost all operators, within correlation functions, they do not commute with the global charges or the Hamiltonian.

To be concrete, we will consider a CFT on $S^{d-1}\times R$. The black hole is dual to a state $|\Psi\rangle$ in the CFT, with an energy that is much larger than, but of the same order as, $\mathcal{N}$. In this section, we will show how to construct the tildes on this state $|\Psi\rangle$.

1. Regularizing the space of operators

First, let us discuss the operators that we can use to probe the black hole geometry—this is the set $\mathcal{A}$ above. We have some number of light operators in the CFT that correspond to the supergravity fields. In addition, we could probe the black hole geometry with excitations corresponding to stringy states, and perhaps even with brane probes. In the CFT, all of these can be represented by conformal primary operators with a dimension that is much smaller than $\mathcal{N}$. We remind the reader that $\mathcal{N}$ is the central charge. So, in maximally supersymmetric $SU(N)$ theory, $\mathcal{N}\propto N^2$ and even a giant graviton operator has dimension $\Delta = N \ll \mathcal{N}$.

It will be convenient for us to discuss the modes of these operators, which are defined by

$$\mathcal{O}^i_{\omega,\boldsymbol{m}} = \int \mathcal{O}^i(t,\Omega)e^{i\omega t}Y_{\boldsymbol{m}}(\Omega)d^{d-1}\Omega dt,$$

where $Y_{\boldsymbol{m}}$ is the spherical harmonic indexed by the $d-1$ integers in the array $\boldsymbol{m}$.

Now, the relevant spacing of the energy levels around energies of order $\mathcal{N}$ is actually $e^{-S}\sim e^{-\mathcal{N}}$. So, the spectrum of modes of low-dimensional conformal primaries is almost continuous even when the CFT is on a sphere.

Now, consider two energy levels $|E\rangle$ and $|E+\delta_\omega\rangle$. We can consider the precise mode $\mathcal{O}_{\delta\omega,\boldsymbol{m}}$ that causes transitions between these levels. However, if the differences between energies are nondegenerate, as we expect on general grounds for a "chaotic" system, then this mode will have a zero matrix element between *any other states.*

So, we need to "coarse grain" these modes a little to come up with a useful set of operators. We will do this, by introducing a lowest infrared frequency ω_{min}, and bin together the modes of $\mathcal{O}^i$ in bins of this width. More precisely, we define

$$\mathcal{O}^i_{n,\boldsymbol{m}} = \frac{1}{(\omega_{\text{min}})^{1/2}}\int_{n\omega_{\text{min}}}^{(n+1)\omega_{\text{min}}}\mathcal{O}^i_{\omega,\boldsymbol{m}}d\omega. \tag{3.13}$$

These regularized modes $\mathcal{O}^i_{n,\boldsymbol{m}}$ have a smooth behavior in the Hilbert space, and we might reasonably expect them to obey the eigenstate thermalization hypothesis (ETH), as we show in more detail in Sec. V. We will often use

$$\omega_n = n\omega_{\text{min}},$$

and correspondingly also write $\mathcal{O}^i_{\omega_n,\boldsymbol{m}}$.

We can take ω_{min} to go to zero faster than any power of $\mathcal{N}$, but it must be much larger than $e^{-\mathcal{N}}$. So, for example, we could take $\omega_{\text{min}} = e^{-\sqrt{\mathcal{N}}}$. So, the reader may wish to think of the $SU(N)$ theory, with an infrared cutoff that scales like e^{-N}. This is certainly adequate for all purposes of constructing perturbative fields in the interior.

We have now regulated both the maximum dimension of allowed probe operators, and their modes in the manner above. Let us call these various operators $\mathcal{O}^i_{n,\boldsymbol{m}}$ where i refers to the conformal primary, and $n,\boldsymbol{m}$ specify the mode. We now consider the set formed by taking the span of arbitrary products of up to K numbers of these operators

$$\mathcal{A} = \text{span}\{\mathcal{O}^i_{n,\boldsymbol{m}}, \mathcal{O}^{i_1}_{n_1,\boldsymbol{m}_1}\mathcal{O}^{i_2}_{n_2,\boldsymbol{m}_2},\ldots,\mathcal{O}^{i_1}_{n_1,\boldsymbol{m}_1}\mathcal{O}^{i_2}_{n_2,\boldsymbol{m}_2}\ldots\mathcal{O}^{i_k}_{n_K,\boldsymbol{m}_K}\}.$$

The set $\mathcal{A}$ is limited by the constraint that each product occurring in $\mathcal{A}$ satisfies $\omega_{\text{min}}\sum_{i=1}^K n_i \ll \mathcal{N}$ which limits the total energy that can appear in this set.

Note, that as we emphasized in [12], taking the linear span of the products of operators above is exactly the same as thinking of $\mathcal{A}$ as the set of all *polynomials* in the modes of the operators $\mathcal{O}^i$

$$A_\alpha = \sum_N \alpha(N)(\mathcal{O}^i_{n,\boldsymbol{m}})^{N(i,n,\boldsymbol{m})},$$

with the constraint that

$$\sum_{i,n,\boldsymbol{m}} N(i,n,\boldsymbol{m})\omega_{\min} n \leq E_{\max} \ll \mathcal{N}. \qquad (3.14)$$

We also require that the set cannot be too large:

$$\mathcal{D}_{\mathcal{A}} = \dim(\mathcal{A}) \ll e^{\mathcal{N}}. \qquad (3.15)$$

The second constraint is automatically satisfied if we also limit the number of insertions in the polynomials

$$\sum_{i,n,\boldsymbol{m}} N(i,n,m) \leq K_{\max},$$

and do not take $E_{\max}$ to be too large. In fact, there is an interplay between the value of $K_{\max}$, $E_{\max}$, and $\omega_{\min}$ so that (3.15) can be preserved. For example, if we take $\omega_{\min} = e^{-\sqrt{\mathcal{N}}}$, then we must take $K_{\max} \ll \sqrt{\mathcal{N}}$ in order to preserve (3.15). If we take $\omega_{\min}$ to scale just as an inverse power of $\mathcal{N}$, we can take $K_{\max}$ to be larger.

Note that these polynomials, are polynomials in *non-commutative* variables, since the operators do not commute with one another. However, there may be operator relations within the CFT, and as a result it may happen that some particular set of polynomials vanish because of these relations. In taking the set of polynomials above, we must mod out by these relations. For example if for the three operators that appear above, $\mathcal{O}^{i_1}_{n_1,\boldsymbol{m}_1}\mathcal{O}^{i_2}_{n_2,\boldsymbol{m}_2} = \mathcal{O}^{i_3}_{n_3,\boldsymbol{m}_3}$, then, the polynomial $(\mathcal{O}^{i_1}_{n_1,\boldsymbol{m}_1})^2\mathcal{O}^{i_2}_{n_2,\boldsymbol{m}_2}$ must clearly be identified with the polynomial $\mathcal{O}^{i_1}_{n_1,\boldsymbol{m}_1}\mathcal{O}^{i_3}_{n_3,\boldsymbol{m}_3}$.

This set $\mathcal{A}$ consists of all possible probes that we are allowed to make in the black hole geometry. We emphasize that the set of operators in $\mathcal{A}$ is essentially the largest set of operators, for which one might hope to make sense of a semiclassical geometry. For example, if we start including products of up to $\mathcal{N}$ of the conformal primary modes, then there is no reason at all that expectation values of such operators should be reproducible by calculations in a semiclassical geometry.

In this concrete setting, the reader can also see another feature that we discussed in the section above. The set $\mathcal{A}$ is not quite an algebra because of the cutoff (3.14) that has been imposed on the energy of the operators that can appear. On the other hand, it is often possible to multiply elements of $\mathcal{A}$ together to obtain another member of $\mathcal{A}$.

Before we proceed to the definition of the mirror operators, we must impose a final technical constraint on the set $\mathcal{A}$. We do not take the Hamiltonian itself, or any conserved charge (by which we mean any operator, which commutes with the Hamiltonian) to be part of this set. This is equivalent to excluding the *zero modes* of conserved currents. These zero modes to not correspond to propagating degrees of freedom in the bulk and, in any case, we will deal with them separately below.

2. Defining the mirror operators

We now describe how to define the mirror operators. The CFT in a generic thermal state has the following property:

$$A_p|\Psi\rangle \neq 0, \quad \forall\, A_p \in \mathcal{A}. \qquad (3.16)$$

This is simply the statement that the insertion of a small number of light operators cannot annihilate the generic thermal state. We will work with states that satisfy (3.16). States that do not satisfy this condition are a measure-0 subset of the set of all states, and as we discuss below, they may not have a smooth horizon.

We will now define the tilde operators, by specializing the rules that we gave above. The mirror operators are defined by two very simple rules:

$$\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}|\Psi\rangle = e^{-\frac{\beta\omega_n}{2}}(\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger|\Psi\rangle, \qquad (3.17)$$

$$\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}A_p|\Psi\rangle = A_p\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}|\Psi\rangle, \quad \forall\, A_p \in \mathcal{A}. \qquad (3.18)$$

As advertised, we do not need to expand the set of allowed observables $\mathcal{A}$ to $\mathcal{A}^{\text{exp}}$ in the CFT to define the mirror operators.

Note that (3.17) and (3.18) together give us $\mathcal{D}_{\mathcal{A}}$ *linear equations* for the $\tilde{\mathcal{O}}$. However, $\tilde{\mathcal{O}}$ can operate in a space that is $e^{\mathcal{N}}$ dimensional. These equations are all internally consistent because of the condition (3.16). So, there are many possible solutions to these constraints. One explicit solution is shown in (3.10).

All these solutions are equivalent for our purposes, since they do not show any difference at all, except when inserted in very high-point correlators. As we pointed out above, there is also an, in principle, difference between (3.17) and (3.18). While (3.17) needs to be corrected order by order in $\frac{1}{\mathcal{N}}$, (3.18) is already correct at all orders in the $\frac{1}{\mathcal{N}}$ for the correlators that we are interested in.

3. Choice of gauge: Hamiltonian and Abelian charges

We now turn to the issue of a choice of gauge. We are willing to consider cases, where $|\Psi\rangle$ is an energy eigenstate, and certainly it may be possible to put $|\Psi\rangle$ in an eigenstate of some other conserved charge. We first discuss the inclusion of the Hamiltonian, which corresponds to zero modes of the stress tensor, and other Abelian charges, then turn to other kinds of conserved charges including non-Abelian charges in the next subsection.

If $|\Psi\rangle$ is an energy eigenstate, or the eigenstate of some other charge, we still expect it to appear thermal. However, in such cases, we see that we might have

$$(\hat{Q} - Q)|\Psi\rangle = 0,$$

where $\hat{Q}$ is the charge operator and Q is the corresponding eigenvalue. This is the reason that we cannot include $\hat{Q}$ in the set $\mathcal{A}$. If, with this inclusion, we were to also demand (3.18), we would get an inconsistency.

However, this is quite simple to fix. We set $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ to have a nonzero commutator with the *zero mode* of the corresponding conserved current. In fact, this zero mode is not of any interest, except for the fact that it includes the charge itself. So, we append the charge to the set $\mathcal{A}$ and add an additional rule to the set of rules above.

First, since the position space operator $O^i(t,\Omega)$ is Hermitian, we need to reorganize its modes $\mathcal{O}^i_{n,\boldsymbol{m}}$ into operators that transform simply under the charger under consideration. If this charge is just the Hamiltonian or the angular momentum on S^{d-1}, then the modes already transform in a simple manner. But, in any case, we can construct linear combinations $\mathcal{O}^{i,q}_{n,\boldsymbol{m}}$, which have a well-defined charge so that $[\hat{Q},\mathcal{O}^{i,q}_{n,\boldsymbol{m}}] = q\mathcal{O}^{i,q}_{n,\boldsymbol{m}}$. The action of the mirror operators on the original linear combinations can be constructed by using the antilinearity of the mirror map. We now add the following rule to the set of rules above

$$\tilde{\mathcal{O}}^{i,q}_{n,\boldsymbol{m}} A_1\hat{Q}A_2|\Psi\rangle = A_1\hat{Q}A_2\tilde{\mathcal{O}}^{i,q}_{n,\boldsymbol{m}}|\Psi\rangle + qA_1A_2\tilde{\mathcal{O}}^{i,q}_{n,\boldsymbol{m}}|\Psi\rangle. \tag{3.19}$$

In Appendix B, we discuss this issue further. We show how a choice of gauge results in these commutation relations, and how they may be interpreted in terms of Wilson lines. We also explore the fact that these relations already seem to lead to some interesting physical implications. We note that by virtue of this rule we see that $\tilde{\mathcal{O}}^i$ does not really correspond to a local field on the boundary, since such a field would have nonzero commutators for other modes of the current as well. Here, this is not a difficulty, since the bulk fields constructed from $\tilde{\mathcal{O}}^i$ cannot ever be taken close to the boundary to obtain any kind of contradiction. But this also provides a criterion for when the $\tilde{\mathcal{O}}^i$ fields can enter bulk operators, and it explains why they cannot be used in bulk fields below the Hawking Page transition.

Second, notice that since the charge and energy of the $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ can be measured by the CFT Hamiltonian, this tells us that there is not really any "other side" of the collapsing geometry. We return to this at greater length in Appendix B.

4. Non-Abelian charges

We now describe how the mirror operators act on descendants of the state $|\Psi\rangle$ produced by acting with various non-Abelian charges. The main difference with the analysis for the Hamiltonian and Abelian charges above is that in this case, we can have other kinds of null vectors. The analysis of the subsection above is subsumed in the more general analysis of this subsection.

For example, we might want to consider a Schwarzschild black hole, and consider a corresponding ensemble in the CFT, where the states transform in a small representation of some non-Abelian charge, but are yet not charge eigenstates. Now, we may have $J^K_+|\Psi\rangle = 0$, for some "raising operator" J_+. We wish to ensure that our definition of the $\tilde{\mathcal{O}}$ operators is correct in this case. Below, we will denote any *polynomial* in the charges by $\mathcal{Q}_\alpha$. The space of physical states is produced by acting with all such polynomials on the base state $|\Psi\rangle$, and then modding out by the null vectors. The action of $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ must be correct on this quotient space, in that it must annihilate all null vectors.

The set of null vectors.—First, the condition that the action by an observable does not annihilate the state must be refined in the presence of such charges. We will impose the following condition. Consider a set of charge polynomials $\mathcal{Q}_{\alpha_1}\ldots\mathcal{Q}_{\alpha_m}$. Now, we demand

$$\sum_{i=1}^{m}\kappa_i\mathcal{Q}_{\alpha_i}|\Psi\rangle \neq 0, \quad \forall\ \kappa_i \Rightarrow \sum_{i=1}^{m}A_{\beta_i}\mathcal{Q}_{\alpha_i}|\Psi\rangle \neq 0, \quad \forall\ A_{\beta_i}. \tag{3.20}$$

Translated into words, this means that we get various "descendants" by acting on the base state with the charges. If these descendants are linearly independent, then by acting on them with our observables, we cannot "make" them linearly dependent. This is a very natural generation of (3.16) above, and more formally speaking the states that do *not* satisfy (3.20) form a measure-0 space in the Hilbert space. Of course, we can also phrase (3.20) as

$$\sum_{i=1}^{m}A_{\beta_i}\mathcal{Q}_{\alpha_i}|\Psi\rangle = 0 \Rightarrow \exists\kappa_i \in \mathbb{C}, \text{s.t.}$$
$$\sum_{i=1}^{m}\kappa_i\mathcal{Q}_{\alpha_i}|\Psi\rangle = 0.$$

Now, we want to consider the structure of the quotient space that we can get by acting *both* with the $\mathcal{Q}_\alpha$ polynomials and with the A_α polynomials. First note that by using the commutation relations of the operators inside A_α with $\mathcal{Q}_\alpha$, we can always move the $\mathcal{Q}_\alpha$ to the right. So, we start by considering the module produced by acting freely, first with $\mathcal{Q}_\alpha$ and then with A_α.

$$\mathcal{V} = \left\{\sum_{i=1}^{\mathcal{D}_\mathcal{A}}A_{\beta_i}\mathcal{Q}_{\alpha_i}|\Psi\rangle\right\},$$

where the set is formed by considering all possible combinations of A_{β_i} and $\mathcal{Q}_{\alpha_i}$. Some vectors in $\mathcal{V}$ are null, because the leading charge polynomials in the expression have annihilated the base state. Say that a basis of polynomials, which annihilate the state, is given by $\mathcal{Q}_{n_1}\ldots\mathcal{Q}_{n_P}$, all of which satisfy

$$\mathcal{Q}_{n_i}|\Psi\rangle = 0, \quad i = 1\ldots P.$$

For example, we might have null vectors because $|\Psi\rangle$ is an eigenvector of some charge $(\hat{Q}-q)|\Psi\rangle = 0$, as we discussed in the previous subsection. Or, as we mentioned earlier, we might have null vectors because $|\Psi\rangle$ is only finitely separated from the highest-weight state: $J_+^K|\Psi\rangle = 0$, for K greater than some number. All of these types are included in the set above.

Then the set of all null vectors in $\mathcal{V}$ is given by the set of all vectors that are obtained by acting with an element of the $\mathcal{A}$ on the null vectors listed above. More precisely, the null set in $\mathcal{V}$ is

$$\mathcal{N} = \left\{\sum_{i=1}^{P} A_{\beta_i}\mathcal{Q}_{n_i}|\Psi\rangle\right\}, \tag{3.21}$$

where the set is formed by considering all possible A_{β_i}

Let us prove the equivalence of (3.20) and (3.21), which is not immediately obvious. Consider some arbitrary null vector

$$|n\rangle = \sum_{i=1}^{K} A_{\beta_i}\mathcal{Q}_{\alpha_i}|\Psi\rangle. \tag{3.22}$$

We will now prove that (3.20) implies that this can always be written in the form (3.21). We note that (3.20) implies that the set of vectors $\{\mathcal{Q}_{\alpha_1}|\Psi\rangle, \ldots \mathcal{Q}_{\alpha_K}|\Psi\rangle\}$ is not linearly independent. For the sake of generality, we will assume that there are multiple linear dependences in this set, and that some m vector, $\mathcal{Q}_{\alpha_1}|\Psi\rangle \ldots \mathcal{Q}_{\alpha_m}|\Psi\rangle$ are linearly independent. However,

$$|n_j\rangle = \mathcal{Q}_{\alpha_j}|\Psi\rangle - \sum_{i=1}^{m}\kappa_j^i\mathcal{Q}_{\alpha_i}|\Psi\rangle = 0, \quad m+1 \le j \le K, \tag{3.23}$$

which simply states that $\mathcal{Q}_{\alpha_{m+1}}|\Psi\rangle \ldots \mathcal{Q}_{\alpha_K}|\Psi\rangle$ are dependent on the first m vectors. Consequently,

$$|n\rangle = \sum_{i=1}^{m}\left(A_{\beta_i} + \sum_{j=m+1}^{K}\kappa_j^i A_{\beta_j}\right)\mathcal{Q}_{\alpha_i}|\Psi\rangle.$$

From (3.20), we see that for this to hold, each term in the sum over i must vanish individually, and so

$$A_{\beta_i} = -\sum_{j=m+1}^{K}\kappa_j^i A_{\beta_j}, \quad 1 \le i \le m,$$

as an identity. This means that we can write (3.22) as

$$|n\rangle = \sum_{j=m+1}^{K} A_{\beta_j}\left(\mathcal{Q}_{\alpha_j} - \sum_{i=1}^{m}\kappa_j^i\mathcal{Q}_{\alpha_i}\right)|\Psi\rangle.$$

From (3.23), we see that we can write this precisely as

$$|n\rangle = \sum_{j=m+1}^{K} A_{\beta_j}|n_j\rangle,$$

which is of the form (3.21). This proves what we require.

The action of the mirror operators.—The physical space $\mathcal{H}_\psi$ is given by the quotient

$$\mathcal{H}_\psi = \mathcal{V}/\mathcal{N}.$$

Our task is to define the action of $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ on this space in a natural manner, and also ensure that $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ annihilates all elements of $\mathcal{N}$.

First, we define the action of $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ on the space $\mathcal{V}$. Our intuition is just that, $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ should transform the same way as the adjoint of the ordinary operator $(\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger$. For the ordinary operator, we have some commutation relations that are imposed by how the operator transforms under the algebra. In particular, denoting by Q^1 a *single charge* (not a polynomial) we have

$$[(\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger, Q^1] = t^{ij}(\mathcal{O}^j_{n,\boldsymbol{m}})^\dagger, \tag{3.24}$$

where t^{ij} is some matrix that describes the transformation of the operator. Note that, in general, $\mathcal{O}^i_{n,\boldsymbol{m}}$ will not transform in an irreducible representation, because we have chosen conventions where the position space operator $O^i(t,\Omega)$ is Hermitian. As we pointed out above, this does not involve any loss of generality, and the mirror of any operator can be obtained by means of linear combinations, and the use of the antilinearity of the mirror map.

Now, we define the action of $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ on an element of $\mathcal{V}$ as follows:

$$\begin{aligned}\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}A_{\alpha_1}Q^1A_{\alpha_2}\mathcal{Q}_{\alpha_3}\ldots A_{\alpha_n}|\Psi\rangle \\ = t^{ij}A_{\alpha_1}\tilde{\mathcal{O}}^j_{n,\boldsymbol{m}}A_{\alpha_2}\mathcal{Q}_{\alpha_3}\ldots A_{\alpha_n}|\Psi\rangle \\ + A_{\alpha_1}Q^1\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}A_{\alpha_2}\mathcal{Q}_{\alpha_3}\ldots A_{\alpha_n}|\Psi\rangle.\end{aligned} \tag{3.25}$$

Q^1 is the same charge that appears in (3.24), and $\mathcal{Q}_{\alpha_2}\ldots$ are arbitrary polynomials in the charges. We have specified how $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ commutes through a single charge, but clearly we can use this definition recursively to move through the rest of the operators acting on $|\Psi\rangle$ above, as well, and hence define the action of $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ on any element of $\mathcal{V}$.

To ensure that this action is consistent on $\mathcal{H}_\psi$, we simply notice the following simple fact. The definition (3.25) implies

$$\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}A_\beta\mathcal{Q}_{\alpha_i}|\Psi\rangle = A_\beta e^{\frac{-\beta\omega_n}{2}}(\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger\mathcal{Q}_{\alpha_i}|\Psi\rangle. \tag{3.26}$$

In fact, we can use the commutation relations of the charges with the ordinary operators, to always move all the charges immediately next to $|\Psi\rangle$, so (3.26) can be used as an alternate definition of $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ on $\mathcal{V}$, as was done in [12].

However, what (3.26) tells us immediately is that acting on an element of $\mathcal{N}$, $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ returns another element of $\mathcal{N}$. Hence, this linear operator $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$ consistently reduces to a linear operator on the quotient space $\mathcal{H}_\psi$ and transforms in the representation conjugate to $\mathcal{O}^i_{n,\boldsymbol{m}}$.

C. Decoupled harmonic oscillators

Now, having discussed the construction of mirror operators in a general theory, and in the CFT, we will move to a simple and concrete example: a set of decoupled harmonic oscillators. As we will discuss below, we can even use these decoupled harmonic oscillators as a model of the s-waves emitted from the black hole. As a result, this model provides us with significant insight into several recent discussions of the information paradox, some of which [7,9] are basically phrased in this context. We should caution the reader that while this model is very simple and explicit, the flip side is that we will need to use states with a small spread in energy, rather than energy eigenstates, to "mock up" some of the features of the interacting theory and define the mirror operators.

Consider a collection of harmonic oscillators of different frequencies. If we think of this as a model for the s-waves emitted by the black hole, the lowest frequency s-wave is inversely proportional to the Page time. We will write

$$\omega_{\mathrm{IR}} = \frac{2}{3} M_{\mathrm{pl}} \left(\frac{M_{\mathrm{pl}}}{M}\right)^3,$$

where the coefficient of $\frac{2}{3}$ has been chosen by hand, for reasons that will be apparent below. The "gas" of oscillators consists of frequencies $p\omega_{\mathrm{IR}}$ for positive integers p.

As introduced above, M is just a parameter that controls this lowest frequency, but we will take it to be the average total energy in the harmonic oscillators. For consistency with the notation above, in this section, we will use

$$\mathcal{N} = \sqrt{\frac{M}{\omega_{\mathrm{IR}}}}.$$

Now, notice that there is another physical interpretation of this gas of decoupled harmonic oscillators. Let us say that we quantize a massless field outside the black hole with one boundary condition at the Page length R_{Page} and another boundary condition on the field that can be placed a few Schwarzschild radii away from the horizon of the black hole. The exact position of the inner cutoff is not important, and it can be placed far enough from the black hole, that most of the radiation is in outgoing s-waves. We have shown our setup schematically in Fig. 2. This would automatically lead to the set of frequencies that we have above.

Now, consider a configuration of these oscillators with total energy that in a small band $[M - \Delta, M + \Delta]$ where, in this analysis, we will have to take $\Delta \propto \mathcal{N}$, for a reason that we explain below.

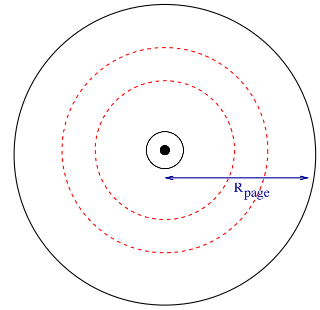

FIG. 2 (color online). A toy model of a black hole (small black circle in the center) emitting Hawking radiation. We are quantizing a massless field between the two solid circles, one of which is a Page distance away. The emission is mostly in s-waves if the inner cutoff is far enough from the horizon.

We see that the number of such configurations is given by the number of sets $\{n_p\}$ that satisfy the integer equation

$$M - \Delta < \sum_p p n_p \omega_{\mathrm{IR}} < M + \Delta.$$

Since $\Delta \ll M$, to leading order, we are just counting the number of solutions to the Diophantine equation $\sum_{p=1}^{\mathcal{N}^2} p n_p = \mathcal{N}^2$. The log of the leading term in the number of solutions, N_{sol}, is given by Cardy's formula

$$\log(N_{\mathrm{sol}}) = 2\pi \sqrt{\frac{\mathcal{N}^2}{6}} = \pi \frac{M^2}{M_{\mathrm{pl}}^2} \equiv S.$$

So, this gas of s-waves has the right entropy up to a numerical factor that we have inserted by hand in choosing the lowest IR frequency.

Now, let us consider the field outside the black hole. Again, neglecting the higher angular momenta, this has an expansion in terms of outgoing s-waves and can be written as

$$\phi(t, r) = \sum_{p=1}^{\mathcal{N}^2} \left[\frac{a_p}{2\pi\sqrt{p}} \frac{e^{-i\omega_p(t-r)}}{r} + \frac{b_p}{2\pi\sqrt{p}} \frac{e^{-i\omega_p(t+r)}}{r} + \mathrm{H.c.} \right], \tag{3.27}$$

where, $\mathcal{N}$ is defined above. The modes a_p correspond to the outgoing modes, and b_p correspond to the "ingoing" modes.

We want to consider an excitation of this field, that comprises purely outgoing modes. So, we consider a *pure state* in our gas of decoupled harmonic oscillators made out of the states in the energy band that we discussed above,

$$
|\Psi\rangle = \sum_{n_p} \alpha(n_p)|n_p\rangle,
$$
$$
M - \Delta < \sum p n_p \omega_{\rm IR} < M + \Delta,
$$
$$
|n_p\rangle \equiv \prod_p \frac{(a_p^\dagger)^{n_p}}{\sqrt{n_p!}}|\Omega\rangle,
$$

where the $\alpha(n_p)$ are some randomly chosen coefficients,[8] and the sum runs over all states that live in this energy band. We can associate a "temperature" to the state $|\Psi\rangle$, take $\beta = \frac{\partial S}{\partial M} = \frac{2\pi M}{M_{\rm pl}}$

We want to find the mirror operators $\tilde{a}_p$ and $\tilde{a}_p^\dagger$. We could simply take an appropriate state from the ensemble discussed above and follow the general rules for defining the mirror operators. However, this model is so simple that it is useful to derive them from scratch.

We would like operators that act in this theory but which "mimic" the thermofield state

$$
|\Psi\rangle_{\rm tfd} = Z_\beta^{-1} \sum_{\{n_p\}} e^{-\frac{\beta}{2}\sum_p n_p \omega_p} |\tilde{n_p}\rangle |n_p\rangle,
$$

where the sum is taken over all functions n_p, and Z_β is the partition function, which normalizes the state.

The operators that act on the "other side" of this entangled state are

$$
\begin{aligned}
\tilde{a}_m^{\rm tfd}|\Psi\rangle_{\rm tfd} &= Z_\beta^{-1} \sum_{\{n_p\}} e^{-\frac{\beta}{2}\sum_p n_p \omega_p} \sqrt{n_m} |\tilde{n_p} - \delta_{pm}\rangle |n_p\rangle \\
&= Z_\beta^{-1} \sum_{\{n_p'\}} e^{-\beta\frac{\omega_m}{2} - \frac{\beta}{2}\sum_p n_p' \omega_p} \sqrt{n_m' + 1} |\tilde{n_p'}\rangle |n_p' + \delta_{pm}\rangle \\
&= e^{\frac{-\beta\omega_m}{2}} a_m^\dagger |\Psi\rangle_{\rm tfd}.
\end{aligned}
$$

In the second line above, we changed the sum from n_p to $n_p' = n_p - \delta_{pm}$, which allowed us to notice that the action of $\tilde{a}_m^{\rm tfd}$ on this state was simply related to the action of $a_m^\dagger$. We can easily derive the same result by following the prescription of (A3).

Using a very similar calculation, we find that

$$
\tilde{a}_m^{\dagger,{\rm tfd}}|\Psi\rangle_{\rm tfd} = e^{\frac{\beta\omega_m}{2}} a_m |\Psi\rangle_{\rm tfd}.
$$

For each pair of operators $a_m, a_m^\dagger$, we now *define* the operators

$$
\begin{aligned}
\tilde{a}_m|\Psi\rangle &= e^{-\frac{\beta\omega_m}{2}} a_m^\dagger |\Psi\rangle, \\
\tilde{a}_m a_{m_1} \dots a_{m_{n_1}} a_{m_{n_1+1}}^\dagger \dots a_{m_{n_2}}^\dagger |\Psi\rangle &= a_{m_1} \dots a_{m_{n_1}} a_{m_{n_1+1}}^\dagger \dots \\
&\quad \times a_{m_{n_2}}^\dagger \tilde{a}_m |\Psi\rangle. \qquad (3.28)
\end{aligned}
$$

[8]The band defines a finite Hilbert space, and we can choose the $\alpha(n_p)$ by using the Haar measure on this space.

We define $\tilde{a}_m^\dagger$ in a similar manner,

$$
\begin{aligned}
\tilde{a}_m^\dagger|\Psi\rangle &= e^{\frac{\beta\omega_m}{2}} a_m |\Psi\rangle, \\
\tilde{a}_m^\dagger a_{m_1} \dots a_{m_{n_1}} a_{m_{n_1+1}}^\dagger \dots a_{m_{n_2}} |\Psi\rangle &= a_{m_1} \dots a_{m_{n_1}} a_{m_{n_1+1}}^\dagger \dots \\
&\quad \times a_{m_{n_2}} \tilde{a}_m^\dagger |\Psi\rangle. \qquad (3.29)
\end{aligned}
$$

In the formulas above, the product of operators in the second line of both (3.29) and (3.28) is, as usual, limited to cases where $n_2 \ll \mathcal{N}$.

We see, once again, that these equations are consistent provided that the set of products that we consider must have the property that no linear combination of these products must annihilate the state $|\Psi\rangle$. Otherwise, we run into the difficulties mentioned above, and the linear equations defining $\tilde{a}_m, \tilde{a}_m^\dagger$ may fail to have a solution.

We now see the importance of the band Δ. It serves to ensure that the operator

$$
\left(\sum m a_m^\dagger a_m - \mathcal{N}^2\right)|\Psi\rangle \neq 0.
$$

In fact to annihilate $|\Psi\rangle$, we need to take a product of $\mathcal{N}$ such operators, which is the width of the energy band. In the CFT, we required no such restriction because even an energy eigenstate in the CFT has a spread of occupation numbers of single-trace operators.

With these restrictions, the tilde operators can be used to construct a mirror "field"

$$
\tilde{\phi}(t, r) = \sum_p \frac{\tilde{a}_p}{2\pi\sqrt{p}} \frac{1}{r} \frac{e^{i\omega_p(t-r)}}{r} + \text{H.c.} \qquad (3.30)
$$

Note that we cannot reconstruct the $\tilde{a}_p^\dagger$ for very high $p \propto \mathcal{N}^2$ very well, because acting even a few times with the corresponding a_p can annihilate the state. However, these operators are negligible within correlation functions.

The field (3.30) commutes with the ordinary field in (3.27), within low-point correlators evaluated on $|\Psi\rangle$ and has the same correlators as one expects from the thermofield doubled state up to corrections that are expected in changing ensembles. Note, as usual, that the wave function multiplying $\tilde{a}_p$ has been conjugated.

D. Mirror operators in the spin chain

To aid the reader, we finally describe our construction in a second simple example: a simple spin-chain model. In Appendix E, we present a numerical computation of the mirror operators in this model. The reader may choose to directly consult that appendix and the included computer program to see how the various features of the mirror operators work out in an absolutely concrete setting.

We considered this toy model first in [1]. Consider a spin chain consisting of $\mathcal{N}$ spin $1/2$ particles labeled by

$i = 1, \ldots, \mathcal{N}$. Each spin i has a set of associated spin observables, $\mathbf{s}^i_a$, which satisfy

$$[\mathbf{s}^i_a, \mathbf{s}^j_b] = \frac{1}{2} i\epsilon_{abc}\delta^{ij}\mathbf{s}^i_c.$$

The simultaneous eigenstates of the $\mathbf{s}^i_z$ operators in this theory can be specified in terms of a single number from 0 to $2^{\mathcal{N}-1}$ using the eigenvalues of the operator $B = \sum_{i=1}^{\mathcal{N}} (\mathbf{s}^i_z + \frac{1}{2})2^{i-1}$. In this basis of B eigenstates, satisfying $B|n\rangle_B = n|n\rangle_B$, consider a state

$$|\Psi\rangle = \sum \alpha_n |n\rangle_B, \tag{3.31}$$

where the a_n can be picked randomly using the Haar measure on $\mathcal{C}P^{2^{\mathcal{N}}-1}$.

One commonly considered model of Hawking evaporation has been to imagine these spins "breaking off" from the spin chain one by one to constitute the outgoing Hawking radiation. This model should not be taken too seriously, but we will use it to illustrate our ideas.

The key issue in Hawking radiation is that bits are emitted in "pairs." After p bits have evaporated, the "outside observer" can make measurements involving the $\mathbf{s}^1_a \ldots \mathbf{s}^p_a$ operators. For each such measurement that the outside observer can make, there is a commuting measurement $\tilde{B}$ that the "inside" observer can make, and moreover the results of the two experiments are exactly correlated.

Here, we are interested in identifying the mirrored measurements. This means that we would like to find operators $\tilde{\mathbf{s}}^i_a$, which also satisfy

$$[\tilde{\mathbf{s}}^i_a, \tilde{\mathbf{s}}^j_b] \doteq \frac{1}{2} i\epsilon^{abc}\delta^{ij}\tilde{\mathbf{s}}^i_c,$$

where, as we have mentioned above, the $\doteq$ indicates that this algebra will be satisfied in the state of the theory, and not as an operator algebra. Moreover, we would like

$$[\tilde{\mathbf{s}}^i_a, \mathbf{s}^j_b] \doteq 0, \tag{3.32}$$

and that, in the state under consideration (and its descendants obtained by acting with these Pauli-spin matrices), these measurements to be perfectly correlated,

$$\langle\Psi|\tilde{\mathbf{s}}^i_a \mathbf{s}^j_b|\Psi\rangle = -\delta^{ij}\delta_{ab}. \tag{3.33}$$

These conditions together imply that for measurements of low-point correlators of $\mathbf{s}^i_a$ and $\tilde{\mathbf{s}}^j_b$, a given state $|\Psi\rangle$, on which they are defined, *looks like* the thermofield doubled state

$$|\Psi\rangle_{\text{tfd}} = \sum_B |B\rangle|2^N - 1 - B\rangle, \tag{3.34}$$

in the notation above. This thermofield doubled state is just a direct product of $\mathcal{N}$-entangled Einstein-Podolsky-Rosen pairs: $|\Psi\rangle_{\text{tfd}} = (|0\tilde{1}\rangle + |1\tilde{0}\rangle)^{\mathcal{N}}$, where the exponentiation by $\mathcal{N}$ means we need to take the direct product of this state with itself $\mathcal{N}$ times.

We now show how the operators $\tilde{\mathbf{s}}^i_a$ can be obtained very simply in a given state. As usual, we define these operators by specifying their action on a set of vectors. First, we describe how $\tilde{\mathbf{s}}^i_a$ acts on $|\Psi\rangle$,[9]

$$\tilde{\mathbf{s}}^i_a|\Psi\rangle = -\mathbf{s}^i_a|\Psi\rangle. \tag{3.35}$$

Next, we describe how it acts on states that differ from the action of $|\Psi\rangle$ by an action of up to K-ordinary $\tilde{\mathbf{s}}^i_a$ operators. For any product of operators, where p below satisfies $p < K$, we demand

$$\tilde{\mathbf{s}}^i_a \prod_{j=1}^{p} \mathbf{s}^{i_1}_{a_1} \ldots \mathbf{s}^{i_p}_{a_p}|\Psi\rangle = \left(\prod_{j=1}^{p} \mathbf{s}^{i_1}_{a_1} \ldots \mathbf{s}^{i_p}_{a_p}\right)\tilde{\mathbf{s}}^i_a|\Psi\rangle. \tag{3.36}$$

Note that $\tilde{\mathbf{s}}^i_a$ can be a $2^{\mathcal{N}} \times 2^{\mathcal{N}}$ matrix, and to describe the operator, we need to specify its action on $2^{\mathcal{N}}$ linearly independent vectors. The rules (3.35) and (3.36) together specify the action of $\tilde{\mathbf{s}}^z_m$ on

$$\mathcal{D}_{\mathcal{A}} = \sum_{j=0}^{K} \binom{\mathcal{N}}{j} 3^j$$

basis vectors. Provided that we do not take K to scale with $\mathcal{N}$, we have $n_K \ll 2^{\mathcal{N}}$. In fact, the precise condition we need in order to be able to construct the mirror operators is just $\mathcal{D}_{\mathcal{A}} < 2^{\mathcal{N}}$. So, there is a $(2^{\mathcal{N}} - \mathcal{D}_{\mathcal{A}})^2$-parameter family of choices of operators $\tilde{\mathbf{s}}^i_a$ that satisfy (3.35) and (3.36). If we like, we can restrict this ambiguity by increasing K, but the action of all of these operators coincides *exactly* within low-point correlation functions.

This prescription guarantees the correct behavior of $\tilde{\mathbf{s}}^i_a$ within low-point correlators, with up to K insertions, as specified by (3.32) and (3.33). For example, (3.36) tells us that within a low-point correlator $\tilde{\mathbf{s}}^i_a$ commutes *exactly* with $\mathbf{s}^j_b$, $\forall\, a, b, i, j$.

Note that the fact that our operators are state dependent is quite important here. For example, it is easy to prove that there is no operator $\tilde{\mathbf{s}}^j_b$ in the theory, except for the identity operator, with commutes exactly with all the $\mathbf{s}^i_a$ matrices. Our point is that, within low-point correlators, we can

[9]The careful reader may have noticed that we have switched conventions a little from the setting of Sec. III A, by inserting the additional minus sign in (3.35). This is because the thermofield doubled state, which we are mimicking here, has anticorrelated eigenvalues. It also allows us to ensure that the mirror operators obey the same, rather than the conjugated algebra. In the spin-chain setting, the convention we use here is more natural, and we hope that this will not confuse the reader.

produce operators that *look* like they achieve this zero commutator.

IV. RESOLVING VARIOUS PARADOXES

We now explain how our construction of the previous section resolves *all* the recent paradoxes that have been brought up in the recent literature on the information paradox. In particular, we resolve the following issues in order.

(1) The strong subadditivity paradox in Sec. IV A.
(2) The apparent issue of nonvanishing commutators between the early radiation and measurements inside the black hole in Sec. IV B.
(3) The apparent problem with the lack of a left-inverse for "creation" operators inside the black hole in Sec. IV C.
(4) The apparent argument that the infalling observer measures a nonzero particle number in Sec. IV D.
(5) The apparent "theorem" that small corrections cannot unitarize Hawking radiation in Sec. IV E.

A. Resolution to the strong subadditivity puzzle

We now describe how our operators resolve the strong subadditivity puzzle of Mathur [10] and Almheiri *et al.* (AMPS) [2]. This puzzle was proposed, while tacitly keeping in mind the picture of the spin chain, where Hawking evaporation is understood to be simply the detachment of individual spins from the chain. So, we first resolve the puzzle in this context.

We then try and formulate the strong subadditivity paradox, as carefully as we can, in terms of CFT correlators—something that, to our knowledge, has not been done so far. We then resolve it in this more precise context.

Summary of resolution.—Before we proceed to our detailed resolution, let us summarize the naive formulation of the strong subadditivity puzzle. We think of three subsystems: (1) the early radiation E, and a Hawking pair that is just being emitted, which consists of (2) B—the particle just outside the horizon and (3) its $\tilde{B}$—the particle just inside the horizon. For an old enough black hole, B must be entangled with E, and for the horizon to be smooth B is entangled with $\tilde{B}$.

Our resolution to this paradox is simple: the system $\tilde{B}$ is *not* independent of E. However, this overlap is cleverly designed so that commutators of operators on the early radiation E and the bit $\tilde{B}$ vanish $[E, \tilde{B}] \doteq 0$, when inserted within a low-point correlation function evaluated on the state of the system $|\Psi\rangle$.

We have already discussed this resolution in our previous paper [1]. The key new point is to ensure that such a resolution does not lead to a situation, where the observer outside can transmit messages to the observer inside, within the regime where effective field theory should be reliable. Our construction, where the $\tilde{B}$ operators are cleverly constructed to have exactly vanishing commutator with the operators in E, within all low-point correlators, ensures this.

We now discuss the resolution of the strong subadditivity puzzle in the spin chain, where the puzzle can be formulated most clearly. We then discuss some of the subtleties of formulating the puzzle in the CFT, attempt to formulate it as precisely as possible, and then resolve it in that context.

1. Resolution to the strong subadditivity puzzle in the spin chain

Let us first describe how the "strong subadditivity" of the entropy puzzle of [2,10] is resolved in the spin chain. Given the state $|\Psi\rangle$ in (3.31), if we consider the reduced density matrix of the first n qubits then for most choices of the coefficients a_i, we expect that [24]

$$\begin{aligned} S_n &= -\mathrm{Tr}(\rho_n \ln \rho_n) \\ &= \left[n\theta\left(\frac{\mathcal{N}}{2} - n\right) + (\mathcal{N} - n)\theta\left(n - \frac{\mathcal{N}}{2}\right)\right] \\ &\quad + \mathrm{O}(2^{-\frac{\mathcal{N}}{2}}) \ln 2. \end{aligned}$$

This curve is shown in Fig. 3.

The well-known interpretation of this equation is as follows. Consider the case where $k - 1 > \frac{\mathcal{N}}{2}$ bits have evaporated and we are considering the evaporation of the kth qubit. Then it is possible to find a set of operators—which we will call $\hat{\mathbf{s}}_a^k$—that obey the usual $SU(2)$ algebra and satisfy

$$[\hat{\mathbf{s}}_a^k, \mathbf{s}_b^k] = 0, \qquad \hat{\mathbf{s}}_a^k |\Psi\rangle = -\mathbf{s}_a^k |\Psi\rangle,$$

Hence, these $\hat{\mathbf{s}}_a^k$ operators effectively realize the algebra of the kth spin, without acting on that spin at all. This is the statement that these operators are "entangled" with the kth spin.

The choice of $\hat{\mathbf{s}}_a^k$ as operators on the first $(k-1)$ bits is not unique: we can take it to act on any selection of $\frac{\mathcal{N}}{2}$ qubits in these $(k-1)$ qubits. Nevertheless, the strong subadditivity condition, in this context, can be stated by saying that any such operator that we find *cannot* commute with the spin operators on the first $(k-1)$ bits:

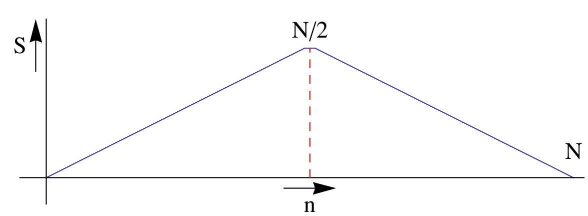

FIG. 3 (color online). Behavior of the entanglement entropy S_n with n.

$$[\hat{\mathbf{s}}^k_a, \mathbf{s}^m_b] \neq 0, \quad \text{for some m with } 1 \leq m \leq k-1.$$

How is this consistent with our explicitly constructed operator $\tilde{\mathbf{s}}^k_j$, which *appears* to commute with all the ordinary spins? The point is that, as an *operator*, it is indeed true that $[\tilde{\mathbf{s}}^k_j, \mathbf{s}^m_i] \neq 0$, for some $m \in 1 \ldots k-1$. But, nevertheless, this commutator annihilates the state $|\Psi\rangle$, and its descendants produced by acting with the insertion of up to K-ordinary and mirror operators,

$$[\tilde{\mathbf{s}}^k_i, \mathbf{s}^m_j]\mathbf{s}^{m_1}_{a_1} \ldots \mathbf{s}^{m_p}_{a_p}|\Psi\rangle = 0, \quad \forall\{m, j, i, a_1, \ldots a_p, m_1, \ldots m_p\}. \tag{4.1}$$

The equation continues to be true if we replace either some or all of the ordinary $\mathbf{s}^{m_p}_{a_p}$ matrices with the tilde counterparts.

Thus, within this model, our construction provides a precise realization of black hole complementarity. After the Page time, the operators in the interior of the black hole secretly act on the early radiation as well. Nevertheless, this action is exactly "local" within K-point correlators because of the vanishing of the commutator, as displayed in (4.1). The physical interpretation is that locality can be preserved exactly unless we try and consider correlators with $O(\mathcal{N})$ insertions.

2. *Resolving the strong subadditivity puzzle in the CFT*

We now resolve the strong subadditivity puzzle within the CFT. First, we need to formulate the puzzle precisely, and even this exercise suffers from some subtleties as we describe here. One possible precise formulation is to use the "plasma-ball" construction of [25], and this is what we use. After formulating the paradox in terms of plasma-ball evaporation in the boundary CFT, we then describe a resolution that is identical in spirit to the resolution for the spin chain demonstrated above.

Subtlety in formulating the strong subadditivity paradox in quantum gravity.—Making the strong-subadditivity paradox precise within quantum gravity is actually somewhat subtle. We summarize this difficulty and then attempt to reformulate the strong subadditivity paradox in terms of the CFT in an independent manner, and resolve it in that context.

The naive formulation of strong subadditivity relies on the idea that S_E rises, and then falls to zero. Within local quantum field theory, S_E could be defined as the entanglement entropy between the region "outside" and "inside" an imaginary barrier that is placed at a fixed distance from the black hole.

The subtlety is that the entanglement entropy of these regions even in the vacuum is infinite. This may not be the case in a fully theory of quantum gravity, but we do not understand how quantum gravity effects automatically resolve this divergence, in any detail. One could try and define a "renormalized" entanglement entropy, by considering the "excess" entanglement entropy in the state $|\Psi\rangle$ over the vacuum

$$S^{\text{ren}}_E = S^{\Psi}_E - S^{\Omega}_E.$$

However, now we run into the following difficulty: the definition above is very sensitive to the precise definition of the region E, since both terms on the right-hand side are divergent. Since the metric is changing as the black hole evaporates, it does not make sense to define E to be the region inside a given coordinate distance. Depending on how precisely we define the region, we can make S^{ren}_E increase, decrease, or stay constant, even as we cross the Page time.

It may be possible to avoid this subtlety by defining the entanglement entropy on $\mathcal{I}^+$, but below we explore an alternate formulation, which avoids our having to go to asymptotic infinity.

Formulation of the strong subadditivity paradox in terms of plasma-ball evaporation.—So, we now describe an alternate formulation that helps us sidestep this issue of the backreaction of Hawking radiation on the metric. Let us imagine a plasma ball in the conformal field theory, which is a lump of a deconfined phase that is localized in the boundary directions as depicted in Fig. 4. Gravity solutions corresponding to such a configuration were found in [25]. To consider these solutions, we switch back to the picture of the CFT living on R^4.

The picture of Hawking radiation is shown in the figure below. On the boundary, we expect that the quark-gluon plasma will decay via the emission of glueballs. These glueballs propagate away freely from the original plasma ball.

An *intuitive* way to think of this process, which is valid at large N, is to imagine a "plasma-ball operator" $\mathcal{P}_M$ which creates a plasma ball of energy M, but no glueballs. We can now consider operators that create wave packets of glueballs,

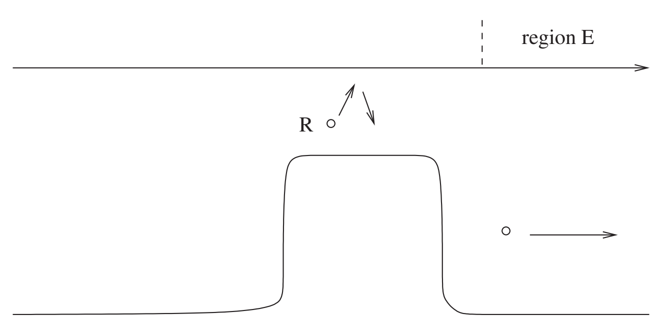

FIG. 4. Modeling Hawking radiation in a "localized black hole." Hawking quanta R emitted towards the boundary reflect and fall back, but quanta emitted in the noncompact directions escape to infinity.

$$O^i(f) = \int O^i(\boldsymbol{x}) f(\boldsymbol{x}) d^d\boldsymbol{x}, \tag{4.2}$$

where $f(\boldsymbol{x})$ is some function that controls the profile of the wave packet on the boundary. At large N, we can imagine that the evaporation of the plasma ball can be thought of as the Schrodinger evolution of the state from a pure plasma-ball state to a plasma ball of lower energy M' and some glueball wave packets,

$$\mathcal{P}_M|\Omega\rangle \longrightarrow \sum_{\{N\}} \alpha_{\{N\}}(t)(O^i(f_n))^{N(i,n)}|\mathcal{P}_{M'}\rangle, \tag{4.3}$$

where the sum runs over all functions $N(i,n)$. The functions f_n are some suitably regularized basis of glueball wave-packet profiles, and the α are some coefficients. We caution the reader again that the equation above is valid only at large $\mathcal{N}$, where we can clearly differentiate the plasma nall from the glueballs, and we provide it only for intuition.

The advantage of formulating this puzzle in the field theory is that we can make a much more precise statement. Consider the entanglement entropy of the region E on the *boundary*. We can regulate this entanglement entropy in some time-independent way, without having to worry about gravitational backreaction. Then we can consider the behavior of the entanglement entropy of region S_E with time, and we expect that this has the form expected from Page's general analysis, which is shown in Fig. 3.

Now, we can rephrase the strong subadditivity paradox as follows in this setting. First, we allow the plasma ball to form, and then allow it to evaporate for a while according to the process (4.3) until the emitted glueball spreads out over some region. In the bulk, we expect the geometry settles down and we reach a state of approximate thermal equilibrium. The evaporation time scale of the plasma ball scales with $\mathcal{N}$, and so if we are interested in processes that occur over a parametrically smaller time scale we can also consistently associate a temperature to the system.

We also need to be a little careful about the fact that the concept of temperature is local in space as well, and restricted to the neighborhood of the plasma ball. However, if we start with a plasma ball with an energy that scales with $\mathcal{N}$, allow it to evaporate, and then consider processes that have a spatial extent that is parameterically smaller, then again there does not seem to be a problem in associating a temperatue to the region of interest.

If we further consider observables in this region of approximate thermal equilibrium, and include them in our set $\mathcal{A}$, then the plasma ball also satisfies (3.16). Note that far away from the plasma ball where the CFT state looks locally like the vacuum, we can write down localized operators that annihilate the state, but this is not the case in the region near the plasma ball which is populated by thermal radiation.

The alert reader will have noticed that we have introduced some imprecision into our discussion because of this need to delineate scales in both space and time, so as to get a notion of approximate thermal equilibrium, and also to restrict the observables that we are interested in. We do not see any insuperable obstacle to making this precise. But we emphasize that our primary purpose in introducing this plasma ball is to phrase the strong subadditivity paradox in a setting where the entanglement entropy of the early radiation, the late radiation, and the black hole can be defined unambiguously—something that cannot be done directly in a theory of quantum gravity—and it is clear that this system meets that objective.

With this caveat, we can, as usual, map the boundary fields to bulk fields, and construct the mirror operators inside the black hole, and use them to construct the bulk fields. Now, we can consider some "wave packet" of the bulk operator

$$\phi^i_{\text{CFT}}(g) = \int \phi^i_{\text{CFT}}(z,\boldsymbol{x}) g(z,\boldsymbol{x}) d^{d+1}\boldsymbol{x},$$

where g is a function in $d+1$ dimensions with support in some region entirely inside the bulk black hole, behind the horizon. We might expect that this corresponds to some localized excitation entirely inside the Plasma-ball. However, let i' run over the set of glueball primaries that correspond to the supergravity modes. Then, strong subadditivity implies that

$$\exists i', h(x) \text{ such that } [\mathcal{O}^{i'}(h), \phi^i_{\text{CFT}}(g)] \neq 0, \tag{4.4}$$

where $O^{i'}(h)$ is defined by (4.2), and the function h on the boundary has the property that it vanishes everywhere *inside* the (past) light cone of the domain of g.

We can also phrase this as a property of the function $\tilde{\mathcal{O}}^i(\mathbf{x})$. Consider a "wave packet" of this mirror operator on the boundary, $\tilde{\mathcal{O}}^i(g_{\text{bound}})$ defined by (4.2), where g_{bound} has support on a region after the "Page time" of the plasma ball. Then we find that strong subadditivity implies that

$$\exists i', h', \text{ such that } [\mathcal{O}^{i'}(h'), \tilde{\mathcal{O}}^i(g_{\text{bound}})] \neq 0,$$

where h' is localized on some region that is spacelike separated from the domain of g_{bound}. So, $\tilde{\mathcal{O}}^i(\boldsymbol{x})$ cannot be a local operator on the boundary.

In other words, this is telling us that the operators inside the black hole after the Page time $\tilde{\mathcal{O}}_{\boldsymbol{m}_0,\boldsymbol{n}_0}$ must act on the glueball modes in the region E. Since these glueball modes also constitute the Hawking particles outside the horizon, we see that we have a precise version of the statement that the interior of the black hole has support on the degrees of freedom outside.

As we have seen several times above, however, the statement (4.4) is an operator statement. What we really want is that within low-point correlators built on the state $|\Psi\rangle$,

$$[O^{i'}(h), \phi^i_{\rm CFT}(g)] \doteq 0,$$

and there is absolutely no contradiction between this statement and (4.4).

Distilling the entangled bit?—It is worthwhile to briefly comment on another version of the strong subadditivity paradox. Could the observer outside "distill" the part of the outgoing radiation that is entangled with the near-horizon mode, and then jump into the black hole to obtain a contradiction.

It is simple to see, as we now show, that to "distill" the entangled bit, the infalling observer has to measure a correlator where the energy of the insertions scales with $\mathcal{N}$. First note that to "distill" the entangled bit is the same as finding an operator that is a polynomial in the ordinary operators $\mathcal{O}^i_{n,\boldsymbol{m}}$, but one that does *not* commute with the operators $\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}$. More precisely, calling this extraordinary operator $\mathcal{E}$, we need

$$\mathcal{E} = P(\mathcal{O}^i_{n,\boldsymbol{m}}), [\mathcal{E}, \tilde{\mathcal{O}}^{i_0}_{n_0,\boldsymbol{m}_0}]|\Psi\rangle \neq 0, \quad \text{for some } i_0, n_0, m_0, \tag{4.5}$$

where P is some polynomial.

Now, in (3.2), we have already ensured that the operators $\tilde{\mathcal{O}}^{i_0}_{n_0,\boldsymbol{m}_0}$ commute with all elements of $\mathcal{A}$, while acting on the state $|\Psi\rangle$. In particular, this includes all polynomials in which the the energy of every monomial does not scale with $\mathcal{N}$. So, we see that the polynomial P in (4.5) must include a term that violates (3.14) to have a nontrivial commutator with $\tilde{\mathcal{O}}^{i_0}_{n_0,\boldsymbol{m}_0}$ when acting on the state $|\Psi\rangle$.

This could happen, if for example, we consider a measurement that has $\mathrm{O}(\mathcal{N})$ insertions of supergravity fields. Or else, we could run into this difficulty if we take $\mathrm{O}(\sqrt{\mathcal{N}})$ insertions of fields with an energy $\mathrm{O}(\sqrt{\mathcal{N}})$ each. Translated to the supersymmetric $SU(N)$ theory, these measurements correspond to correlators with N^2 insertions of supergravity fields, or N insertions of giant graviton operators.

Our point is that these *correlators do not have an interpretation in terms of fields propagating on a perturbative spacetime*, and so it is not surprising that our intuitive concepts of spacetime—such as the idea that the interior and exterior of the black hole are distinct and well separated regions—break down for such correlators.

We would like to make a few more comments on this issue. One way of getting around the difficulty above has been to "couple" the CFT to another large system, and then perform the measurement in that large system. This does not affect our conclusions here at all. To the extent that the CFT coupled to the large system has a spacetime interpretation, this interpretation breaks down for measurements in this extended system that correspond to inserting N^2 supergravity fields.

We emphasize that our argument here is entirely independent of the bounds from quantum computing that have been discussed in this context [5]. This argument has been criticized in the later versions of [2], and we refer the reader to that paper. But, in any case we do not feel that these bounds are crucial to the discussion on the information paradox.

Finally, it is amusing to note that, in any case, "distilling" the entangled bit requires a state-dependent measurement (see Appendix C). Hence, if state-dependent measurements are disallowed even in principle then an observer who is part of the bulk spacetime in the first place, and then evolves autonomously with this spacetime, cannot make the required measurement.

B. The $[E,\tilde{B}] \neq 0$ paradox

An immediate objection to the picture of "complementarity" that we have outlined above is that the commutator of measurements on the radiation outside, and on measurements inside will not vanish. This is based on the observation that generically the commutator of two qubits is $O(1)$.

Let us briefly explain this objection, although it obviously does not apply to our construction. The point is that if we take the operator $\mathcal{O}^i_{n_0,\boldsymbol{m}_0}$ and "scramble" it using some generic $e^S \times e^S$ unitary matrix $U_{\rm scram}$ then it is generically true that

$$[U_{\rm scram}\mathcal{O}^i_{n_0,\boldsymbol{m}_0}U^\dagger_{\rm scram}, \mathcal{O}^i_{n_0,\boldsymbol{m}_0}] \sim \mathrm{O}(1),$$

in the sense that the generic size of the eigenvalue of the matrix on the left is $\mathrm{O}(1)$. This nonzero commutator can be detected within low-point correlators.

We emphasize that our construction of the mirror operators is *not* of this sort, and so the argument above fails completely. As we have emphasized many times above, our entire construction is designed to ensure that the commutator $C = [\tilde{\mathcal{O}}^i_{n_0,\boldsymbol{m}_0}, \mathcal{O}^i_{n_1,\boldsymbol{m}_1}]$ is undetectable within low-point correlators. More precisely, we have $CA_\alpha|\Psi\rangle = 0$ and hence any low-point correlator involving C or even $C^\dagger C$ vanishes.

So, our version of complementarity cannot be used to send messages across spacelike distances, at least within the approximation that the spacetime geometry makes sense at all.

The fact that this commutator vanishes, in this effective sense, is an extremely important consistency check for our construction. It is also the key element that is required to make the "complementarity resolution" of the strong subadditivity paradox viable. We also wish to point out that, in our construction, this lack of independence between the degrees of freedom outside and inside the black hole is *not* restricted to evaporating black holes. In fact, even for a big black hole in AdS, if we probe the exterior finely enough, we can see that the interior degrees of freedom are not independent.

C. The lack of a left-inverse paradox

Now, let us turn to some of the other arguments of [7]. One of these arguments goes as follows. Consider some conformal primary corresponding to a supergravity field, and consider the action of $\tilde{\mathcal{O}}^i_{-n,\boldsymbol{m}}$ on the pure state $|\Psi\rangle$, where n is any positive integer. This operator acts like a "creation" operator for the field behind the horizon. For this subsection, we adopt the following shorthand notation:

$$G_\beta(n,\boldsymbol{m}) = \langle\Psi|[\mathcal{O}^i_{n,\boldsymbol{m}}, (\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger]|\Psi\rangle,$$
$$b = \frac{1}{\sqrt{G_\beta(n,\boldsymbol{m})}}\mathcal{O}^i_{n,\boldsymbol{m}}; \quad b^\dagger = \frac{1}{\sqrt{G_\beta(n,\boldsymbol{m})}}(\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger,$$
$$\tilde{b} = \frac{1}{\sqrt{G_\beta(n,\boldsymbol{m})}}\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}; \quad \tilde{b}^\dagger = \frac{1}{\sqrt{G_\beta(n,\boldsymbol{m})}}(\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}})^\dagger.$$

However, by the relation (3.19), we have $[H_{\text{CFT}}, \tilde{b}^\dagger] = -\omega_n\tilde{b}^\dagger$ so the action of this operator *lowers* the energy in the CFT and maps a state of average energy E to a state of average energy $E-\omega_n$. (Recall that ω_n was defined to be $n\omega_{\text{IR}}$ in Sec. III B.) Nevertheless, some simple algebra shows us that this operator satisfies the following relation to leading order in the $\frac{1}{N}$ expansion:

$$\begin{aligned}\langle\Psi_0|(\tilde{b}\tilde{b}^\dagger - \tilde{b}^\dagger\tilde{b})|\Psi_0\rangle &= \langle\Psi_0|(\tilde{b}e^{\frac{\beta\omega_n}{2}}b - \tilde{b}^\dagger e^{\frac{-\beta\omega_n}{2}}b^\dagger)|\Psi_0\rangle \\ &= \langle\Psi_0|(bb^\dagger - b^\dagger b)|\Psi_0\rangle \\ &= 1 + \mathrm{O}\left(\frac{1}{N}\right).\end{aligned}$$

This allows us to write

$$\tilde{b}\tilde{b}^\dagger \doteq 1 + \tilde{b}^\dagger\tilde{b}. \tag{4.6}$$

We have been careful to put a $\doteq$ above, once again indicating that this relation holds within low-point correlation functions.

The "lack of a left-inverse" is simply the claim that had we had a true operator equality in (4.6) then since the right-hand side is a manifestly positive operator, $\tilde{b}^\dagger$ should have a left-inverse paradox. But this appears to be impossible, since there are fewer states in the smaller energy range.

Of course, we do not have any contradiction with our state-dependent construction, where (4.6) is not satisfied as an operator equation but as a relation that holds within low-point correlators constructed on $|\Psi\rangle$.

In fact, we can choose the $\tilde{b}^\dagger$ operators to be rather sparse on the full Hilbert space. This is because the linear equations of (3.17) and (3.18) are not in contradiction with multiple null vectors for $\tilde{b}^\dagger$. For example, as we pointed out in Sec. III A, we could choose $\tilde{b}_m$ so that it obeys Eqs. (3.17) and (3.18) within the space $\mathcal{H}_\Psi$, but annihilates all vectors that are orthogonal to this subspace. By construction, this would not create any contradiction with low-point correlators.

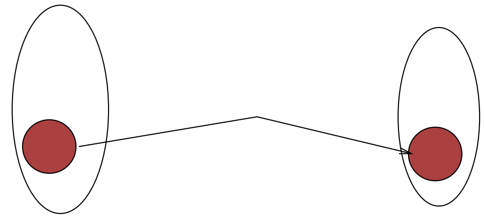

FIG. 5 (color online). $\tilde{b}^\dagger$ is a sparse operator, and it maps the intersection of $\mathcal{H}_\Psi$ with the space of states of average energy E to the intersection of $\mathcal{H}_\Psi$ with the states of average energy $E-\omega_n$. The precise domain and range depend on the base state $|\Psi\rangle$.

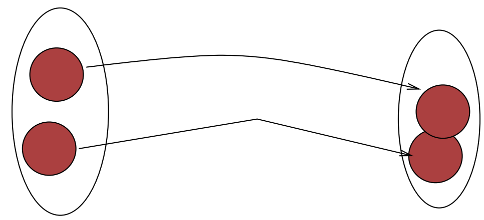

FIG. 6 (color online). Extending the domain of $\tilde{b}^\dagger$ by taking the union of the maps corresponding to different base states leads to a many-to-one function.

What the argument of [7] tells us is whatever action we choose for $\tilde{b}^\dagger$ outside the space $\mathcal{H}_\Psi$ this operator must have null vectors. However, within low-point correlators, these null vectors are completely unobservable and it appears that these operators obey the algebra (4.6).

Pictorially, we can depict the action of $\tilde{b}^\dagger$ by Fig. 5.

Union of all constructions?—The paper [7] contained a further argument to try and account for the situation described in Fig. 5. The argument was that, if we consider the "union of all constructions," we could get a contradiction with the expectation of (4.6).

This argument was not spelled out in detail, but by this we understand the following: the operator $\tilde{b}^\dagger_\omega$ provides a map between states of higher and lower energy, as shown in Fig. 5. This map depends on the state. Perhaps, the authors of [7] meant to suggest that by considering different states, and by considering the union of all these maps, we could obtain operators that satisfied (4.6) as an operator equation, rather than just on the states under consideration.

Here, we wish to point out that the "union of all constructions" does not help in this. If one tries to take the maps corresponding to different base states $|\Psi\rangle$, so as to completely cover the space with energies in a band about E, then we invariably end up *overcovering* the space with energies in a band about $E-\omega_n$. This is shown in Fig. 6.

D. The "$N_a \neq 0$" paradox

We now turn to an argument made in [9], leading to the apparent conclusion that AdS/CFT cannot describe the interior of the black hole. First, we summarize the argument and then show why it fails in our construction.

Summary of the Marolf-Polchinski argument.—We start by defining two "number operators"

$$N_b = b^\dagger b,$$
$$N_a = \frac{1}{1-e^{-\beta\omega_n}}[(b^\dagger - e^{\frac{-\beta\omega_n}{2}}\tilde{b})(b - e^{\frac{-\beta\omega_n}{2}}\tilde{b}^\dagger) + (\tilde{b}^\dagger - e^{\frac{-\beta\omega_n}{2}}b)(\tilde{b} - e^{\frac{-\beta\omega_n}{2}}b^\dagger)]. \quad (4.7)$$

We see that N_b measures the number of particles at frequency ω_n as seen by the Schwarzschild observer outside. The operator N_a is the standard number operator as seen by the infalling observer and the factors of $e^{-\beta\omega_n}$ come from the standard Bogoliubov transformations between these two frames [26].

Note that (4.7) is also relevant in Rindler space, where N_b could be the number operator measured by a Rindler observer and N_a the number operator measured by the Minkowski observer. However, now we come to a crucial difference between the Rindler and the AdS/CFT case. We see that

$$[H_{\rm CFT}, N_b] = 0 + {\rm O}(\omega_{\min}).$$

On the other hand, the commutator between the Minkowski Hamiltonian and the Rindler number operator clearly does not vanish. So, the CFT Hamiltonian behaves like the Rindler Hamiltonian.

As a consequence of the fact above, we can consider a set of eigenstates of N_b, which we will denote by $|\bar{E}, n_b\rangle_i$, which have the property that

$$H_{\rm CFT}|\bar{E}, n_b\rangle_i = \bar{E}|\bar{E}, n_b\rangle_i + {\rm O}(\omega_{\min}),$$
$$N_b|\bar{E}, n_b\rangle_i = n_b|\bar{E}, n_b\rangle_i.$$

The two conditions above, which specify the energy up to an accuracy $\omega_{\min}$, and the N_b eigenvalue still leave an enormous degeneracy, and the index i is meant to denote the different states that can satisfy this property. Now, consider the span of all such states that have mean energy in some range

$$\mathcal{S} = {\rm span}\{|\bar{E}, n_b\rangle_i : \bar{E}_0 - \Delta \leq \bar{E} \leq \bar{E}_0 + \Delta\}.$$

It seems clear that no element of the *basis* of $\mathcal{S}$, that we used above, has a smooth horizon. If we reconstruct the bulk, for such a state, using the bulk-boundary map, and evaluate the stress tensor as we approach what would have been the horizon, it will diverge. This is entirely consistent with the fact that the states in this basis do not satisfy (3.2), and so we cannot construct the mirror operators on them. However, we can consider the following harder question:

"Consider a typical state in $\mathcal{S}$, picked with the Haar measure on this space. Does such a state have a smooth horizon, or not?"

The authors of [9] claim that for $\Delta \sim {\rm O}(\beta^{-1})$, the set $\mathcal{S}$ covers almost the entire microcanonical ensemble with width β^{-1} centered on $\bar{E}_0$. We do not entirely understand the basis for this estimate of the width, or the subtleties in determining whether $\mathcal{S}$ really contains almost all states in the microcanonical ensemble. Neither of these details are provided in the paper [9]. As a consequence, the reader should note that there may be a subtle difference between the question above, and the question of whether a typical state in the microcanonical ensemble has a smooth horizon or not.

The authors of [9] argued that the answer to the question above is negative. We will now review their argument, and then show that it fails for state-dependent operators, and that typical states in the span of $\mathcal{S}$ do have a smooth horizon.

Let us say that some *state-independent* operator in the CFT could tell us the particle number as measured by the infalling observer. We will call such an operator $N_a^{\rm univ}$. Then, we could compute

$$\langle N_a^{\rm univ}\rangle = \frac{1}{\dim(\mathcal{S})}{\rm Tr}_{\mathcal{S}}(N_a^{\rm univ}) = \frac{1}{\dim(\mathcal{S})}\,{}_i\langle\bar{E}, N_b|N_a^{\rm univ}|N_b, \bar{E}\rangle_i = {\rm O}(1).$$

The last equality follows because $N_a^{\rm univ}$ is a positive operator. Moreover since the state with $N_a^{\rm univ} = 0$ has a thermal distribution of N_b, the expectation value of $N_a^{\rm univ}$ in any N_b eigenstate is ${\rm O}(1)$.

This is consistent with the fact that typical states with a definite Rindler energy are not regular as we cross the Rindler horizon.

Failure of the argument for state-dependent operators.—First, we point out the following simple fact. Consider a typical state $|\Psi\rangle \in \mathcal{S}$. With respect to the usual set of observables $\mathcal{A}$ defined in Sec. III B, we would expect such a state to satisfy (3.16), and so we can define the mirror operators. Now, it is immediately clear from (3.17) and (3.18) and the definitions (4.7) that

$$N_a|\Psi\rangle = 0,$$

which follows from the simple observation that both

$$(b - e^{\frac{\beta\omega_n}{2}}\tilde{b}^\dagger)|\Psi\rangle = 0, \quad \text{and} \quad (\tilde{b} - e^{\frac{\beta\omega_n}{2}}b^\dagger)|\Psi\rangle = 0.$$

The reason that the argument above fails is that our operator N_a is state dependent, and in fact, it is partly designed to ensure that $N_a = 0$ in a typical state $|\Psi\rangle$. For such an operator, the change of basis in the trace clearly fails.

Consider another simple example of this sort. Let us say that $\rho_\psi = |\Psi\rangle\langle\Psi|$ is the *density operator* corresponding to the state $|\Psi\rangle$. Clearly, we have $(\rho_\psi - 1)|\Psi\rangle = 0$, and this is true for any state $|\Psi\rangle$. On the other hand ${\rm Tr}_{\mathcal{S}}(\rho_\psi - 1) = 1 - \dim(\mathcal{S}) \neq 0$. These two statements are not in any contradiction, because ρ_ψ is a state-dependent operator precisely like our N_a.[10]

E. Unitarizing Hawking radiation with small corrections

We now address the claim that "small corrections" cannot unitarize Hawking radiation [10]. Before, we address this claim, it is extremely important to specify what, precisely, is meant by "small corrections." From our perspective, the size of corrections is estimated by the size of corrections to low-point correlation functions of light local operators compared to the results that we would get from ordinary effective field theory in the black hole background.

Thus, for example, if there is structure behind the horizon then we might expect large corrections to correlators involving insertions on either side of the horizon. Similarly, if the process of Hawking radiation is modified significantly, then we might expect large corrections even to correlators outside the black hole, because the state will not be well approximated by the Unruh vacuum.

We stress that it is important to adopt the definition above, rather than one that looks at, say, whether the full wave function at the end of Hawking evaporation is close to that predicted by the Hawking calculation. We can see the error in this kind of approach even if we consider the set of states that are dual to a large black hole in AdS. The wave functions of these states differ widely, but from a geometric perspective, or equivalently from the perspective of expectation values of elements in the set $\mathcal{A}$, these states are almost impossible to distinguish.

With this prelude, we now consider two cases, and show how the Hawking evaporation process is consistent with

(1) small corrections outside the horizon, and

(2) small corrections across the horizon.

We will phrase our arguments in this subsection in terms of the spin-chain toy model of III D, since this is the context in which the claim of [10] was formulated.

1. *Small corrections outside the horizon*

As we showed in our previous paper [1], it is perfectly consistent with unitarity for correlators of local fields outside the black hole to be very close to their semiclassical values, as calculated in the Unruh vacuum. In the spin-chain model that we have described, this is the following simple statement. For a correlator made up of products of spin operators, where the number of insertions does not scale with K, we have

$$\langle\Psi|\mathbf{s}_{i_1}^{a_1}\ldots\mathbf{s}_{i_p}^{a_p}|\Psi\rangle = \mathrm{Tr}(\rho_{i_1\ldots i_p}\mathbf{s}_{i_1}^{a_1}\ldots\mathbf{s}_{i_p}^{a_p}) = \frac{1}{2^p}\,\mathrm{Tr}(I_{2^p\times 2^p}\mathbf{s}_{i_1}^{a_1}\ldots\mathbf{s}_{i_p}^{a_p}) + \mathrm{O}(2^{\frac{-\mathcal{N}}{2}}),$$

where $\rho_{i_1\ldots i_p}$ is the reduced density matrix for the spins $i_1\ldots i_p$ in the state $|\Psi\rangle$.

The ordinary $\mathbf{s}_i^a$ operators correspond to measurements made outside the horizon. So, the interpretation of this equation is as follows: for the purposes of computing correlation functions with a small number of insertions outside the black hole, it is always possible to use the thermal density matrix—which, in our toy model, is just the identity matrix.

In our previous paper, we discussed this issue in a slightly different language by pointing out that a seemingly thermal density matrix could be unitarized by a correction matrix, that was exponentially suppressed.

This means that it is possible to have a situation where the exact density matrix ρ_{exact}, which comes from unitary evolution, differs from the Hawking density matrix, which is just the identity here, by a correction matrix ρ_{corr} whose elements, in some basis, are very small,

$$\rho_{\mathrm{exact}} = \rho_{\mathrm{hawk}} + 2^{-N}\rho_{\mathrm{corr}}.$$

This is consistent with the relation above. Correlators computed in the two density matrices vary by a factor of $2^{\frac{-N}{2}}$, since the typical contribution of the second term is $2^{\frac{-N}{2}}$.

In this context, we should mention that when $\mathrm{O}(\mathcal{N})$ particles have been emitted, it may look like the correction matrix is comparable to the original Hawking matrix. This is just an indication of the fact that the Hawking computation cannot reliably predict the amplitude for any given configuration of $\mathrm{O}(\mathcal{N})$ emitted particles, since this amplitude is exponentially suppressed. When we focus on correlators with a small number of insertions, then we invariably end up focusing on a reduced density matrix with a much smaller dimension, and the difference between the unitary and thermal density matrix vanishes.

So, to conclude this subsection, small corrections to correlators of an O(1) number of fields outside the black hole are completely consistent with unitarity.

2. *Small corrections to correlators across the horizon*

The proof of [10] focused not just on the density matrix outside the horizon, but also on the evolution of the wave function of the theory during Hawking evaporation.

The assumption is that the full wave function evolves as

$$|\Psi\rangle_{t+1} = \frac{1}{\sqrt{2}}|U\Psi\rangle_t \otimes (|0\rangle_B|1\rangle_{\tilde{B}} + |1\rangle_B|0\rangle_{\tilde{B}}. \tag{4.8}$$

Here $|\Psi\rangle$ encodes the state of the black hole and the radiation at a given time. Equation (4.8) should be understood as the statement that the wave function evolves by the

[10]One difference, of course, is that whereas ρ_ψ is defined for all states $|\Psi\rangle$, including those that are eigenstates of N_b, our construction of mirror operators works for typical states that satisfy (3.16). As we mentioned above, for eigenstates of N_b, the construction fails, and this is consistent with the physical understanding that such states have no "interior."

addition of two entangled particles: of which one falls into the black hole and the other falls out, while the extant black hole and radiation evolve autonomously through the unitary matrix U.

Now (4.8) clearly cannot be taken literally in a theory of quantum gravity, especially if we are to take the lessons from AdS/CFT seriously. For example, as we explored above, a clear setting in which Hawking radiation can be observed is given by the localized plasma ball in AdS/CFT [25]. As shown in Fig. 4, the plasma ball is a localized lump of quark-gluon plasma that gradually evaporates via the emission of glueballs. What this teaches us is that Hawking radiation should be modeled as a process in which the black hole loses some of its energy to the emission of an external particle, while the remaining degrees of freedom reorganize themselves so that it *appears* that a particle has been created within the black hole.

Hence, what we must demand is that, while the underlying dynamics may be quite complicated, the wave function *effectively* evolves as in (4.8). What does this mean in terms of correlators? In our spin-chain model of Hawking evaporation, after k spins have evaporated and we can measure measure operators $\mathbf{s}_k^a$ on the emitted Hawking particle and there is an additional operator which effectively commutes with the spin operators for the emitted Hawking quanta that is exactly correlated with measurements of $\mathbf{s}_k^a$.

The punch line is that after $\mathcal{N}$ steps for the purposes of low-point correlators involving the operators $\tilde{\mathbf{s}}_n^a$ and $\mathbf{s}_m^a$, the state effectively looks like a collection of $\mathcal{N}$-Bell pairs, as shown below Eq. (3.34). But, in reality, it is only an entangled state involving $\mathcal{N}$, and not $2\mathcal{N}$, spins.

We should emphasize two important factors in our construction, which were *not accounted* for in the construction of [10]. One of them is that, loosely speaking, the interior particle may be constructed partly out of the previously emitted radiation. The precise version of this statement is, of course, given by our operators $\tilde{\mathbf{s}}_k^a$ above, which for $k > \frac{\mathcal{N}}{2}$, must necessarily act on some of the first $\frac{\mathcal{N}}{2}$ particles as well.

The second point has to do with the state dependence of our construction. The papers [10,27] described models of "burning paper." In these models, the qubit $\tilde{B}$ was identified with some specific qubit constructed in the remaining spin chain. It is clear that after $\frac{\mathcal{N}}{2}$ bits have evaporated, there is no state-independent operator that can be perfectly correlated with the emitted spin. Even, for the first $\frac{\mathcal{N}}{2}$ bits, the correlation between the qubit-$\tilde{B}$ and the qubit B can be maintained only by fine-tuning the state.

Let us say this more precisely. Consider some *fixed* state-independent operators $\underline{\mathrm{s}}_a$ and try and make this play the role of the mirror to the first spin. Then in some given state of the spin chain, we require $\underline{\mathrm{s}}_a|\Psi\rangle = -\mathbf{s}_a^1|\Psi\rangle$. Clearly, for some fixed operator $\underline{\mathrm{s}}_a$, this condition will not be met for a generic state $|\Psi\rangle$. So, for a generic state $|\Psi\rangle$, the correlator $\langle\Psi|\underline{\mathrm{s}}_a\mathbf{s}_a^1|\Psi\rangle \approx 0$, whereas we would like it to have the value -1. This is what leads to the suggestion that in models of burning paper there are large corrections to correlators between operators "inside" and "outside" the black hole.

These conclusions *do not* hold for our state-dependent operators $\tilde{\mathbf{s}}_a^i$. We clearly have $\tilde{\mathbf{s}}_a^i|\Psi\rangle = -\mathbf{s}_a^i|\Psi\rangle,\ \forall\, i$. Moreover, we see that *for a generic state*, we have $\langle\Psi|\tilde{\mathbf{s}}_a^i\mathbf{s}_b^j|\Psi\rangle = -\delta^{ij}\delta_{ab}$, precisely as we need. So, with state-dependent operators, we can arrange to have small corrections across the horizon as well, while remaining within a unitary framework.

Before, we conclude this section we would like to point out that the argument of [9] can, in some sense, be understood to be a rephrasing of the argument that small corrections cannot unitarize the black hole. In that discussion, the argument was that if we selected the operators corresponding to $\tilde{b}_m$ and $\tilde{b}_m^\dagger$ to be some *fixed* operators in the CFT Hilbert space, then it is clear that their action on the state will generically not be correlated with the action of b_m and $b_m^\dagger$. This argument fails for state-dependent operators as we showed above.

The conclusion is that, provided local fields in the interior of the black hole are constructed in a state-dependent manner, we can consistently reconcile unitary evolution with small corrections to correlation functions in effective field theory.

Numerically large nonlocality.—At this point, we also briefly address another criticism made in [6]. The "nonlocalities" that we mentioned do indeed spread out over the Page sphere of the black hole. For a solar-sized black hole, the Page sphere is huge: 10^{77} km. Nevertheless, we wish to emphasize that the nonlocality is incredibly difficult to measure.

In the construction that we have described, we have to measure a correlation function involving of order $\exp[10^{77}]$ points, before we can detect this nonlocality. So, to the extent that actual numbers are relevant to these conceptual issues, it is clear that we do not have any contradiction with either any observed or possible-to-observe physics.

V. NONEQUILIBRIUM SCENARIOS

So far, in this paper, we have discussed how to define the mirror operators in an equilibrium state. In this section, we briefly discuss the nonequilibrium scenario, leaving a more detailed study to further work.

Let us phrase the question that we are interested in more precisely. Our construction of the previous section *already* gives us interesting time-dependent correlators of fields $\langle\Psi|\phi_{\rm CFT}^{i_1}(t_1,\Omega_1)\ldots\phi_{\rm CFT}^{i_m}(t_m,\Omega_m)|\Psi\rangle$. However, when any of these operators is behind the horizon, we need to use the mirror operators to define it, and these mirror operators are defined in an equilibrium state. Now, let us say that someone gives us an out-of-equilibrium state $|\Psi'\rangle$, perhaps produced by exciting an equilibrium state with some sources. What, then is the correct way to define the mirror operators so as to get the results expected from semi-classical field theory?

Stated briefly, our proposal is that to deal with a nonequilibrium state that is produced by turning on sources dual to a small number of local operators on an equilibrium state, we "strip off" the excitations that create this nonequilibrium state from a thermal state. We now define the mirror operators on this thermal state, and then use these *unchanged* operators in the nonequilibrium state. We describe this more precisely below.

Our construction automatically addresses a technical objection that has been made to state-dependent proposals, which is sometimes called the "frozen vacuum." This is simply the claim that defining the mirror operators using the rules for equilibrium states always leads to a featureless horizon, even though one could manually excite the horizon by injecting some matter before the infalling observer falls in. Our proposal below for nonequilibrium states does not lead to any such issue.

It is true that one can, by hand, ensure that the infalling observer perishes at the horizon, by aiming a focused laser beam which intersects the observer just as he crosses it. In this section we show how to describe correlators outside and inside the horizon in such a scenario. However, our construction makes another unambiguous connection. Just as one would expect from semiclassical field theory, any such excitation soon shares the fate of the observer and falls into the singularity in a short amount of time leaving behind a featureless horizon once again.

A. Detecting nonequilibrium states

First, we discuss how to differentiate equilibrium states from nonequilibrium states. Let us say we are given some state $|\Psi'\rangle$. Can we detect, by measuring expectation values in this state, whether the state is in equilibrium or out of it?

The first point to note is that this, itself, is a manifestly time-dependent question. Consider a state that consists of superpositions of different energy eigenstates.

$$|\Psi'\rangle = \sum_i c_i |E_i\rangle.$$

Now, consider an element A_p of $\mathcal{A}$. It is very natural that, in an interacting CFT like supersymmetric Yang-Mills theory, the elements of A_p will obey the eigenstate thermalization hypothesis

$$\langle E_i|A_p|E_j\rangle = A(E_i)\delta_{ij} + e^{-\frac{1}{2}S(\frac{E_i+E_j}{2})}R_{ij}. \quad (5.1)$$

Here $S(\frac{E_i+E_j}{2})$ is the density of states at the mean energy, and below we will write just S for this quantity to lighten the notation. Note that $S \propto \mathcal{N}$ for the systems that we have considered above. A is a "smooth" function of their arguments, but R_{ij} is a matrix comprising erratically varying phases but a smoothly varying magnitude. (In the papers in [28], sometimes another function is introduced to capture this magnitude, but we have no need for it here.)

We will need a further technical assumption on the matrix R_{ij}. Note that because $\mathrm{Tr}(R^\dagger R) = \mathrm{O}(\#1)[e^{2S}]$ and R has e^S eigenvalues $r_1, \ldots r_{e^S}$, we expect that the typical magnitude of each eigenvalue will be $|r_i| = \mathrm{O}(e^{\frac{S}{2}})$. We also need to assume that *no eigenvalue* of R is much greater than this $|r_i|e^{\frac{-S}{2}} = \mathrm{O}(1)$, $\forall\, i$. On the other hand, the phase of r_i will generically be arbitrary.

This form is very natural and follows from the following simple assumption. The eigenvectors of the operator $\mathcal{O}^i$ are not correlated with the exact energy eigenstates. This is quite common in interacting field theories. If the two sets of eigenvectors are related by some "random" unitary transformation, then (5.1) follows.

For example, consider our "regularized frequency modes" $\mathcal{O}_{n,\boldsymbol{m}}$[11] which include a band of modes of width $\omega_{\min}$, and are defined in (3.13). We see that between two energy eigenstates, the following statements hold:

$$\langle E_i|\mathcal{O}_{\omega,\boldsymbol{m}}|E_j\rangle = \langle E_i|\mathcal{O}_{\boldsymbol{m}}(0)|E_j\rangle\delta(E_i - E_j - \omega),$$
$$\langle E_i|\mathcal{O}_{n,\boldsymbol{m}}|E_j\rangle = \frac{1}{(\omega_{\min})^{\frac{1}{2}}}\langle E_i|\mathcal{O}_{\boldsymbol{m}}(0)|E_j\rangle\theta(E_i - E_j - (n-1)\omega_{\min})\theta(E_j + n\omega_{\min} - E_i),$$

where $\mathcal{O}_{\boldsymbol{m}}(0) = \int \mathcal{O}(0,\Omega)Y^*_{\boldsymbol{m}}(\Omega)d^{d-1}\Omega$, which is a natural notation. The normalization factor of $\sqrt{\omega_{\min}}$ ensures that the diagonal elements of the operator

$$\langle E_i|\mathcal{O}_{n,\boldsymbol{m}}\mathcal{O}^\dagger_{n,\boldsymbol{m}}|E_i\rangle = \sum_{E_j=E_i+(n-1)\omega_{\min}}^{E_i+n\omega_{\min}} |\langle E_i|\mathcal{O}_{n,\boldsymbol{m}}|E_j\rangle|^2 = \mathrm{O}(1),$$

since the sum runs over $e^S\omega_{\min}$ states, and each term is of order $e^{-S}\omega_{\min}^{-1}$, assuming the ETH for the operator $\mathcal{O}_{\boldsymbol{m}}(0)$. So, we see that $\mathcal{O}_{n,\boldsymbol{m}}$ also obeys the ETH up to the additional normalization of $(\omega_{\min})^{-\frac{1}{2}}$, which is O(1) in the accounting of (5.1) and will not be important in the discussion below.

This analysis leads to

$$\chi_p(t) = \langle\Psi'|e^{iHt}A_p e^{-iHt}|\Psi'\rangle = \sum |c_i|^2(A(E_i) + e^{-\frac{S}{2}}R_{ii}) + \sum_{i\neq j} c_j c_i^* e^{-\frac{S}{2}} e^{i(E_i-E_j)t} R_{ij}.$$

The second term is manifestly time dependent. Now note, that by the assumption on the maximum size of the eigenvalues of R above, we see that at most we can get

[11]Sometimes we omit the superscript "i" of the operators ϕ^i_{CFT} and $\mathcal{O}^i_{n,\boldsymbol{m}}$ in order to lighten the notation.

an $\chi_p(t) - \chi_p(0) \leq \mathrm{O}(1)$ time dependence. However, in the generic situation where the coefficients c_i and time t are not carefully selected, the time-dependent term is $\chi_p(t) - \chi_p(0) = \mathrm{O}(e^{-\frac{S}{2}})$—exponentially suppressed in $\mathcal{N}$.

We can use this as a diagnostic of whether the state is in equilibrium or not. As we mentioned above, this is a time-dependent question, and in the CFT, we are interested in the issue of whether the state is in equilibrium from some starting time $t = 0$ to some other long time $t = \omega_{\mathrm{min}}^{-1}$. Recall that we introduced ω_{min} to regulate the frequency modes of the CFT, and we can even take $\omega_{\mathrm{min}} = e^{-\sqrt{N}}$, if we are interested in a big black hole that has a much longer lifetime.

So, to precisely evaluate whether a state is in equilibrium or not, we consider the following quantity:

$$\nu_p = \omega_{\mathrm{min}} \int_0^{\omega_{\mathrm{min}}^{-1}} |(\chi_p(t) - \chi_p(0))| dt.$$

We will declare that a state is in equilibrium if

$$\nu_p = \mathrm{O}(e^{-\frac{S}{2}}), \quad \forall\ p,$$

i.e. this property holds for all observables in $\mathcal{A}$. Otherwise we will classify it as a nonequilibrium state.

We emphasize that this is a much finer distinction than we require in practice. In fact, when we consider small black holes in AdS, they have a lifetime that is only polynomial in $\mathcal{N}$. This is also true of the plasma balls that we considered previously. (See Page 3 of [25].) In such cases, it may be useful to consider a slightly modified definition of an equilibrium state, where the time scale of evolution scales with $\mathcal{N}$ are effectively in equilibrium, while those that change over a much shorter time are not. For simplicity here, however, we restrict ourselves to large black holes in AdS.

B. Near equilibrium states

We now discuss a class of states that we will call "near equilibrium states." These are states that are produced by acting on an equilibrium state $|\Psi\rangle$, with a unitary matrix produced by exponentiating a Hermitian element of A_p

$$|\Psi'\rangle = U|\Psi\rangle, \qquad U = e^{iA_p}. \tag{5.2}$$

where $|\Psi\rangle$ is in equilibrium and A_p is Hermitian. Recall that although we have written the basis of the algebra $\mathcal{A}$ in terms of Fourier modes, we are allowed to take arbitrary linear combinations. So, the set above includes states that are produced by coupling a source to the boundary field for a limited amount of time. For example, we could take $U = e^{i\int J(t,\Omega)\mathcal{O}(t,\Omega)}$. We can also consider sources that couple to stringy modes or brane probes.

Although our construction can be generalized to several other statistical-mechanics systems, our presentation in this section will focus on the CFT. In the CFT, it is true that if $|\Psi\rangle$ is in "equilibrium" then any state $|\Psi'\rangle$ of the form (5.2) is out of equilibrium. The logic behind this claim is as follows. Consider turning on a source for some local operator in the CFT by adding $\int J(t,\Omega)\mathcal{O}(t,\Omega)$ to the Hamiltonian (the logic easily generalized to bilocal and k-local operators). It is possible to find another operator $\Pi(t,\Omega)$ so that

$$\langle\Psi|[\mathcal{O}(t,\Omega), \Pi(t',\Omega')]|\Psi\rangle = k(t-t',\Omega,\Omega') \neq 0.$$

Now, we see that

$$\begin{aligned}&\langle\Psi'|e^{iHt}\Pi(0,\Omega)e^{-iHt}|\Psi'\rangle \\ &\quad = \langle\Psi|\Pi(t,\Omega)|\Psi\rangle + \int J(t',\Omega')k(t-t',\Omega,\Omega')dt'd\Omega'.\end{aligned}$$

The second term above tells us that we have turned on the source, and can be detected.

If we turn on sources for operators in *momentum space*, this is still possible, although it may be a little confusing. This follows from an examination of (3.13), which tells us that our momentum-space operators are effectively defined over a time range of length $\omega_{\mathrm{min}}^{-1}$ in the CFT, and so turning on sources for these operators is just like turning on a slow-acting source for a position space operator.

For example, consider a state

$$|\Psi'\rangle = e^{i\lambda(\mathcal{O}^i_{n,\boldsymbol{m}} + (\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger)}|\Psi\rangle, \tag{5.3}$$

for some particular conformal primary, and some particular modes n and $\boldsymbol{m}$, where λ is a constant. In fact, such a state is out of thermal equilibrium since we can see that

$$\begin{aligned}&\langle\Psi'|((\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger - \mathcal{O}^i_{n,\boldsymbol{m}})|\Psi'\rangle \\ &\quad = \langle\Psi|e^{-i\lambda(\mathcal{O}^i_{n,\boldsymbol{m}} + (\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger)}((\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger - \mathcal{O}^i_{n,\boldsymbol{m}})e^{i\lambda(\mathcal{O}^i_{n,\boldsymbol{m}} + (\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger)}|\Psi\rangle \\ &\quad = \langle\Psi|((\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger - \mathcal{O}^i_{n,\boldsymbol{m}}) + 2\lambda|\Psi\rangle + \mathrm{O}\left(\frac{1}{\mathcal{N}}\right) \\ &\quad = 2\lambda + \mathrm{O}(e^{-\frac{S}{2}}).\end{aligned}$$

Here we used the Baker-Campbell-Hausdorff lemma in going from the second to the third line together with the fact that $[\mathcal{O}^i_{n,\boldsymbol{m}}, (\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger] = 1 + \mathrm{O}(\frac{1}{\mathcal{N}})$. However, now if we evaluate $\chi(t)$, we can already see that by evolving for a time $t_0 \approx \frac{i\pi}{n\omega_{\mathrm{min}}}$ that $\chi(t_0) = -2\lambda$. However, the long term value of $\chi(t)$ is 0. Over time scales larger than $\omega_{\mathrm{min}}^{-1}$, we see that the approximate commutation relations between $\mathcal{O}^i_{n,\boldsymbol{m}}$ and $(\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger$ break down because the different oscillators in (3.13) that comprise these operators decohere.

It is very hard to detect that the state (5.3) is out of equilibrium partly because of the nature of the source that we turned on. In terms of local operators, this corresponds to a slow-acting source that acts over a time scale of $\omega_{\mathrm{min}}^{-1}$.

We can consider a harder example: $|\Psi'\rangle = e^{i\lambda \mathcal{O}^i_{n,\boldsymbol{m}}(\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger}|\Psi\rangle$. In fact, even here it is possible to detect the action of this source as a *subleading order* in $\frac{1}{N}$. We need to find an operator Π so that $\langle\Psi|[\mathcal{O}^i_{n,\boldsymbol{m}}(\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger, \Pi]|\Psi\rangle \neq 0$. In an interacting CFT such an operator should exist on general grounds, although we cannot write down its explicit form here without knowing the operator product expansion coefficients in detail. Given such an operator, we can again use the logic above to detect this slow-acting source, and also the fact that the state is slightly out of equilibrium.

To summarize, this discussion implies that if a nonequilibrium state can be written in the form (5.2), then U is essentially uniquely fixed. Any other U' would only take the state out of equilibrium again. In some more detail: suppose that there are two different equilibrium states $|\Psi_1\rangle, |\Psi_2\rangle$ such that we can write the near-equilibrium state $|\Psi'\rangle$ as $|\Psi'\rangle = e^{iA_1}|\Psi_1\rangle$ and also $|\Psi'\rangle = e^{iA_2}|\Psi_2\rangle$. From these two we find $|\Psi_2\rangle = e^{-iA_2}e^{iA_1}|\Psi_1\rangle$. But we argued above that it is not possible for both $|\Psi_1\rangle$ and $|\Psi_2\rangle$ to be equilibrium states, unless $A_1 = A_2$.

C. Mirror operators for near-equilibrium states

We now describe how to construct mirror operators for nonequilibrium states. Given a nonequilibrium state $|\Psi'\rangle$, we have described above how we can detect that it is not in equilibrium and also find how it is related to the equilibrium state $|\Psi\rangle$ by

$$|\Psi'\rangle = U|\Psi\rangle.$$

We now define the action of the mirror operators by the following modified recursive rules in the CFT:

$$\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}|\Psi'\rangle = U e^{-\frac{\beta\omega_n}{2}}(\mathcal{O}^i_{n,\boldsymbol{m}})^\dagger U^\dagger|\Psi'\rangle, \tag{5.4}$$

$$\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}} A_p|\Psi'\rangle = A_p\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}}|\Psi'\rangle, \quad \forall\ A_p \in \mathcal{A}. \tag{5.5}$$

As usual, the factor of $e^{-\frac{\beta\omega_n}{2}}$ in (5.5) must be corrected at subleading orders in $\frac{1}{N}$, but the fact that the mirror operators commute through the ordinary operators should hold at all orders in perturbation theory.

Keeping this in mind, we can define the action of $\tilde{\mathcal{O}}_{n,\boldsymbol{m}}$ on the state $|\Psi'\rangle$ and its descendants in a single compact equation as

$$\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}} A_p|\Psi'\rangle = A_p U e^{-\frac{\beta\omega_n}{2}}\mathcal{O}^i_{-n,\boldsymbol{m}} U^\dagger|\Psi'\rangle.$$

D. The frozen vacuum

We now address the "frozen vacuum" objection to state-dependent proposals that was articulated by Bousso [11]. The argument of [11] was made in the context of the proposals of [4], which also use state-dependent operators. We do not understand some of the details in [11], but we translate what we think is the relevant part of the argument, albeit in somewhat more prosaic language.

The point is simply that we cannot use the rules (3.17) and (3.18) in a nonequilibrium state like $|\Psi'\rangle$ in (5.2) and expect to get the right semiclassical correlators. For example, as we saw above in Sec. IV D, the rules (3.17) tell us that the particle number observed by the infalling observer is zero, if there are no additional excitations i.e., $\langle\Psi|N_a|\Psi\rangle = 0$.

We do not expect this in nonequilibrium states. For a nontrivial U, notice, for example, that generically we have

$$(\tilde{\mathcal{O}}^i_{n,\boldsymbol{m}} - e^{\frac{\beta\omega_n}{2}}\mathcal{O}^i_{-n,\boldsymbol{m}})|\Psi'\rangle \neq 0,$$

and so with the operation of the mirror operators defined in (5.4) and (5.5), we generically obtain $\langle\Psi'|N_a|\Psi'\rangle \neq 0$. The precise expectation value depends on the kind of perturbation that we have made to the state.

We wish to emphasize that even in the equilibrium construction of the previous sections, or of our previous paper [1], it was perfectly possible to excite the horizon of the black hole. What we have done here is simply to explain how to construct the operators $\tilde{O}_{n,\boldsymbol{m}}$ when the base state that we are given is out of equilibrium. It is clear that our procedure of "stripping off" the U, and then restoring it, gives us exactly the same answers as one would get from effective field theory in the bulk.

Acting with a unitary behind the horizon?—We should also mention a second issue raised by van Raamsdonk [8]. In Ref. [8], van Raamsdonk considers a case where an autonomous unitary transformation is made on the second sided CFT. In fact, in the eternal black hole, a small perturbation made early enough in the second CFT can lead to a highly boosted shock wave just behind the horizon that separates regions I and II [29].

van Raamsdonk's argument is not directly relevant to our construction since we do not really have a second side in the collapsing geometry. These states do not really have an existence as states that are autonomously created by collapsing matter. In a collapsing geometry such a state can only be created indirectly, by pumping in an excitation from the outside. In such a situation we can detect the excitation outside the black hole, and use the more precise rules for the definition of the mirror operators given above.

Nevertheless, we point out that our discussion of near-equilibrium states does not exhaust the set of all states in the CFT, and so we would like to revisit this issue in future work.

E. An example: A beam from the boundary

Let us now consider an example in some detail, where we turn on a source at the boundary dual to some operator as depicted in Fig. 7. We wish to check the following qualitative conclusions. First, the correlators across the horizon should not be affected before the beam has time to

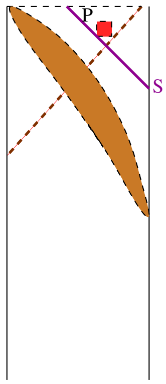

FIG. 7 (color online). A nonequilibrium state: a laser beam sent from the boundary point S intersects the patch of interest at P

reach P. Then, the correlators should be affected for some time, in a way that is determined by effective field theory. Finally, once we wait for a scrambling time, the correlations should go back to their previous values. We wish to consider

$$C_{12} = \langle \Psi' | \phi_{\text{CFT}}(t_1, \Omega_1, z_1) \phi_{\text{CFT}}(t_2, \Omega_2, z_2) | \Psi' \rangle,$$
$$|\Psi'\rangle = e^{i \int J(t,\Omega) \mathcal{O}(t,\Omega)} |\Psi\rangle,$$

where $J(t, \Omega)$ is a source that is sharply peaked around the origin of boundary coordinates $(t = 0, \Omega = 0)$, the point (t_1, Ω_1, z_1) is in front of the horizon, and (t_2, Ω_2, z_2) is *behind* the horizon, and $|\Psi\rangle$ is an equilibrium state.

Using the expansion (2.4), we see that

$$\begin{aligned} C_{12} = \sum_{m} \int_{\omega>0} \frac{d\omega}{2\pi} [&\langle \Psi | U^\dagger \phi_{\text{CFT}}(t_1, \Omega_1, z_1) \\ &\times (\mathcal{O}_{\omega,m} g^{(1)}_{\omega,m}(t_2, \Omega_2, z_2) + \text{H.c.}) U | \Psi \rangle \\ &+ \langle \Psi | U^\dagger \phi_{\text{CFT}}(t_1, \Omega_1, z_1) U (\tilde{\mathcal{O}}_{\omega,m} g^{(2)}_{\omega,m}(t_2, \Omega_2, z_2) \\ &+ \text{H.c.}) | \Psi \rangle], \end{aligned}$$

where we used that $[U, \tilde{\mathcal{O}}_{\omega,m}] = 0$ for the tildes defined with respect to the equilibrium state $|\Psi\rangle$.

Now, we see that the properties of C_{12} we inferred above follow directly from the properties of *ordinary* local fields under conjugation by U. We know that $[\mathcal{O}(0,0), \phi_{\text{CFT}}(t_1, \Omega_1, z_1)] = 0$, when the bulk point (t_1, Ω_1, z_1) is spacelike separated from the origin of the boundary coordinates, and this commutator also becomes small when the point is in the *far future* of the origin. However this commutator is appreciably nonzero, when the bulk point is near the light cone that extends from the origin of the boundary. The same result holds for the commutator $\sum_m \int_{\omega>0} \frac{d\omega}{2\pi} [\mathcal{O}(0,0), \mathcal{O}_{\omega,m} g^{(1)}_{\omega,m}(t_2, \Omega_2, z_2) + \text{H.c.}]$. These properties follow from an analysis of Green functions for perturbative fields in the bulk.

Just to clarify this point, we remind the reader that in 4-dimensional flat space the commutator for a scalar field ψ of mass m is [30]

$$\begin{aligned} [\psi(\boldsymbol{x}), \psi(\boldsymbol{y})] = &\frac{i}{2\pi} s(x^0 - y^0) \delta(\lambda) \\ &- \frac{im}{4\pi\sqrt{\lambda}} \theta(\lambda) s(x^0 - y^0) J_1(m\sqrt{\lambda}), \end{aligned}$$

where $\lambda = (\boldsymbol{x} - \boldsymbol{y})^2$, and s is the sign function in this equation. This commutator always vanishes at spacelike separation. For a massless field, the commutator is nonzero only on the light cone, but even for a massive field, this commutator vanishes for large timelike separation as well. The explicit expressions in the AdS-Schwarzschild geometry are much more difficult to write down, but the same qualitative properties hold. Note that this involves an interplay between the CFT commutators $\mathcal{O}_{\omega,m}$ and the transfer function.

So, in the case where the bulk points are in the far future, or spacelike separated from the source at the boundary, we can just commute the U through the ordinary operators to annihilate the $U^\dagger$ and so C_{12} reduces to the correlator in the state $|\Psi\rangle$. However, when either of the bulk points are near the light cone from the origin of the boundary, we expect that this correlator will receive appreciable corrections. This is exactly what we had inferred.

VI. LINKS WITH TOMITA-TAKESAKI THEORY

In this section we provide an additional (though mathematically equivalent) perspective to the construction of the mirror operators. We also discuss the relation of the current proposal to that of [1], which was based on a coarse/fine decomposition of the Hilbert space. Finally we present some intriguing mathematical connections of our construction with the Tomita-Takesaki theory of operator algebras.

Many of the ideas described in this section have already been discussed in Sec. III. We summarize them again for the convenience of the reader and slightly modify the presentation in order to connect with the Tomita-Takesaki theory.

A. Another intuitive explanation of our construction

In Sec. III we emphasized that the operators can be found by solving equations (3.8). We argued that since the number of equations is much smaller than the size of the Hilbert space, we can always find solutions, and we explicitly wrote down a solution in (3.10). Here we expand on a slightly different perspective, which was already mentioned at the end of Sec. III A. This perspective leads to a "constructive" definition of the mirror operators and is more suitable to make contact with the mathematical discussion of the next subsection.

The intuition is very simple. As mentioned several times in the previous sections of the paper, we imagine that we have a complicated quantum system with Hilbert space $\mathcal{H}$, which is in a particular pure state $|\Psi\rangle$. Also, we imagine that we can only probe the system by using a small set of observables $\mathcal{A}$. Since we will be computing correlation functions of these observables on the state $|\Psi\rangle$, it is very natural to define the span of states of the form

$$A_i|\Psi\rangle, \quad A_iA_j|\Psi\rangle, \quad A_iA_jA_k|\Psi\rangle, \text{etc.}$$

where $A_i \in \mathcal{A}$. We introduced the linear span of states of this form at the end of Sec. III A and we called it

$$\mathcal{H}_\Psi = \{\text{span of } A|\Psi\rangle, \quad A \in \mathcal{A}\}.$$

The space $\mathcal{H}_\Psi$ is a subspace of the full Hilbert space $\mathcal{H}$, which obviously depends on the choice of the initial state $|\Psi\rangle$. This is schematically depicted in Fig. 8.

The main shift in perspective from the discussion in Sec. III is the following: in that section we constructed the operators $\tilde{A}_p$, element by element, for each operators A_p in $\mathcal{A}$. Now, we will describe a more formal "one shot" construction of the mirror operators.

We will see that we can define the mirror operators in a very natural way by concentrating on how they act on the subspace $\mathcal{H}_\Psi$. Their action on this subspace is extremely natural. Their action on the "orthogonal subspace" $\mathcal{H}_\Psi^\perp$ is not completely specified—this is related to the fact that the equations in (3.8) have more than one solution—but this ambiguity has no effect on the computation of low-point correlation functions. We stress again, that this is equivalent to Eq. (3.8); the new perspective offered here provides some additional intuition and demonstrates a canonical solution of these equations.

As before, the starting point for the existence of the mirror operators is that, if the size of $\mathcal{A}$ is small relative to the size of the Hilbert space, then on general grounds we expect that a typical state $|\Psi\rangle$ *cannot be annihilated by nonvanishing elements of the set* $\mathcal{A}$ or, in equations, for $A \in \mathcal{A}$ we have

$$A|\Psi\rangle = 0 \Leftrightarrow A = 0. \tag{6.1}$$

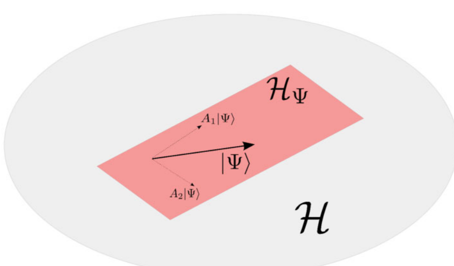

FIG. 8 (color online). A quantum system with Hilbert space $\mathcal{H}$ placed in the pure state $|\Psi\rangle$ is probed by a set of observables $A_i \in \mathcal{A}$. We define the subspace $\mathcal{H}_\Psi = \{\text{span of A}|\Psi\rangle, \text{A} \in \mathcal{A}$ which is relevant for computing correlation functions of observables in $\mathcal{A}$ on the state $|\Psi\rangle$ and the construction of the mirror operators.

In the case of the big black hole in AdS/CFT, the idea is that $\mathcal{A}$ is the set of products of a small number of single-trace operators, while $|\Psi\rangle$ is a typical quark-gluon-plasma (QGP) microstate. It is clear that such a typical QGP microstate cannot be annihilated by a small number of single-trace operators.

Equation (6.1) expresses the point that the state $|\Psi\rangle$ *looks entangled from the point of view of the algebra* $\mathcal{A}$. Usually we define entanglement in situations where the Hilbert space of the system has a bipartite structure, but here we generalize the concept of entanglement, by discussing how the state *appears to be entangled* in terms of certain *observables*. This is expressed by Eq. (6.1). We will argue that whenever we have such a situation, in which a quantum state looks (sufficiently) entangled when probed by a set of observables $\mathcal{A}$, then the set of observables $\mathcal{A}$ is "doubled." This doubling explains the origin of the dual modes behind the horizon.

We now start with these assumptions, i.e. that we have a big quantum system probed by a small number of observables $\mathcal{A}$, such that (6.1) is satisfied and we show how we define the mirror operators.

The starting point is that there is a natural way to define the action of a *second copy* of the observables $\mathcal{A}$ acting on the subspace $\mathcal{H}_\Psi$. This can be achieved by defining, effectively, an action of observables in $\mathcal{A}$ "from the right." This can be compactly described by introducing an antilinear map $S\colon \mathcal{H}_\Psi \to \mathcal{H}_\Psi$ defined by

$$SA|\Psi\rangle = A^\dagger|\Psi\rangle, \tag{6.2}$$

which obviously satisfies $S^2 = 1$ and also

$$S|\Psi\rangle = |\Psi\rangle.$$

Notice that condition (6.1) is crucial in order for (6.2) to be well defined.

Then it is easy to check that the operators defined by

$$\hat{A} = SAS \tag{6.3}$$

satisfy the following two properties:

(i) Their algebra is isomorphic to that of operators in $\mathcal{A}$, since $S^2 = 1$.
(ii) The hatted operators commute with operators in $\mathcal{A}$ when acting on elements of $\mathcal{H}_\Psi$.

To see this, notice that any vector in $\mathcal{H}_\Psi$ can be written as $C|\Psi\rangle$ for some $C \in \mathcal{A}$ and we have

$$\hat{A}BC|\Psi\rangle = SASBC|\Psi\rangle = SAC^\dagger B^\dagger|\Psi\rangle = BCA^\dagger|\Psi\rangle = B\hat{A}C|\Psi\rangle.$$

Hence

$$[\hat{A}, B]C|\Psi\rangle = 0,$$

for all $A, B, C \in \mathcal{A}$.

Notice the following important point: the subspace $\mathcal{H}_\Psi$ was defined as the span of states of the form $A|\Psi\rangle$. While the operators $\hat{\mathcal{A}}$ commute with those in $\mathcal{A}$, they are still acting on the same space $\mathcal{H}_\Psi$.

The operators $\hat{A}$ can be extended in the full Hilbert space $\mathcal{H} = \mathcal{H}_\Psi \oplus \mathcal{H}_\Psi^\perp$. One naive possibility would be by defining them to be "zero" on the orthogonal subspace $\mathcal{H}_\Psi^\perp$, but there are many other possibilities. This issue was already discussed at the end of Sec. III A.

In the previous steps we have identified a "second copy" $\hat{\mathcal{A}}$ of the observables acting on the space $\mathcal{H}$. This already captures the essence of the "doubling." However, to finalize the construction of the mirror operators and make contact with the conventional "thermofield doubling," it is convenient to perform a small redefinition of the operators $\hat{\mathcal{A}}$. The issue is that the mapping S is—in general—not (anti-)unitary. Hence the "normalization" of the operators $\hat{\mathcal{A}}$ is not the same as those of the $\mathcal{A}$. In order to fix this we can rescale the magnitude of the antilinear operator S by defining

$$S = J\Delta^{1/2},$$

where J is anti-uitary with $J^2 = 1$ and Δ positive and Hermitian. The precise definition of Δ will be discussed later. Then we can define the conventionally normalized mirror operators by

$$\tilde{A} = JAJ.$$

While it is obvious that the hatted operators (6.3) commute with elements of $\mathcal{A}$, it is less obvious that the $\tilde{A}$'s commute with operators in $\mathcal{A}$, due to the factors of $\Delta^{1/2}$. Nevertheless, it is a fact that they do commute and this will be explained in more detail later.

Finally let us notice that in most situations—and certainly in the case of the large N gauge theory—the "set of observables" $\mathcal{A}$ is not a closed algebra in the strict mathematical sense. For instance, if we attempt to define $\mathcal{A}$ as the set of "small number of insertions of single-trace operators," then this set is not strictly closed under operator multiplication.

This point is at the heart of black hole complementarity: while $\mathcal{A}$ is not an exact algebra, it behaves approximately like an algebra for certain low-point correlators. Hence the construction of the commuting mirror operators, as outlined above, approximately works for such low-point correlators. The existence of the mirror operators for low-point functions is sufficient in order to reconstruct the experience of the infalling observer.

On the other hand, if we act with too many of them, we will either "get out of $\mathcal{A}$," or we will have to allow the set $\mathcal{A}$ of "accessible observables" to be large enough so that it becomes an algebra and it contains all possible products. In this case the state $|\Psi\rangle$ is depleted of any entanglement with respect to $\mathcal{A}$ and condition (6.1) is not satisfied any more. Then the dual-operator construction does not work and the black hole interior ceases to make sense.

This is all in agreement with the idea of complementarity and the validity of effective field theory in the bulk.

So far these ideas were motivated by physical considerations. Intriguingly, the mathematical language in which the dual-operator construction was phrased above appears in surprisingly similar form in the theory of operator algebras as we explain in the next subsection.

B. Relation to the Tomita-Takesaki modular theory

We now describe an extremely interesting link between our construction of the mirror operators behind the horizon, and an area in the study of von Neumann algebras that goes by the name of Tomita-Takesaki theory. The existing reviews of this subject in the literature are somewhat formal, so we will summarize the main ideas here. The reader interested in a more sophisticated mathematical discussion can refer to [31–33].

Exactly like in our physically motivated construction mentioned above, the Tomita-Takesaki construction involves building the *commutant* of an algebra $\mathcal{A}$, and uses an appropriate state vector to do so. For example, given the set of operators on a finite interval, one could use the construction to generate the operators outside the light cone of this interval that, in a local quantum field theory, should commute with the original algebra. Here, we will use it to construct operators "behind" the horizon, given the operators in front of it.

The Tomita-Takesaki construction starts with an algebra $\mathcal{A}$ acting on a Hilbert space $\mathcal{H}$ and a state vector $|\Psi\rangle$ that is *cyclic* and *separating*. For a state to be *cyclic* means that the space $\mathcal{H}_\psi = \mathcal{A}|\Psi\rangle$ is dense in the Hilbert space $\mathcal{H}$. The statement that the state is separating is simply the condition (6.1): $A|\Psi\rangle \neq 0$, $\forall\ A \in \mathcal{A}$.

The reader can satisfy herself that these conditions are easily met, for example, in relativistic QFT, if one takes $\mathcal{A}$ to be the algebra of operators on an open ball of space time and $|\Psi\rangle$ to be the vacuum state. Part of this statement is the so-called Reeh-Schlieder theorem, that we also discuss in Appendix C.

Here, we are interested in a different situation. For us $|\Psi\rangle$ is a typical pure state that looks like it is close to thermality, whereas $\mathcal{A}$ is the *set* (not, necessarily, an algebra) of low-point correlation functions. Consequently, $\mathcal{H}_\psi$ is not dense in the larger Hilbert space $\mathcal{H}$, but this will not be an obstacle either, as we will now see. For the remainder of this section, and in order to state the Tomita-Takesaki theorem in simple form, we will just assume that $\mathcal{A}$ is an algebra and we will think of $\mathcal{H}_\Psi$ as the entire Hilbert space, so that the assumptions that $|\Psi\rangle$ is cyclic and separating are satisfied. In other words, in the following part we will imagine that $\mathcal{H}_\Psi$ plays the role of the entire Hilbert space $\mathcal{H}$ and will simply call it $\mathcal{H}$. We will discuss the important modifications necessary for the case of the large N gauge theory later. We also assume that $\mathcal{A}$ is closed under the Hermitian conjugation operation.

This means that we have a von Neumann algebra $\mathcal{A}$ acting on a Hilbert space $\mathcal{H}$, which has a cyclic and separating vector $|\Psi\rangle$. The Tomita-Takesaki theorem states that in this case the "commutant" $\mathcal{A}'$ of the algebra $\mathcal{A}$ can be constructed by an antilinear conjugation, which can be identified with the "tilde" mapping used in thermofield theory.

Like in the discussion of the previous section, the Tomita-Takesaki construction starts by constructing the *antilinear* map S that appeared above,

$$SA|\Psi\rangle = A^\dagger|\Psi\rangle.$$

We consider the polar decomposition of S as

$$S = J\Delta^{1/2},$$

where J is anti-uitary, and Δ is Hermitian and positive. This can also be understood as follows. For an antilinear map S we define the Hermitian conjugate as

$$(|A\rangle, S^\dagger|B\rangle) \equiv (|B\rangle, S|A\rangle), \tag{6.4}$$

where (,) denotes the inner product. Then we have

$$\Delta = S^\dagger S. \tag{6.5}$$

It is not too difficult to prove the useful relations

$$J\Delta^{1/2} = \Delta^{-1/2}J, \qquad J^2 = 1,$$

and

$$S|\Psi\rangle = J|\Psi\rangle = \Delta|\Psi\rangle = |\Psi\rangle.$$

Finally, under the previous conditions, the Tomita-Takesaki theorem states that

$$J\mathcal{A}J = \mathcal{A}', \tag{6.6}$$

and

$$\Delta^{is}\mathcal{A}\Delta^{-is} = \mathcal{A}, \qquad s \in \mathbb{R}. \tag{6.7}$$

Equation (6.6) implies that the commutant $\mathcal{A}'$ can be recovered by conjugating the operators in $\mathcal{A}$ with the antilinear map J.

To interpret Eq. (6.7), let us first write $\Delta = e^{-K}$ where K is a Hermitian operator. Then Eq. (6.7) means that the set $\mathcal{A}$ is "closed under time evolution" with respect to the "modular Hamiltonian" K. As we will see later in the case of the large N gauge theory, and in the large N limit, the analogue of the operator K behaves like $\beta(H_{\mathrm{CFT}} - E_0)$, where $E_0 = \langle\Psi|H_{\mathrm{CFT}}|\Psi\rangle$. Hence by identifying $s = t/\beta$ we see that this equation expresses the closure of the algebra $\mathcal{A}$ under time evolution.

In order to provide some additional intuition, let us consider the usual thermofield-double construction. We start with a quantum system with spectrum $H|E_i\rangle = E_i|E_i\rangle$. We consider the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ of two identical copies of this system and place it in a special entangled state

$$|\Psi\rangle_{\mathrm{tfd}} = \frac{1}{\sqrt{Z}}\sum_i e^{-\beta E_i/2}|E_i, E_i\rangle,$$

where $Z = \sum_i e^{-\beta E_i}$. We call H_1, H_2 the two Hamiltonians. We also introduce the "thermofield Hamiltonian" defined as

$$H_{\mathrm{tfd}} = H_1 - H_2,$$

which satisfies

$$H_{\mathrm{tfd}}|\Psi\rangle_{\mathrm{tfd}} = 0.$$

It should be an easy exercise for the reader to verify the following. If we take as our algebra of "accessible observables" $\mathcal{A}$ to be the operators acting on system 1, then the conditions of the Tomita-Takesaki theorem are satisfied, i.e. the state $|\Psi\rangle_{\mathrm{tfd}}$ is cyclic and separating for the algebra $\mathcal{A}$ in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. We can thus define the operators S, J, Δ as described above. A few lines of algebra show that J turns out to be the *antilinear* map that takes

$$J\colon\ |E_i, E_j\rangle \to |E_j, E_i\rangle,$$

and

$$\Delta = \exp(-\beta(H_1 - H_2)) = \exp(-\beta H_{\mathrm{tfd}}).$$

Hence, for any operator $A \in \mathcal{A}$, i.e. for any operator acting on the first copy of the system, the "mirror operator" JAJ given by the Tomita-Takesaki construction is an operator acting on the second system and precisely coincides with what we would have defined as the dial via the usual thermofield doubling. The relation between the Tomita-Takesaki construction and the thermofield doubling has been noted before in the literature; for instance see [34].

Let us now consider the conformal field theory and consider the case where the elements of $\mathcal{A}$ are just modes of a generalized free field. The last result that we wish to show here is that Δ really does reproduce the factors of $e^{-\frac{\beta\omega_n}{2}}$ that we introduced above, at least for typical pure states. First consider the state $|\Psi'\rangle = \mathcal{O}^i{}_{\omega_n,\boldsymbol{m}}|\Psi\rangle$, where $|\Psi\rangle$ is a typical equilibrium pure state. Using expression (6.5) we have

$$\begin{aligned}\langle\Psi'|\Delta|\Psi'\rangle &= \langle\Psi|(\mathcal{O}^i_{\omega_n,\boldsymbol{m}})^\dagger\Delta\mathcal{O}_{\omega_n,\boldsymbol{m}}|\Psi\rangle \\ &= \langle\Psi|(\mathcal{O}^i_{\omega_n,\boldsymbol{m}})^\dagger S^\dagger S\mathcal{O}^i_{\omega_n,\boldsymbol{m}}|\Psi\rangle \\ &= \langle\Psi|(\mathcal{O}^i_{\omega_n,\boldsymbol{m}})^\dagger S^\dagger(\mathcal{O}^i_{\omega_n,\boldsymbol{m}})^\dagger|\Psi\rangle.\end{aligned}$$

Using the definition of the adjoint $S^\dagger$ of an antilinear operator given in (6.4) we find

$$\langle\Psi'|\Delta|\Psi'\rangle = \langle\Psi|\mathcal{O}^i_{\omega_n,\boldsymbol{m}}(\mathcal{O}^i_{\omega_n,\boldsymbol{m}})^\dagger|\Psi\rangle.$$

Now we remind the reader that typical equilibrium states in a large N CFT satisfy the Kubo-Martin-Schwinger (KMS) condition, which for the modes of generalized free fields reads

$$\langle\Psi|\mathcal{O}^i_{\omega_n,\boldsymbol{m}}(\mathcal{O}^i_{\omega_n,\boldsymbol{m}})^\dagger|\Psi\rangle = e^{-\frac{\beta\omega_n}{2}}\langle\Psi|(\mathcal{O}^i_{\omega_n,\boldsymbol{m}})^\dagger\mathcal{O}^i_{\omega_n,\boldsymbol{m}}|\Psi\rangle.$$

This was extensively reviewed in [1] where the reader can find more details. So all in all we find

$$\langle\Psi'|\Delta|\Psi'\rangle = e^{-\frac{\beta\omega_n}{2}}\langle\Psi'|\Psi'\rangle.$$

Moreover if we have two different states of the form $|\Psi_1'\rangle = \mathcal{O}_{\omega_1,\boldsymbol{m}_1}|\Psi\rangle$, $|\Psi_2'\rangle = \mathcal{O}_{\omega_2,m_2}|\Psi\rangle$ with $1 \neq 2$ (in hopefully obvious notation) we have $\langle\Psi_1{}'|\Delta|\Psi_2{}'\rangle = 0$. The reader can easily verify that, using the KMS condition and the large N factorization of the CFT, then for any two states of the form

$$|\Psi_1{}'\rangle = \mathcal{O}^{i_1}_{\omega_1,\boldsymbol{m}_1}\ldots\mathcal{O}^{i_m}_{\omega_m,\boldsymbol{m}_m}|\Psi\rangle,$$

and

$$|\Psi_2{}'\rangle = \mathcal{O}^{i_1{}'}_{\omega_1{}',\boldsymbol{m}_1{}'}\ldots\mathcal{O}^{i_n{}'}_{\omega_n{}',\boldsymbol{m}_n{}'}|\Psi\rangle,$$

we have[12]

$$\langle\Psi_1{}'|\Delta|\Psi_2{}'\rangle = e^{-\frac{\beta}{2}\sum_{i=1}^m\omega_m}\langle\Psi_1{}'|\Psi_2{}'\rangle + \left(\frac{1}{N}\text{ corrections}\right). \tag{6.8}$$

[12]Notice that Eq. (6.8) is consistent, even though it seems to break the symmetry between the number of insertions m, n in the two states $|\Psi_{1,2}{}'\rangle$. The point is that in the large N limit, both sides of the equation are zero, unless $m = n$ and the frequencies/momenta of state 1 are a permutation of those of 2.

If we are concerned with the action of Δ only in $\mathcal{H}_\Psi$, then this set of matrix elements completely specifies the operator. However, we see that the statement above is precisely the KMS condition for the state $|\Psi\rangle$. So we see that in a state $|\Psi\rangle$, in which the correlators are close to being thermal, the operator Δ behaves precisely as $e^{-\beta(H_{\rm CFT}-E_0)}$, where $E_0 = \langle\Psi|H_{\rm CFT}|\Psi\rangle$, and this produces the $e^{-\frac{\beta\omega}{2}}$ factors that we required above.

We should caution the reader that in a real state in the CFT, which might correspond to a black hole, the condition (6.8) might receive corrections at subleading order in $\frac{1}{N}$. These corrections might have an effect on the eigenvalues of Δ in the case where the set $\mathcal{A}$ itself has a size that scales with N. We leave an investigation of these $\frac{1}{N}$ effects to further work.

C. Finite-dimensional algebras

In this subsection we specialize to the case where $\mathcal{A}$ is a finite-dimensional closed subalgebra, which is acting on a system with Hilbert space $\mathcal{H}$. This Hilbert space may be infinite dimensional. We assume that the algebra $\mathcal{A}$ is closed under Hermitian conjugation. We will find that the Tomita-Takesaki construction reduces to the construction of the mirror operators defined in [1]. We *do not assume* that the system has necessarily a bipartite structure.

The system is taken to be in a pure state $|\Psi\rangle$. We consider the span $\mathcal{H}_\Psi$ of states of the form $\mathcal{A}|\Psi\rangle$. If the dimensionality of the algebra $\mathcal{A}$ is n, then $\mathcal{H}_\Psi$ is an n-dimensional subspace of the full Hilbert space.

The interesting part of everything that follows will take place in this finite-dimensional space. While the algebra $\mathcal{A}$ is acting on $\mathcal{H}_\Psi$ it is clear that there are many other operators which can act on the space $\mathcal{H}_\Psi$. In fact, the dimensionality of the algebra $\mathcal{A}$ is n while the dimensionality of the algebra $\mathcal{B}(H_\Psi)$ of *all* operators acting on $\mathcal{H}_\Psi$ is n^2. We will argue that this is precisely related to the fact that *on the same space* $\mathcal{H}_\Psi$ we can naturally define the action of a second, commuting copy of the algebra $\mathcal{A}$; let us call it $\mathcal{A}'$ such that

$$B(\mathcal{H}_\Psi) = \mathcal{A}\otimes\mathcal{A}'.$$

The construction proceeds exactly as before. In this case, all operators that we encounter are finite dimensional, so it is very easy to check all steps in our argument explicitly; see Appendix D for technical details.

Again, we assume that the state $|\Psi\rangle$ appears sufficiently entangled with respect to the algebra $\mathcal{A}$, which means that

$$A|\Psi\rangle = 0 \Leftrightarrow A = 0.$$

This allows us to define the antilinear map $S{:}\,\mathcal{H}_\Psi \to \mathcal{H}_\Psi$ defined by

$$SA|\Psi\rangle = A^\dagger|\Psi\rangle.$$

We also introduce its adjoint $S^\dagger$, which—due to the fact that S is antilinear—is defined by

$$(|A\rangle, S^\dagger|B\rangle) = (|B\rangle, S|A\rangle).$$

Using these two operators we consider the linear operator $\Delta\colon \mathcal{H}_\Psi \to \mathcal{H}_\Psi$ defined by

$$\Delta = S^\dagger S.$$

It is easy to show that Δ is a positive, Hermitian operator. It is related to what we would get from the polar decomposition of S as

$$S = J\Delta^{\frac{1}{2}},$$

with J anti-uitary. Equivalently we can just define the antilinear operator

$$J = S\Delta^{-\frac{1}{2}}.$$

As explained in Appendix D we can check that J satisfies

$$J^2 = 1.$$

We also have the important relations

$$S|\Psi\rangle = J|\Psi\rangle = \Delta|\Psi\rangle = |\Psi\rangle.$$

Now we define the mirror operators as operators acting on the Hilbert space $\mathcal{H}_\Psi$ by the relation

$$\tilde{A}_i = JA_iJ.$$

Using the fact that $J^2 = 1$ we find that the antilinear "tilde" mapping

$$\tilde{}\colon\ A \to \tilde{A}$$

is an algebra *-isomorphism, i.e. the mirror operators satisfy the same commutation relations as the original operators up to a conjugation of the structure constants, so if

$$[A_i, A_j] = f_{ij}^k A_k,$$

then

$$[\tilde{A}_i, \tilde{A}_j] = (f_{ij}^k)^* \tilde{A}_k.$$

Moreover, as we demonstrate in Appendix D, the operators in $\mathcal{A}$ commute with the mirror operators

$$[A_i, \tilde{A}_j] = 0.$$

Hence $J\mathcal{A}J \in \mathcal{A}'$. We can also prove that any operator in $A' \in \mathcal{A}'$ can be written in the form $A' \in JAJ$ for some $A \in \mathcal{A}$. Hence we have that $J\mathcal{A}J = \mathcal{A}'$.

Let us now consider correlation functions. First, using $J^2 = 1$ and $J|\Psi\rangle = |\Psi\rangle$, we find that the dual-dual correlators are related to the original correlators by

$$\langle\Psi|\tilde{A}_1 \ldots \tilde{A}_n|\Psi\rangle = \langle\Psi|A_1 \ldots A_n|\Psi\rangle^*,$$

and the mixed correlators obey

$$\langle\Psi|A\tilde{B}|\Psi\rangle = \langle\Psi|A\Delta^{\frac{1}{2}}B^\dagger|\Psi\rangle.$$

The summary is that we can have on the subspace $\mathcal{H}_\Psi$ the action of the original algebra $\mathcal{A}$ together with the action of a second *-isometric copy $\tilde{\mathcal{A}} = J\mathcal{A}J$, which commutes with $\mathcal{A}$. As discussed earlier, the mirror operators can be extended in the full Hilbert space $\mathcal{H}$ in more than one way (for instance by taking them to be "zero" on $\mathcal{H}_\Psi^\perp$.) The details of this extension do not affect correlation functions of $\mathcal{A}$'s and $\tilde{\mathcal{A}}$'s evaluated on the state $|\Psi\rangle$.

1. Bipartite system

We now demonstrate that, in the case where the system is bipartite, the S, J, Δ construction above is equivalent to the more direct construction of the mirror operators described [1].

Suppose we have a system which is bipartite $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, with $\dim \mathcal{H}_A \leq \dim \mathcal{H}_B$. We take the algebra $\mathcal{A}$ to be the algebra of operators acting on $\mathcal{H}_A$. We take the system in a pure state $|\Psi\rangle$, which is generally entangled. As before we assume that the entanglement is sufficiently large so that

$$A|\Psi\rangle = 0 \Leftrightarrow A = 0.$$

As will become clear below, this condition is equivalent to the condition that the reduced density matrix

$$\rho_A = \mathrm{Tr}_B(|\Psi\rangle\langle\Psi|)$$

is of maximal rank.

Definition of mirror operators according to [1].—The pure state of the entire system can be expanded in a general orthonormal basis as

$$|\Psi\rangle = \sum_{ij} c_{ij} |i\rangle_A \otimes |j\rangle_B.$$

We consider a (state-dependent) change of basis to bring the state in the Schmidt form

$$|\Psi\rangle = \sum_{i=1}^{n_A} d_i |i\rangle_A^\Psi \otimes |i\rangle_B^\Psi,$$

where we can take d_i to be real and ≥ 0. Here $n_a = \dim \mathcal{H}_A$. We have explicitly written the Ψ superscript in $|i\rangle_A^\Psi, |i\rangle_B^\Psi$ to denote that these states depend on the choice of the pure state $|\Psi\rangle$.

The reduced density matrix for system A is

$$\rho_A = \begin{pmatrix} d_1^2 & 0 & \dots \\ 0 & d_2^2 & \dots \\ \dots & & \\ \dots & \dots & d_{n_A}^2 \end{pmatrix}.$$

We assumed that the entanglement of the original state $|\Psi\rangle$ is sufficiently large, so that the matrix ρ_A has maximal rank. Hence $d_i > 0$.

In this case we defined the mirror operators as follows. For any operator acting on system A of the form

$$A = A_{ij} |i\rangle_A^{\Psi} {}_A^{\Psi}\langle j| \otimes \mathcal{I}_B,$$

we defined the mirror operator acting on B as

$$\tilde{A} = A_{ij}^* \mathcal{I}_A \otimes |i\rangle_B^{\Psi} {}_B^{\Psi}\langle j|. \tag{6.9}$$

Notice that this operator has nonvanishing matrix elements only along a ($\dim \mathcal{H}_A$)-dimensional subspace of the Hilbert space $\mathcal{H}_B$—i.e. it is a sparse operator. In the language of the previous subsections, this corresponds to the choice of taking the mirror operators to be zero on the subspace orthogonal to $\mathcal{H}_\Psi$.

The mirror operators according to the S, J, Δ Tomita-Takesaki construction.—Let us now see how the S, J, Δ construction leads to the same result. First we start with the state $|\Psi\rangle$ and consider the linear space

$$\mathcal{H}_\Psi = \mathcal{A}|\Psi\rangle.$$

This is an n_A^2-dimensional subspace of the full Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. Everything which will follow will be defined on this space. First we define the antilinear map

$$SA|\Psi\rangle = A^\dagger |\Psi\rangle, \tag{6.10}$$

and we also introduce $\Delta = S^\dagger S$. We then define the antilinear $J = S\Delta^{-1/2}$ and the mirror operators as

$$\tilde{A} = JAJ.$$

We will show that this definition coincides with the one above.

This construction is manifestly basis independent. We can thus apply it on a convenient basis. We select the Schmidt basis

$$|\Psi\rangle = \sum_{i=1}^{n_A} d_i |i\rangle_A^\Psi \otimes |i\rangle_B^\Psi.$$

Any operator A acting on $\mathcal{H}_A$ acts as

$$A|\Psi\rangle = \sum_{i,k=1}^{n_A} d_i A_{ki} |k\rangle_A^\Psi \otimes |i\rangle_B^\Psi,$$

while the Hermitian conjugate acts as

$$A^\dagger|\Psi\rangle = \sum_{i,k=1}^{n_A} d_i A_{ik}^* |k\rangle_A^\Psi \otimes |i\rangle_B^\Psi.$$

We find that the antilinear operator S which implements the modular conjugation (6.10) is

$$S|i\rangle_A^\Psi \otimes |j\rangle_B^\Psi = \frac{d_i}{d_j} |j\rangle_A^\Psi \otimes |i\rangle_B^\Psi,$$

and we see that

$$S^\dagger|i\rangle_A^\Psi \otimes |j\rangle_B^\Psi = \frac{d_j}{d_i} |j\rangle_A^\Psi \otimes |i\rangle_B^\Psi.$$

Hence we find

$$\Delta|i\rangle_A^\Psi \otimes |j\rangle_B^\Psi = \frac{d_i^2}{d_j^2} |i\rangle_A^\Psi \otimes |j\rangle_B^\Psi.$$

Hence we notice that the states $|i\rangle_A \otimes |j\rangle_B$ in the Schmidt basis are eigenstates of Δ. We can easily see that we can express

$$\Delta = \rho_A \otimes \rho_B^{-1},$$

and we have that J is the antilinear map which is defined by

$$J(|i\rangle_A^\Psi \otimes |j\rangle_B^\Psi) = |j\rangle_A^\Psi \otimes |i\rangle_B^\Psi.$$

Hence the mirror operators are defined by

$$\tilde{A} = JAJ.$$

So again we find

$$A = A_{ij} |i\rangle_A^{\Psi} {}_A^{\Psi}\langle j| \otimes \mathcal{I}_B.$$

We defined the mirror operator acting on B as

$$\tilde{A} = A_{ij}^* \mathcal{I}_A \otimes |i\rangle_B^{\Psi} {}_B^{\Psi}\langle j|.$$

This coincides with the definition (6.9) according to [1].

2. Construction in terms of projection operators

We also present one final (equivalent) way to look at this construction. Consider the algebra $\mathcal{A}$ with which we are probing the system. We would like to select a Cartan subalgebra $\mathcal{A}_a$, which we will use to "label" states by the collective eigenvalues a. Of course there are many possible ways to select the Cartan subalgebra. The point is that the way in which a state $|\Psi\rangle$ is aligned relative to the algebra $\mathcal{A}$ selects a particular preferred choice for the Cartan subalgebra, in which the "entanglement is diagonalized."

Hence, for any given system and algebra $\mathcal{A}$, the preferred choice of the Cartan subalgebra depends on the state $|\Psi\rangle$. This in turn depends on what we call a "typical state" i.e. what is the *ensemble* that we are considering. In particular, if we want to consider a microcanonical ensemble then it is the Hamiltonian which determines the ensemble and which finally selects the preferred orientation of the Cartan subalgebra. Hence the choice of this Cartan subalgebra is a dynamical question. As we will see later, in the case of the large N gauge theory, and if we think of $\mathcal{A}$ as the algebra of single-trace operators, the *dynamics of the CFT* implies that the entanglement of a typical microstate selects as the preferred orientation of the Cartan subalgebra the one generated by the "occupation number operators" $N_{\omega,\boldsymbol{k}}$ of the various modes. After these generalities, let us now see in more detail how the mirror operator construction works in this language. Suppose we select a particular Cartan subalgebra for $\mathcal{A}$. Consider projection operators P_a on the eigenspaces of the Cartan subalgebra. The original state can be written as

$$|\Psi\rangle = \sum_a P_a|\Psi\rangle = \sum_a d_a|a\rangle_\Psi,$$

where[13]

$$|a\rangle_\Psi = \frac{P_a|\Psi\rangle}{||P_a|\Psi\rangle||}, \qquad d_a = ||P_a|\Psi\rangle||.$$

With this normalization, and since states of different eigenvalue a are orthogonal, we have

$${}_\Psi\langle a|a'\rangle_\Psi = \delta_{aa'}.$$

It is clear that by acting with elements of the algebra $\mathcal{A}$, we can map a state of eigenvalues a to a state with eigenvalues b. This can be achieved by acting on the original state with an appropriate combination of operators from $\mathcal{A}$. We call this combination of operators T_{ba}. Since we have assumed that $A|\Psi\rangle = 0$ implies $A = 0$, we can see that for any possible transition $a \to b$, there is a unique choice of T_{ba} (up to overall multiplicative constant). We then define the following set of states:

$$|b,a\rangle_\Psi = T_{ba}|a\rangle_\Psi.$$

We select the normalization of T_{ba} so that all these states have unit norm. However, they are not necessarily orthogonal. In general we have ${}_\Psi\langle b,a|b',a'\rangle_\Psi = \delta_{bb'} f_{aa'}$. The point now is that by a particular choice of the Cartan algebra, we can achieve that the the entanglement is "diagonalized" in the sense that $f_{aa'} = \delta_{aa'}$. Of course this problem is closely related to the Schmidt diagonalization. From now on we assume that we have aligned our Cartan algebra so that

$${}_\Psi\langle b,a|b',a'\rangle_\Psi = \delta_{bb'}\delta_{aa'}.$$

In this case, the original pure state can be written as

$$|\Psi\rangle = \sum_a d_a|a,a\rangle_\Psi,$$

and schematically we see this in Fig. 9. One can check that operators from the algebra $\mathcal{A}$ act on this state as

$$A = \sum_{a,a',b} A_{aa',b}|a,b\rangle_{\Psi\Psi}\langle a',b|.$$

We define the corresponding mirror operator as

$$\tilde{A} = \sum_{a,a',b} A^*_{aa',b}|b,a\rangle_{\Psi\Psi}\langle b,a'|.$$

We see a graphical representation of this in Fig. 10. We hope it is clear to the reader that this definition of the mirror operators is completely equivalent to the previous definitions.

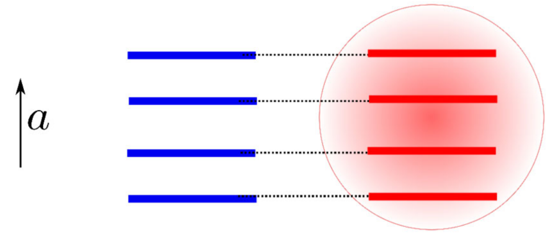

FIG. 9 (color online). A typical pure state $|\Psi\rangle$ expanded in eigenstates of a Cartan subalgebra $\mathcal{A}_a$ of $\mathcal{A}$, selected so that the entanglement appears "diagonal." Here a denotes the collective eigenvalues of $\mathcal{A}_a$.

D. "Truncated algebras" and complementarity

In the previous subsections we described the doubling in the case where the set of "accessible observables" $\mathcal{A}$ forms a closed algebra under multiplication. In that case the construction of the mirror operators was rather straightforward.

In this subsection we come to the more interesting case where the set of observables has a dual role: if it is truncated to a small subset, then we are naturally lead to the "doubling" and the introduction of mirror operators for this subset. If it is not truncated, then the doubling is impossible and we can see that, what used to be the mirror

[13]The fact that all $d_a > 0$ follows from the assumption that $A|\Psi\rangle \neq 0$ for $A \neq 0$.

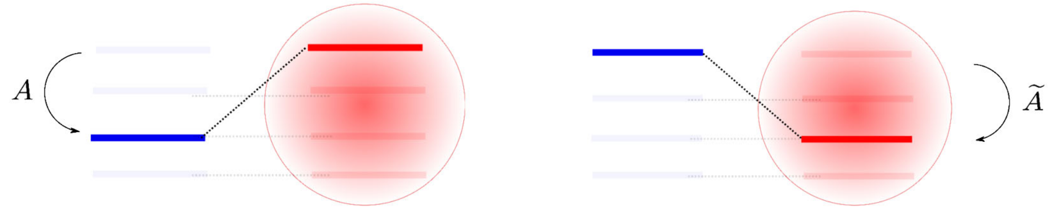

FIG. 10 (color online). Left: while we started with n states, we can construct n^2 states by acting with operators in $\mathcal{A}$. Right: the mirror operators can be defined as causing transitions between these n^2 states.

operators before, can in reality be expressed as complicated combinations of the observables. Of course this is the operator-language version of the idea of black hole complementarity.

To be more precise, the case which is more relevant for us is when we have a set of "basic observables" $\{A_1, \ldots A_n\}$ with which we can probe the system. Since we want to compute correlation functions of these observables, it means that we also have to consider products of them. By considering an unrestricted number of products, we do get a closed algebra generated by A_i.

But suppose that we do not want to consider this entire full algebra, but rather that we want to consider the case where we probe the system only with a "few insertions" of the basic observables. Hence we want to include in our observables products of the A_i's as long as the number of factors does not get "too large." This requires some sort of regularization. A crude regulator would be to impose a hard cutoff k in the number of insertions of the basic observables. For any choice of this regularization, the set $\mathcal{A}$ of allowed combinations of the basic observables is not a proper algebra, since it is not closed under multiplication in a strict mathematical sense.

In a large N gauge theory, the large N scaling provides us with a natural intuitive definition of these ideas: the "basic observables" are the single-trace operators, and the allowed set of observables are products of $O(1)$ numbers of these operators. For any choice of the regulated set $\mathcal{A}$ we can define the mirror operators by following a slight modification of the procedure mentioned in the previous section. This leads to a definition of mirror operators which depends on the size of the regulated set $\mathcal{A}$, or relatedly, on the value of the cutoff k.

We need to be careful about the regime of validity of this construction. We want to take the cutoff k to be large, but not too large—otherwise we "run out of space" in the Hilbert space and we completely deplete the entanglement. Whether this construction is sensible/useful depends on the actual physical system under consideration. What we need to establish is that the mirror operators are "robust" under a change of the cutoff, when the cutoff is large but not too large.

To summarize, the realization of black hole complementarity in operator language is the statement that we have a set of "basic observables" $\{A_1, \ldots A_n\}$, which have the property that

(1) when we are considering low-point correlations of these observables, we can define the dual operators via the aforementioned construction;

(2) when we are considering arbitrarily high-order correlations, then we cannot—this is due to the fact that the full algebra generated by arbitrary products of A_i is so large, that with respect to this large algebra the state $|\Psi\rangle$ *does not look entangled* any more, and there is no space/no need to define the dual operators.

In that sense the dual operators can be understood as very complicated combinations of the basic observables A_i. This is in line with the paradigm of black hole complementarity.

E. Large N gauge theories

In the case of the large N gauge theory, we have $|\Psi\rangle$, a typical black hole microstate [i.e. a state of the CFT with energy of order $O(N^2)$]. The set $\mathcal{A}$ is the vector space spanned by light operators. For example, as we mentioned above and in [12], we could take the set to be spanned by the set of polynomials in the modes of $\mathcal{O}^i_{n,\boldsymbol{m}}$ with an upper bound on the energy, excluding the zero modes of conserved currents, which we return to below.

Alternately, we can also consider polynomials in just single-trace operators, and in an interacting theory, this should produce an equivalent set. As we have mentioned, in the $SU(N)$ theory, we can even consider a product of up to N-single-trace operators.

The set $\mathcal{A}$ is not a proper algebra, because we have imposed the restriction that the number of insertions of single-trace operators should not be too large. Let us call k the effective cutoff in how many operators we allow. This defines the Hilbert space $\mathcal{H}^k_\Psi = \{\text{span}\colon \mathrm{A}|\Psi\rangle\}$ we have included the superscript k and the subscript Ψ to indicate explicitly that this Hilbert space depends on the microstate $|\Psi\rangle$ and the cutoff k. It is a small subspace of the full Hilbert space $\mathcal{H}_{\text{CFT}}$.

We call P^k the projection operator on $\mathcal{H}^k_\Psi$. By acting with these projection operators on the operators of $\mathcal{A}$ (i.e. replacing $A \to P^k A P^k$) we get a deformation of the set $\mathcal{A}$ into an algebra. Using this algebra we can define the

S, J, Δ operators. It is clear that the matrix elements of S, Δ between states which do not carry too many excitations relative to $|\Psi\rangle$ are robust under scaling the cutoff k, and we will discuss this further in upcoming work. In the large N limit the modular operator Δ coincides with $e^{-\beta(H_{\rm CFT}-E_0)}$ where $E_0 = \langle\Psi|H_{\rm CFT}|\Psi\rangle$. Hence the correlators that we will get by following the Tomita-Takesaki construction are *to leading order in large* $\mathcal{N}$, the same as the thermofield correlators.

F. Conserved charges

Now, we describe how S acts on conserved charges, including insertions involving *polynomials* of charges, $\mathcal{Q}_\alpha$. As usual, by a conserved charge, we mean any operator that commutes exactly with the Hamiltonian, but we consider non-Abelian symmetries as well here.

As we noted in Sec. III B 4, we can always move the charges so that they act directly on the state $|\Psi\rangle$. On such states, we define

$$SA_\alpha\mathcal{Q}_\beta|\Psi\rangle = A_\alpha^\dagger S\mathcal{Q}_\beta|\Psi\rangle. \tag{6.11}$$

We emphasize that (6.11) is valid only when the charges are immediately next to the state. We have not yet defined the action of S on the space of states produced by acting with the charge-polynomials on the base state $|\Psi\rangle$, which we discuss below. However, since $S^2 = 1$, we see that even without specifying the action of S on the charge polynomial, we immediately obtain equivalence with (3.26).

First, let us check this fact: (6.11) reproduces (3.26). We have

$$\begin{aligned} JO^i_{n,\boldsymbol{m}}JA_\alpha\mathcal{Q}_\beta|\Psi\rangle &= S\Delta^{-\frac{1}{2}}O^i_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}SA_\alpha\mathcal{Q}_\beta|\Psi\rangle \\ &= S\Delta^{-\frac{1}{2}}O^i_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}A_\alpha^\dagger S\mathcal{Q}_\beta|\Psi\rangle \\ &= A_\alpha\Delta^{\frac{1}{2}}(O^i_{n,\boldsymbol{m}})^\dagger\Delta^{-\frac{1}{2}}\mathcal{Q}_\beta|\Psi\rangle. \end{aligned}$$

Using the fact shown above that $\Delta \approx e^{-\beta(H_{\rm CFT}-E_0)}$ at large N, we see that we precisely reproduce (3.26).

Now, we return to the definition of S on the space produced by acting with charge polynomials on $|\Psi\rangle$. We denote this space by $\mathcal{V}_\mathcal{Q} = \text{span}\{\mathcal{Q}_{\beta_i}|\Psi\rangle\}$. We need to perform three checks on the action of the map $S\colon \mathcal{V}_\mathcal{Q} \to \mathcal{V}_\mathcal{Q}$.

(1) On eigenstates, where $\mathcal{Q}_\beta|\Psi\rangle = Q_\beta|\Psi\rangle$, we have $S\mathcal{Q}_\beta|\Psi\rangle = Q_\beta^*|\Psi\rangle$.

(2) On null states, where $\mathcal{Q}_{n_i}|\Psi\rangle = 0$, we have $S\mathcal{Q}_{n_i}|\Psi\rangle = 0$.

(3) $S^2\mathcal{Q}_\beta|\Psi\rangle = \mathcal{Q}_\beta|\Psi\rangle$.

In fact, these three conditions do not uniquely fix the action of S on $\mathcal{V}_\mathcal{Q}$. For example, one possible definition of S on $\mathcal{V}_\mathcal{Q}$ is as follows. Let the vectors

$$\{|\Psi\rangle, \mathcal{Q}_{n_1}|\Psi\rangle \ldots \mathcal{Q}_{n_P}|\Psi\rangle, \mathcal{Q}_{b_1}|\Psi\rangle \ldots \mathcal{Q}_{b_M}|\Psi\rangle\},$$

form a basis for $\mathcal{V}_\mathcal{Q}$ so that

$$\mathcal{Q}_\beta|\Psi\rangle = \kappa_\beta^1|\Psi\rangle + \left(\sum_{i=1}^P \kappa_\beta^{n_i}\mathcal{Q}_{n_i}|\Psi\rangle\right) + \left(\sum_{i=1}^M \kappa_\beta^{b_i}\mathcal{Q}_{b_i}|\Psi\rangle\right).$$

Then, we can define

$$S\mathcal{Q}_\beta|\Psi\rangle = (\kappa_\beta^1)^*|\Psi\rangle + \left(\sum_{i=1}^M (\kappa_\beta^{b_i})^*\mathcal{Q}_{b_i}|\Psi\rangle\right).$$

This meets all three criterion above, but clearly we can redefine the S map, by changing the basis. Such a redefinition does not affect the definition of the mirror operators.

Abelian conserved charges and eigenstates.—We now specialize to Hermitian $U(1)$ charges, and to the situation where the state $|\Psi\rangle$ is an eigenstate of such a charge. We include the Hamiltonian $H_{\rm CFT}$ in this discussion, and so the discussion here is also applicable to energy eigenstates. As we see below, the action of S simplifies in this situation.

We start by considering a situation where the state $|\Psi\rangle$ satisfies

$$H_{\rm CFT}|\Psi\rangle = E_0|\Psi\rangle.$$

Now, we find that

$$SH_{\rm CFT}A_1|\Psi\rangle = S([H_{\rm CFT}, A_1] + A_1H_{\rm CFT})|\Psi\rangle.$$

Without even assuming that A_1 has a well-defined energy, we can write $[A_1, H_{\rm CFT}] = A_1^H$, which, we assume, is in the set $\mathcal{A}$. So

$$SH_{\rm CFT}A_1|\Psi\rangle = (-(A_1^H)^\dagger + A_1^\dagger E_0)|\Psi\rangle.$$

We see that this follows automatically if we use

$$\begin{aligned} SH_{\rm CFT}A_p|\Psi\rangle &= (2E_0 - H_{\rm CFT})SA_p|\Psi\rangle; \\ JH_{\rm CFT}A_p|\Psi\rangle &= (2E_0 - H_{\rm CFT})JA_p|\Psi. \end{aligned} \tag{6.12}$$

This has an interesting consequence. As we show in Appendix D, we have the relation

$$J\Delta = \Delta^{-1}J.$$

Using the relation between Δ and the CFT Hamiltonian that we described above $\Delta = e^{-\beta(H_{\rm CFT}-E_0)}$, we see that this is precisely consistent with (6.12).

We have the same relation between the map S and any other $U(1)$ conserved charge $\hat{Q}$, when the state is in a charge eigenstate.

$$S\hat{Q}A_p|\Psi\rangle = (2Q_0 - \hat{Q})SA_p|\Psi\rangle,$$

where $\hat{Q}|\Psi\rangle = Q_0|\Psi\rangle$. So, we see that our choice of gauge follows naturally in the Tomita-Takesaki construction.

Now, let us show directly that this is equivalent to (3.19). Moving to the operators that transform simply under the charge, which were defined near (3.19), we first note that

$$\begin{aligned}\tilde{\mathcal{O}}^{i,q}_{n,\boldsymbol{m}} A_1 \hat{Q} A_2 |\Psi\rangle &= J\mathcal{O}^{i,q}_{n,\boldsymbol{m}} J A_1 \hat{Q} A_2 |\Psi\rangle \\ &= J\mathcal{O}^{i,q}_{n,\boldsymbol{m}} J([A_1,\hat{Q}] + \hat{Q}A_1)A_2|\Psi\rangle \\ &= J\mathcal{O}^{i,q}_{n,\boldsymbol{m}} J(A_1^q + \hat{Q}A_1)A_2|\Psi\rangle \\ &= J\mathcal{O}^{i,q}_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}(A_2^\dagger (A_1^q)^\dagger + (2Q_0 - \hat{Q})A_2^\dagger A_1^\dagger)|\Psi\rangle,\end{aligned}$$

where $A_1^q = [A_1, \hat{Q}]$. Now, we use the fact that $[\mathcal{O}^{i,q}_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}, \hat{Q}] = -q\mathcal{O}^{i,q}_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}$, which is precisely how we defined q in (3.19), and we have additionally used that Δ commutes with $\hat{Q}$.

Substituting this relation above, we see that

$$\begin{aligned}\tilde{\mathcal{O}}^{i,q}_{n,\boldsymbol{m}} A_1 \hat{Q} A_2 |\Psi\rangle &= J\mathcal{O}^{i,q}_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}(A_2^\dagger (A_1^q)^\dagger + 2Q_0 A_2^\dagger A_1^\dagger)|\Psi\rangle \\ &\quad - J\hat{Q}\mathcal{O}^{i,q}_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}A_2^\dagger A_1^\dagger|\Psi\rangle \\ &= J\mathcal{O}^{i,q}_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}(A_2^\dagger (A_1^q)^\dagger + (2Q_0 - q)A_2^\dagger A_1^\dagger)|\Psi\rangle \\ &\quad - J\hat{Q}\mathcal{O}^{i,q}_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}A_2^\dagger A_1^\dagger|\Psi\rangle \\ &= J\mathcal{O}^{i,q}_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}(A_2^\dagger (A_1^q)^\dagger - qA_2^\dagger A_1^\dagger)|\Psi\rangle \\ &\quad + \hat{Q}J\mathcal{O}^{i,q}_{n,\boldsymbol{m}}\Delta^{\frac{1}{2}}A_2^\dagger A_1^\dagger|\Psi\rangle \\ &= ((A_1^q)A_2 - qA_1^\dagger A_2^\dagger)\Delta^{\frac{1}{2}}(\mathcal{O}^{i,q}_{n,\boldsymbol{m}})^\dagger|\Psi\rangle \\ &\quad + \hat{Q}A_1 A_2\Delta^{\frac{1}{2}}(\mathcal{O}^{i,q}_{n,\boldsymbol{m}})^\dagger|\Psi\rangle \\ &= A_1(\hat{Q} - q)A_2\Delta^{\frac{1}{2}}(\mathcal{O}^{i,q}_{n,\boldsymbol{m}})^\dagger|\Psi\rangle,\end{aligned}$$

which agrees precisely with (3.19). Note, in particular, that the terms involving Q_0 have canceled, and are not of significance for the definition of the mirror operators, which always come with *two* J's.

Link with the Hartle-Hawking description of the state.— Before we close this section, let us take this opportunity to make a link to the usual Hartle-Hawking state for the bulk modes. The Hartle-Hawking state is sometimes written as an entangled state of free-field "modes" outside and inside the horizon, and we would like to make this precise here.

First, note that given the single-trace operators $\mathcal{O}^i_{n,\boldsymbol{m}}$ in the set $\mathcal{A}$, we can form the number operators: $N_{n,\boldsymbol{m}} = G_\beta^{-1}(n,\boldsymbol{m})\mathcal{O}^\dagger_{n,\boldsymbol{m}}\mathcal{O}_{n,\boldsymbol{m}}$, precisely as we did in Sec. IV C and IV D, where G has the same meaning as there.

These operators effectively commute with the Hamiltonian, $[N_{n,\boldsymbol{m}}, H] \approx \mathrm{O}(\frac{1}{\mathcal{N}})$, and also with each other $[N_{n,\boldsymbol{m}}, N_{n',\boldsymbol{m}'}] \approx \mathrm{O}(\frac{1}{\mathcal{N}})$. In a sense, these operators are describing the excitation of "particles" above the black hole state.

The eigenvalues of these operators $N_{n,\boldsymbol{m}}$, are integral, and for each such eigenvalue p, we can construct the projector $P^p_{n,\boldsymbol{m}}$, which projects onto a state with a definite eigenvalue[14] of $N_{n,\boldsymbol{m}}$

$$|p,n,\boldsymbol{m}\rangle_\Psi = \frac{P^p_{n,\boldsymbol{m}}|\Psi\rangle}{||P^p_{n,\boldsymbol{m}}|\Psi\rangle||}.$$

These states $|p,n,\boldsymbol{m}\rangle$ do *not* satisfy (3.2) or (6.1). We cannot construct the mirror operators on these states, and these are precisely the *firewall* states.

States of different eigenvalues p are approximately orthogonal. The original pure state can be written as

$$|\Psi\rangle = \sum_p P^p_{n,\boldsymbol{m}}|\Psi\rangle = \sum_p d_p|p,n,\boldsymbol{m}\rangle_\Psi.$$

It is important that this superposition of states has an interpretation as a smooth geometry although the individual states in the sum above do not.

It is also useful to estimate the spread of p. At large $\mathcal{N}$, we expect that with $Z = \sum_p e^{-\beta\sum p\omega_n}$ the coefficients d_p satisfy

$$|d_p|^2 = \frac{1}{Z}e^{-\beta\omega_n p}.$$

Now, if we take p to be very large, say $p = N$, then the formula for d_p is not really valid, but this formula suggests that while we expect d_p to be exponentially suppressed it should still be nonzero.

Put another way, we expect that the original state $|\Psi\rangle$ contains a spread of "number eigenvalues" that is rather large. This has an immediate implication for (3.2). For example, if we try and annihilate the state $|\Psi\rangle$ by acting with a polynomial in the number operator, $\prod_{j=1}^{j_{\max}}(N_{n,\boldsymbol{m}} - j)$, then we find that we must take $j_{\max}$ to scale with $\mathcal{N} \propto N^2$ before the polynomial annihilates the state.

Let us also briefly mention the link to the usual Hartle-Hawking state, which is often written as an entangled state of free-field modes. As we mentioned in footnote [14], we can still approximately diagonalize some O(1) set of modes centered around frequencies $\omega_1,\ldots\omega_p$.

As above, we construct mirror operators for $b_{n_1,\boldsymbol{m}_1}\ldots b_{n_p,\boldsymbol{m}_p}$ and for $b^\dagger_{n_1,\boldsymbol{m}_1}\ldots b^\dagger_{n_p,\boldsymbol{m}_p}$. Then, for these modes (ignoring their interaction outside this set), the state of the CFT *appears* to be in the Hartle-Hawking state

$$|\Psi\rangle_{\mathrm{HH}} = \frac{1}{\sqrt{Z}}\sum_{p_i} e^{-\beta\omega_i p_{i/2}}|\{p_i\}\rangle|\{\tilde{p}_i\}\rangle.$$

We have been careful to restrict the set of modes to an O(1) set, to avoid complications that occur with the

[14]We cannot *simultaneously* project all the $N_{n,\boldsymbol{m}}$ onto their eigenstates. There are $\mathrm{O}(\omega_{\min}^{-1})$ different regularized frequencies, and a simultaneous projection requires us to multiply this many projectors, where $\frac{1}{\mathcal{N}}$ effects become very important.

interaction of these modes within themselves. However, as we mentioned in Sec. III C, for the Hawking gas produced by an evaporating black hole there is a description in terms of a Fock space of an $O(N)$ set of modes. It is clear, in that case, that the Hilbert space is not large enough to literally allow for the existence of a mirror operator for each mode. But, as we have discussed many times above, these mirror operators exist in a state-dependent sense and have precisely the correct properties unless we look at correlators with too many insertions.

VII. CONCLUSIONS AND DISCUSSION

In this paper, we have shown that if we allow the mapping between boundary operators and local bulk operators to depend on the state of the theory, then *all* the recently articulated arguments in favor of structure at the horizon are effectively resolved.

We described in Sec. II that the issue of whether the black hole interior is smooth or not could be reduced to an issue of whether the light degrees of freedom of a single CFT could be effectively doubled in a thermal state. We showed explicitly how this could be done.

Our construction relies on the simple philosophy that only low-point correlators of light operators (where the number of insertions does not scale with $\mathcal{N}$) could be interpreted in terms of correlators of local perturbative fields. So, the "doubled" operators that we need also need to have the correct behavior only *within* such correlators.

We showed in Sec. III that this imposes a set of linear constraints on the operators, that is much smaller than the dimension of the Hilbert space that we are working within. These constraints lead to a set of consistent linear equations in a state that is close to a thermal state, since such a state cannot be annihilated by the action of a small number of single-trace operators. We wrote down an explicit solution to these equations in (3.10). Hence, it is possible to effectively double the number of degrees of freedom.

As we showed in Sec. IV, this completely resolves all issues that might suggest the presence of structure at the horizon. We showed how to resolve the strong subadditivity paradox, while making the commutators of operators inside and outside the horizon vanish exactly within low-point correlators. We also explained why the "creation" operators did not need to have a left-inverse inside the horizon, by pointing out that their commutation relations with the corresponding "annihilation" operators had to be obeyed within correlation functions, and not necessarily as operator relations. We also showed that our construction allowed an explicit computation of the expectation value of N_a—the particle number, as observed by the infalling observer—with the result that $N_a = 0$. The argument of [9] breaks down for state-dependent operators.

We can already study time-dependent correlators about equilibrium states with our construction, including those where the horizon of a black hole is excited. However, we also showed how to extend our construction to cases where the mirror operators are built directly on top of nonequilibrium states and showed that this gave us results that were completely consistent with semiclassical intuition.

We also pointed out that our construction, modulo some technical features having to do with the presence of conserved charges in the CFT, was the same as the well-known Tomita-Takesaki construction that has played an important role in the mathematical quantum field theory literature.

We are left with the issue of whether state dependence must be allowed, even in principle. Although various other authors have explored subtleties in these arguments, which may eventually invalidate them, the arguments of [9] strongly suggest that it is not possible to find state-independent operators behind the horizon,

In this paper, we have investigated how these arguments break down, if we allow a state-dependent mapping between the bulk and boundary operators. As we mentioned the state dependence of our operators is somewhat similar to the state dependence of the density matrix in a given state: $\rho = |\Psi\rangle\langle\Psi|$. The density matrix can be treated as an ordinary operator, and given the density matrix for some state $|\Psi\rangle$, nothing prevents us from considering its action on another state $|\Psi'\rangle$, or evaluating $\langle\Psi'|\rho|\Psi'\rangle$. In this sense the density matrix is a usual operator in the Hilbert space. However, it has a useful physical interpretation in a given state $|\Psi\rangle$.

The situation in our case is certainly a little unusual, in that the local quantum field in the bulk $\phi_{\text{CFT}}(x)$ itself seems to depend on the state $|\Psi\rangle$. We should point out, as we pointed out in Sec. II, that if we consider the bulk-boundary mapping outside the horizon then this is also naively state dependent, since the "transfer function" is different in the vacuum, and the thermal state. At the least, it is clear that the $\frac{1}{\mathcal{N}}$ expansion of the mapping depends on the state. The authors of [9] have suggested [35] that it may be possible to write down a state-independent operator that has the correct behavior in a given state, presumably along the lines of the gauge-invariant relational observables of [36]. However, it would be nice to see a precise formulation of this statement, including an analysis that we can make such a construction stable with respect to quantum corrections.

We leave a deeper analysis of state dependence to further work. However, in this paper, we have tried to analyze this state dependence as carefully as possible, and we have found that it does not contradict any expectations from quantum mechanics.

We should mention that state-dependent operators have also appeared, in parallel with our work, in [4]. The reader should consult those papers for an alternate perspective.

Another important direction for future work has to do with the "uniqueness" of our construction. This issue also exists outside the horizon, where it is possible to write down a mapping between boundary and bulk operators. It has been suggested [37] that bulk locality uniquely fixes this mapping, but it would be nice to put this on a firmer footing. We explore this question, to some extent, in Appendix A, but this issue of uniqueness is even more acute for operators behind the horizon and deserves further investigation.

We would also like to address some philosophical issues regarding the relation of our approach to previous perspectives on this problem. Our perspective in this paper has been that low-point correlators in the conformal field theory can be reinterpreted in terms of correlators on a semiclassical spacetime. If we make the number of insertions too large, scaling with the central charge of the theory, then the picture of semiclassical spacetime breaks down. This is the fundamental limitation that must be respected according to us.

This delineates our perspective from some previous approaches to this problem. Earlier perspectives on black hole complementarity posited a picture of "observer complementarity," where the infalling observer and the observer outside the horizon saw different "realities," except that they could never communicate to obtain a contradiction. Some recent modifications of this approach have attempted to suggest that each different light cone might admit its own "reality." In our opinion, these perspectives are not entirely tenable, and we have not used them at all in this paper.

Our perspective is that there is a *global picture* of reality, which can be accessed by a super-observer in the CFT. As we have shown, this picture is consistent with correlation functions, where the number of insertions does not scale with $\mathcal{N}$. If we exceed this bound it is approximate locality that breaks down. This perspective on black hole complementarity—where we do not attempt to find semiclassical bulk interpretations for $\mathcal{N}$-point correlators—resolves much of the confusion surrounding the information paradox in AdS/CFT.

ACKNOWLEDGMENTS

We would like to thank Ahmed Almheiri, Ofer Aharony, Steven Avery, Souvik Banerjee, Jose Barbon, Jan de Boer, Ramy Brustein, Jan-Willem Bryan, Borun Chowdhury, Bartlomiej Czech, Frederik Denef, James Drummond, Roberto Emparan, Ben Freivogel, Cesar Gomez, Rajesh Gopakumar, Daniel Harlow, Tom Hartman, Gary Horowitz, Norihiro Iizuka, Irfan Ilgin, Dileep Jatkar, Jared Kaplan, Andreas Karch, Elias Kiritsis, Hong Liu, R. Loganayagam, Javier Magan, Gautam Mandal, Samir Mathur, Shiraz Minwalla, Juan Maldacena, Don Marolf, Tim Morris, Rob Myers, Boris Pioline, Andrea Puhm, Joe Polchinski, Eliezer Rabinovici, Slava Rychkov, Shubho Roy, Aninda Sinha, Ashoke Sen, Kostas Skenderis, Marika Taylor, Sandip Trivedi, David Turton, Erik Verlinde, Herman Verlinde, Mark van Raamsdonk, Spenta Wadia, Amos Yarom, Xi Yin, and especially Luboš Motl for very helpful discussions. We are grateful to Rajesh Gopakumar, Shiraz Minwalla, Luboš Motl, and Spenta Wadia for comments on a draft of this manuscript. K. P. and S. R. are grateful to the Seventh regional meeting in String Theory (Crete) for hospitality. S. R. is grateful to the participants of the "Cosmological Frontiers in Fundamental Physics" conference at Perimeter Institute, the "Fuzz, Fire, or Complementarity" workshop at Kavli Institute for Theoretical Physics (KITP) (Santa Barbara), and ICTS discussion meeting on the "Information Paradox, Entanglement and Black Holes" for useful discussions. S. R. is also grateful to the Perimeter Institute, HRI (Allahabad), Princeton University, KITP (Santa Barbara), UCLV (Santa Clara), and ICIMAF (Havana) for their hospitality, while this work was being completed. S. R. is partly supported by a Ramanujan fellowship of the Department of Science and Technology. K. P. is grateful to the organizers and participants of the XVIII IFT Xmas Workshop (Madrid), the "Black hole horizons and quantum information" TH-institute at CERN, the January 2013 Discussion Meeting on String theory in TIFR (Mumbai), the 13th Itzykson Conference (Saclay), the GR20-Amaldi10 conference (Warsaw), and the Spanish-Portuguese Relativity meeting 2013 (Benasque). K. P. would like to thank the Perimeter Institute, DAMTP (Cambridge), the University of Padova, and the University of Southampton for hospitality during the completion of this work. K. P. acknowledges support from the Royal Netherlands Academy of Sciences (KNAW). We thank Rajesh Gopakumar for a discussion on this issue. We are grateful to Ashoke Sen for emphasizing this point to us, in several discussions. We are grateful to Shiraz Minwalla for emphasizing the utility of this plasma ball perspective to us.

APPENDIX A: $\frac{1}{\mathcal{N}}$ CORRECTIONS, ALTERNATE PURIFICATIONS, AND UNIQUENESS

There are two parts to our construction of the mirror operators behind the horizon. One of them is (3.18), which tells us that the mirror operators commute with the ordinary operators. The second is (3.17), which tells us their correlations with ordinary operators. As we have mentioned many times above, we expect (3.18) to hold unchanged when $\frac{1}{\mathcal{N}}$ corrections are included, but (3.17) should receive corrections at first nontrivial order in $\frac{1}{\mathcal{N}}$.

To compute these corrections is a formidable task, even for simple Witten diagrams outside the black hole particularly in a state with energy that scales with $O(\mathcal{N})$. Unlike vacuum Witten diagrams, where $\frac{1}{\mathcal{N}}$ corrections correspond to bulk loops, here, we also have to be careful about the ensemble (canonical vs microcanonical) in which we are working. Nevertheless, in the first part of this section, we discuss how we would modify our prescription if someone were to *give us* the right answer. This brings up the issue of the uniqueness of our construction that we discuss next.

1. Accounting for $\frac{1}{\mathcal{N}}$ corrections

In principle one could compute $\frac{1}{\mathcal{N}}$ corrections to the Bogoliubov coefficients that translate between quantization on the Schwarzschild slices and the nice slices. To our knowledge, no such computation has actually been performed in anti-de Sitter space. However, let us say that such a computation of Bogoliubov coefficients in the *bulk* tells us that we should use not the thermofield doubled state, but the state

$$|\Psi\rangle_{\rm doub} = \sum_{i,j} C_{ij} |E_i\rangle_{\rm out} |\tilde{E}_j\rangle_{\rm in}, \tag{A1}$$

to do bulk computations. This indicates that the state $|E_i\rangle_{\rm out}$ in the Hilbert space of the field theory of the outside Schwarzschild observer is entangled with the state $|E_j\rangle_{\rm in}$ in the field theory of the inside observer.

Here C_{ij} is a matrix that tells us the entanglement between the two sides. Obviously, if we take $C_{ij} = \frac{1}{\sqrt{Z}} e^{-\frac{\beta E_i}{2}} \delta_{ij}$, we get the thermofield doubled state. We are interested in matrices C_{ij} that are close to this form, and differ from it by $\frac{1}{\mathcal{N}}$ corrections. However, below, we will not assume anything about C_{ij} except that it is invertible.

Note, however, that given a pure state in the theory $|\Psi\rangle$, and a set of observables $\mathcal{A}$, we cannot mimic the state $|\Psi\rangle_{\rm doub}$ for any arbitrary matrix C_{ij}. We have the very important consistency condition, that for correlators of ordinary operators in $\mathcal{A}$

$${}_{\rm doub}\langle\Psi|A_p|\Psi\rangle_{\rm doub} = \langle\Psi|A_p|\Psi\rangle, \quad \forall A_p \in \mathcal{A}, \tag{A2}$$

where these elements of $\mathcal{A}$ can, of course, be products of smaller elements. We see why the thermofield doubled state is generically a good choice to leading order in $\frac{1}{\mathcal{N}}$. In this state

$${}_{\rm tfd}\langle\Psi|A_p|\Psi\rangle_{\rm tfd} = {\rm Tr}(e^{-\beta H} A_p) = \langle\Psi|A_p|\Psi\rangle + {\rm O}\left(\frac{1}{\mathcal{N}}\right),$$

for almost any equilibrium state $|\Psi\rangle$. However the ${\rm O}(\frac{1}{\mathcal{N}})$ corrections above also tell us that in general, at subleading orders in $\frac{1}{N}$, this consistency condition requires us to use a more general state of the form (A1). We show how to now correct (3.17) for a state of the form (A1).

Just as in (3.7) and (3.5), as usual, for each operator, $A_p|E_i\rangle = (A_p)_{ji}|E_j\rangle$, we have the mirror operator, which acts on the other side: $A_p^{\rm doub}|\tilde{E}_i\rangle = (A_p)^*_{ji}|\tilde{E}_j\rangle$. This is the operator, that *in a physical sense* acts in the same way on the other side, because the $|E_i\rangle$ form a privileged energy eigenbasis. For example, we could go to the Schmidt basis, in which the entanglement is diagonal and then ask for the operators that act on the other side of the Schmidt basis in the same way, as we did in Eq. (6.9). The reader should note that we are asking a slightly different question here.

Now, note that, in the state $|\Psi\rangle_{\rm doub}$, we can convert the action of $A_p^{\rm doub}$, which acts only on the tilde states, to an action of operators that act only on the ordinary states

$$\begin{aligned} A_p^{\rm doub}|\Psi\rangle_{\rm doub} &= C_{ij}(A_p)^*_{kj}|E_i\rangle|\tilde{E}_k\rangle \\ &= C_{ij}(A_p)^*_{kj}(C^{-1})_{kl}C_{lm}|E_i\rangle|\tilde{E}_m\rangle \\ &= (C^{-1}A_p^\dagger C)_{il}C_{lm}|E_i\rangle|\tilde{E}_m\rangle = \breve{A}_p|\Psi\rangle_{\rm doub}, \end{aligned}$$

where all repeated indices are summed

$$\breve{A}_p = C^{-1}A_p^\dagger C,$$

and $C^{-1}_{il}C_{lm} = \delta_{im}$.

Now, to mimic the action of the mirror operators in a state $|\Psi\rangle_{\rm doub}$, we expand the set of observables $\mathcal{A}$ to include the observables $\breve{A}_p$ and then we simply define our tilde operators to satisfy

$$\tilde{A}_p|\Psi\rangle = \breve{A}_p|\Psi\rangle. \tag{A3}$$

Once we can define the tildes to have an action on the state of some product of ordinary operators, simply by commuting them to the right

$$\tilde{A}_{p_1}A_{p_2}\ldots A_{p_m}|\Psi\rangle = A_{p_2}\ldots A_{p_m}\tilde{A}_{p_1}|\Psi\rangle,$$

and use this, by induction, to define the action of a product of tildes as well.

We can again check that this definition works correctly to reproduce correlators of products of simple operators. We see, from an application of the rules above, that

$$\tilde{A}_{p_1}\tilde{A}_{p_2}|\Psi\rangle = \breve{A}_{p_2}\breve{A}_{p_1}|\Psi\rangle.$$

On the other hand, we can also check that

$$\begin{aligned} A_{p_1}^{\rm doub}A_{p_2}^{\rm doub}|\Psi\rangle_{\rm doub} &= (A_{p_1})^*_{lk}(A_{p_2})^*_{kj}C_{ji}|E_i\rangle|\tilde{E}_l\rangle \\ &= C_{ij}(A_{p_1})^*_{lk}(A_{p_2})^*_{kj}C^{-1}_{lt}C_{tm}|E_i\rangle|\tilde{E}_m\rangle \\ &= CA_{p_2}^\dagger A_{p_1}^\dagger C^{-1}|\Psi\rangle_{\rm doub}. \end{aligned}$$

So

$$\begin{aligned} \langle\Psi|\tilde{A}_{p_1}\tilde{A}_{p_2}|\Psi\rangle &= \langle\Psi|\breve{A}_{p_2}\breve{A}_{p_1}|\Psi\rangle \\ &= {}_{\rm doub}\langle\Psi|A_{p_1}^{\rm doub}A_{p_2}^{\rm doub}|\Psi\rangle_{\rm doub}. \end{aligned}$$

By an extension of this to higher products we can check that, just as desired,

$$\begin{aligned} &\langle\Psi|\tilde{A}_{p_1}\ldots\tilde{A}_{p_m}A_{p_{m+1}}\ldots A_{p_n}|\Psi\rangle \\ &\quad = {}_{\rm doub}\langle\Psi|A_{p_1}^{\rm doub}\ldots A_{p_m}^{\rm doub}A_{p_{m+1}}\ldots A_{p_n}|\Psi\rangle_{\rm doub}. \end{aligned}$$

2. Uniqueness

This discussion brings up another point. If we are given a state, and bulk correlators, how do we fix the matrix C. For an equilibrium state, it is reasonable to choose C to be diagonal in the energy eigenbasis.

Geometrically, this is the following statement. Consider a black hole that has reached thermal equilibrium. If the black hole was formed from the collapse of a state with a narrow band of energies, it *may not* be well represented by the thermofield doubled state. However, it should still be well represented by the state

$$\psi_{\mathrm{doub}} = \sum_i C_{ii} |E_i\rangle|\tilde{E}_i\rangle.$$

For example, if the original black hole is well represented by the microcanonical ensemble, then we could take C_{ii} above to be constant for a given range of energies, and zero outside. Geometrically this also corresponds to an "eternal black hole," but where the entanglement corresponds to the microcanonical ensemble. This geometry differs at $\mathrm{O}(\frac{1}{\mathcal{N}})$ from the canonical eternal black hole geometry.

Both these geometries share the property that they are invariant if we evolve forward in time on the right, and backward in time on the left. In the bulk, this is an isometry which rotates a spacelike slice passing through the bifurcation point.

If we do make the assumption that C is diagonal in the energy eigenbasis, then our tilde operators are essentially fixed. This is because the eigenvalues C_{ii} can be set by measuring expectation values of ordinary operator A_p in the state $|\Psi\rangle$ and demanding (A2).

However, a note of caution is in order here. Even if $|\Psi\rangle$ is in equilibrium, as defined in V, and (A2) holds, it is *not* necessary for C to be diagonal. This is because we see

$${}_{\mathrm{doub}}\langle\Psi|e^{iHt}A_p e^{-iHt}|\Psi\rangle_{\mathrm{doub}} = C_{ij}C^*_{jk}A_{ik}e^{i(E_i-E_k)t}.$$

Now, it is easy to see that even for a generic matrix C, that satisfies $\mathrm{Tr}(C^\dagger C) = 1$, the time dependence above is extremely small.

Note that this question could also be raised about the correspondence between the eternal black hole and the thermofield doubled state. What sets the precise form of the entanglement there to be diagonal in the energy eigenbasis? One answer would be that the bulk theory has the isometry above, where we can rotate a spacelike slice about the bifurcation point. However, this isometry exists to excellent precision even if we change the structure of the entanglement. This issue is related to the issue of the uniqueness of our construction. We leave a more detailed study to further work.

APPENDIX B: CHOICE OF GAUGE

We now briefly discuss our choice of gauge in (3.19). The construction of local operators corresponding to charged fields was also discussed in recent papers [16,37,38].

First, we briefly remind the reader of the nonlocal commutators that result from working in a fixed gauge. We take the example of scalar QED in curved space, although non-Abelian gauge theories lead to similar results, and we believe that our qualitative conclusions should also hold for gravity.

Let us put the metric in the standard ADM $\mathrm{d}+1$ form:

$$ds^2 = -N^2dt^2 + h_{ij}(dx_i + N_i dt)(dx_j + N_j dt).$$

With this split, we have

$$\sqrt{-g} = \sqrt{h}N,$$

and the components of the inverse metric are

$$g^{00} = -1/N^2; \qquad g^{ij} = h^{ij} - N^iN^j/N^2; \qquad g^{0i} = \frac{N^i}{N^2}. \tag{B1}$$

The Lagrangian density for scalar QED is given by

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J_\mu A^\mu + \mathcal{L}_{\mathrm{matter}},$$

where J^μ is composed of the matter fields, but we are not interested in the matter Lagrangian here, except for the Poisson brackets it will induce with the matter field.

We see that we can write

$$\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}F_{0i}F_{0j}(g^{00}g^{ij} - g^{0j}g^{0i}) + \frac{1}{2}F_{0i}F_{kl}(g^{0k}g^{il} - g^{0l}g^{ik}) + \frac{1}{4}F_{mn}F_{kl}g^{mk}g^{nl}, \tag{B2}$$

where all Latin indices run only over the spatial direction. This can be simplified by using the form of the inverse metric given above

$$\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = -\frac{h^{ij}}{2N^2}F_{0j}F_{0i} + \frac{N^k}{N^2}h^{li}F_{kl}F_{0i} + \frac{1}{4}F_{kl}F_{mn}g^{kl}g^{ln}.$$

Now, we go over to the Hamiltonian formalism to make contact with quantum mechanics. We use i for the spatial directions only. We find that

$$\Pi^i(x) = \frac{\partial L}{\partial(\partial_0 A_i(x))} = -F^{0i}(x).$$

Note that we have

$$F^{0i} = g^{0\mu}g^{\rho i}F_{\mu\rho} = (g^{00}g^{ji} - g^{0j}g^{0i})F_{0j} + g^{0k}g^{li}F_{kl}$$
$$= -\frac{h^{ij}}{N^2}F_{0j} + \frac{N^k}{N^2}h^{li}F_{kl},$$

which is entirely consistent with the expansion of the Lagrangian above in (B2).

Just from the structure of the Lagrangian, we have the "primary" constraint

$$\phi_1 = \Pi^0 = 0. \tag{B3}$$

Following Dirac [39], we will use the notation ϕ_n to denote the various constraints that will arise.

We proceed to work out the Hamiltonian. As usual, the sign of the term quadratic in F_{0i} is reversed and the the term linear in F_{0i} drops out. We see that

$$\begin{aligned}\Pi^i\partial_0 A_i + \frac{1}{4}F_{\mu\nu}F^{\mu\nu} &= \Pi^i(F_{0i} + \partial_i A_0) + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\\ &= \frac{1}{2}\frac{h^{ij}}{N^2}F_{0j}F_{0i} + \frac{1}{4}F_{kl}F_{mn}g^{kl}g^{ln}\\ &\quad + \Pi^i\partial_i A_0\\ &= \frac{1}{2}h_{ij}\left(\Pi^i + \frac{N^k}{N^2}h^{li}F_{kl}\right)\\ &\quad \times\left(\Pi^j + \frac{N^m}{N^2}h^{nj}F_{mn}\right)\\ &\quad + \frac{1}{4}F_{kl}F_{mn}g^{km}g^{ln} + \Pi^i\partial_i A_0.\end{aligned}$$

Using the form of the inverse metric given in (B1), we find

$$\begin{aligned}g^{km}g^{ln}F_{mn}F_{kl} &= \bigg(h^{km}h^{ln} - \frac{1}{N^2}h^{km}N^lN^n - \frac{1}{N^2}h^{ln}N^kN^m\\ &\quad + N^lN^kN^mN^n\bigg)F_{mn}F_{kl}\\ &= F_{mn}F^{mn} - \frac{2}{N^2}F_{mn}h^{km}N^lN^nF_{kl},\end{aligned}$$

where as usual, the spatial indices have been raised using h. We see that the second term above precisely cancels with the term that appears when the whole square involving the momentum in the Hamiltonian is expanded out. So, we find that finally

$$\Pi^i\partial_0 A_i + \frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}\Pi_i\Pi^i + \Pi^iN^kF_{kl} + \frac{1}{4}F_{mn}F^{mn} + \Pi^i\partial_i A_0.$$

This leads to the Hamiltonian density

$$\begin{aligned}\mathcal{H}_0 &= [\Pi^i\partial_0 A_i - \mathcal{L} + U_1\Pi^0]d^3x\\ &= \frac{1}{2}\Pi_i\Pi^i + \Pi^iN^kF_{kl} + \frac{1}{4}F_{mn}F^{mn} + \Pi^i\partial_i A_0 + J^\mu A_\mu.\end{aligned}$$

Of course, the Hamiltonian is given by $H_0 = \int\sqrt{-g}\mathcal{H}_0$.

We have called this Hamiltonian density $\mathcal{H}_0$, since we will have to modify it systematically to get consistency with the constraints as laid down in Dirac's procedure. To start with, we also need to include a term $U_1\Pi^0$ for the constraint, as specified by Dirac. Here U_1 can be an arbitrary function of the A_i and the conjugate momenta Π^i. After adding this term, we have the modified Hamiltonian $\mathcal{H}_1 = \mathcal{H}_0 + U_1\phi_1$.

Now, to preserve the constraint we require

$$\{\Pi^0, H_1\} = 0.$$

Recall that when we compute the Poisson brackets, by definition, we have

$$\{A_0(x), \Pi^0(x')\} = \frac{1}{\sqrt{-g}}\delta^d(x - x');$$
$$\{A_i(x), \Pi^j(x')\} = \frac{1}{\sqrt{-g}}\delta_i^j\delta^d(x - x'),$$

where we have suppressed the time coordinate, which is always equal in the quantities in Poisson bracket. The additional factor of $\sqrt{-g}$ appears because of the way we defined our Lagrangian, without the $\sqrt{-g}$.

So, we see that the Poisson bracket above immediately leads to the Gauss law

$$\phi_2 = \frac{1}{\sqrt{-g}}\partial_i[\sqrt{-g}\Pi^i] + J^0 = 0. \tag{B4}$$

So, we have obtained ϕ_2 as a secondary constraint. We see that (B4) does not lead to any further constraints because

$$\begin{aligned}&\{\partial_i\sqrt{-g}(x)\Pi^i(x), F_{lm}(y)\}\\ &= \partial_{x^i}\sqrt{-g}(x)\bigg[\delta^i_m\partial_{y^l}\frac{1}{\sqrt{-g}(y)}\delta^d(x - y)\\ &\quad - \delta^i_l\partial_{y^m}\frac{1}{\sqrt{-g}}\delta^d(x - y)\bigg] = 0,\end{aligned} \tag{B5}$$

since we can convert the $(-g)^{-\frac{1}{2}}(y)$ to a $(-g)^{-\frac{1}{2}}(x)$, using the delta function, pull it out of the derivative, and then cancel it with the $(-g)^{\frac{1}{2}}(x)$ that accompanies the momentum.

As a result, we see that we have

$$\{\phi_2, H_1\} = (\partial_i\sqrt{-g}J^i + \{\sqrt{-g}J_0, H_1\}) = 0,$$

where we have not displayed terms that vanish because of (B5). We are implicitly assuming that when we write down the matter Lagrangian, it gives rise to

$$\{\sqrt{-g}J^0, H_1\} + \partial_i\sqrt{-g}J^i = 0,$$

as an identity.

We now write a second Hamiltonian as

$$H_2 = \int \sqrt{-g}(\mathcal{H}_1 + U_2\phi_2),$$

where U_2, for now, is another arbitrary parameter.

However, we see that we cannot fix the Hamiltonian uniquely, and that U and U_2 are left undetermined. This is because (B3) and (B4) have zero Poisson bracket with each other, so they are *first class* constraints.

At this point, in principle, we could restrict ourselves to only gauge invariant operators. In this language, the analogue of local fields would be fields with Wilson lines attached to them. Here, we will take a slightly cruder approach of simply fixing the gauge, since that is more convenient from the point of view of constructing local bulk observables.

1. Gauge fixing

To convert these first class constraints into second class constraints, we will consider a set of "algebraic gauges" which are fixed by imposing

$$\phi_3 = A_a = 0; \qquad \phi_4 = \Pi_a + \frac{N^k}{N^2}F_{ka} - \partial_a A_0 = 0. \tag{B6}$$

The second constraint is meant to impose $F_{a0} + \partial_a A_0 = 0$. For example, in flat space, we could take $a = 3$ to get the axial gauge. The reader should keep in mind that a is not a dummy index in this section but is fixed to be the index of some particular *spatial* coordinate.

Note that we now have the following matrix of Poisson brackets between the constraints

$$C_{mn}(x,y) = \{\phi_m(x), \phi_n(y)\} = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{\sqrt{-g(x)}}\frac{\partial}{\partial y^a}\delta^d(x-y) \\ 0 & 0 & -\frac{1}{\sqrt{-g(y)}}\frac{\partial}{\partial x^a}\delta^d(x-y) & 0 \\ 0 & \frac{1}{\sqrt{-g(x)}}\frac{-\partial}{\partial y^a}\delta^d(x-y) & 0 & \frac{h_{aa}}{\sqrt{-g}}\delta^d(x-y) \\ \frac{-1}{\sqrt{-g(y)}}\frac{\partial}{\partial x^a}\delta^d(x-y) & 0 & \frac{-h_{aa}}{\sqrt{-g}}\delta^d(x-y) & 0 \end{pmatrix}.$$

We write down a third Hamiltonian:

$$H_3 = H_2 + \int \sqrt{-g}(U_3\phi_3 + U_4\phi_4)d^dx. \tag{B7}$$

For consistency, we need to ensure that (B3), (B4), (B6) are all consistent with the Hamiltonian (B7),

$$\{\phi_m, H_3\} = 0, \quad \text{for } m = 1,2,3,4.$$

where all the equations have to hold in a weak sense.

The main thing to calculate is

$$\{\phi_4, H_3\}.$$

We note that

$$\begin{aligned}\{\Pi_a(x), F_{kl}(y)\} &= h_{ab}(x)\{\Pi^b(x), F_{kl}(y)\} \\ &= h_{ab}(x)\Big[\delta^b_l \frac{1}{\sqrt{-g(x)}}\frac{\partial}{\partial y^k}\delta^d(x-y) \\ &\quad + \delta^b_k \frac{1}{\sqrt{-g(x)}}\frac{\partial}{\partial y^l}\delta^d(x-y)\Big] \\ &= \frac{h_{al}(x)}{\sqrt{-g(x)}}\partial_{y^k}\delta^d(x-y) \\ &\quad + \frac{h_{ak}(x)}{\sqrt{-g(x)}}\partial_{y^l}\delta^d(x-y).\end{aligned}$$

Consequently,

$$\begin{aligned}\left\{\Pi_a(x), \int \sqrt{-g}(y)F_{kl}(y)F^{kl}(y)\right\} &= -4\frac{h_{al}(x)}{\sqrt{-g(x)}} \\ &\quad \times \partial_k\sqrt{-g}F^{kl}(x), \\ \left\{\Pi_a(x)\int \sqrt{-g}N^k\Pi^l F_{kl}dy\right\} &= -\frac{h_{al}(x)}{\sqrt{-g(x)}}\partial_k\sqrt{-g} \\ &\quad \times (N^k\Pi^l - \Pi^l N^k).\end{aligned}$$

Putting all this together, we see that

$$\{\phi_4, H_3\} = -\frac{h_{al}(x)}{\sqrt{-g}(x)}\partial_k\sqrt{-g}F^{kl}(x) - \frac{h_{al}(x)}{\sqrt{-g}(x)}\partial_k\frac{\sqrt{-g}}{N^2}(N^k\Pi^l - \Pi^l N^k) - J_a + \frac{N^k}{N^2}\left(\partial_k\left(\Pi_a + \frac{N^p}{N^2}F_{pa} + \partial_a A_0\right) - \partial_a\left(\Pi_k + \frac{N^p}{N^2}F_{pk} + \partial_k A_0\right)\right) - \partial_a U_1 - U_3.$$

We also see see that

$$\{\phi_3, H_3\} = \phi_4 - \partial_a U_2 - h_{aa}U_4, \qquad \{\phi_1, H_3\} = \partial_3 U_4, \qquad \{\phi_2, H_3\} = \partial_3 U_3.$$

We can solve the equations above as follows:

$$U_2 = 0, \qquad U_3 = 0, \qquad U_4 = 0,$$
$$U_1 = \int_{z_0}^{z}\{\phi_4(\zeta), H_3\}d\zeta.$$

In the last line, we have explicitly displayed the dependence of the quantities on the spacetime coordinates. These solutions are *not unique*. For example, in the first line above, we set $U_2 = 0$, although, technically we could still set U_2 to be a function of only the x, y coordinates; this is a symptom of the residual gauge invariance after our gauge fixing.

So, finally we end up with a nonlocal Hamiltonian, which is just H_3 with the solutions for $U_1 \ldots U_4$ substituted,

$$H_4 = \int N\sqrt{h}\left[\frac{1}{2}\Pi_i\Pi^i + \Pi^l\frac{N^k}{N^2}F_{kl} + \frac{1}{4}F_{mn}F^{mn} + \Pi^i\partial_i A_0 + J^\mu A_\mu + \Pi^0\int_{z_0}^{z}\{\phi_4(\zeta), H_0\}d\zeta\right]d^d x.$$

a. Quantization

Finally, we turn to the nonlocal commutators we get by quantizing this theory. To do this we have to use Dirac's prescription. First, we need to find the inverse of the matrix C_{mn}. Note that this is defined by

$$\int D_{mn}(x-y)C_{np}(y-z)dy = \delta_{mp}\delta^d(x-z).$$

We need solutions to the following differential equations:

$$\frac{\partial}{\partial x^a}\left(\frac{1}{\sqrt{-g}}\mathcal{A}(x,y)\right) = \delta^d(x-y),$$
$$\frac{\partial}{\partial x^a}\frac{1}{\sqrt{-g}}\mathcal{B}(x,y) = \frac{h_{aa}}{\sqrt{-g}}\mathcal{A}(x,y).$$

We see that

$$D_{mn}(x-y) = \delta^d(x_1-y_1)\delta^d(x_2-y_2) \times \begin{pmatrix} 0 & \mathcal{B}(x,y) & 0 & \mathcal{A}(x,y) \\ -\mathcal{B}(x,y) & 0 & \mathcal{A}(x,y) & 0 \\ 0 & -\mathcal{A}(x,y) & 0 & 0 \\ \mathcal{A}(x,y) & 0 & 0 & 0 \end{pmatrix}.$$

A solution to these differential equations is given by

$$\mathcal{A}(x,y) = \sqrt{-g}(\theta(x^a - y^a) + A(\hat{x},\hat{y}))$$
$$\mathcal{B}(x,y) = \sqrt{-g}\left(\int dx^a h_{aa}\frac{1}{\sqrt{-g}}\mathcal{A}(x,y) + B(\hat{x},\hat{y})\right).$$

where the dependence on $\hat{x}$ and $\hat{y}$ means that A and B do *not* depend on x^a. So, we see that we have two arbitrary functions A and B. This is because our gauge-fixing condition does not completely fix the gauge. We will ignore these functions for now.

Now, the Dirac prescription is to consider Dirac brackets given by

$$[\mathcal{F}, \mathcal{G}]_{\mathrm{D.B.}} = \{\mathcal{F}, \mathcal{G}\} - \{\mathcal{F}, \phi_m\}D^{mn}\{\phi_n, \mathcal{G}\}.$$

We are finally in a position to compute commutators between the electric field and the scalar field. When we write down the matter Lagrangian, we get a current that should satisfy

$$\{J^0(x), \Phi(y)\} = \frac{1}{\sqrt{-g}}\delta^d(x-y)\Phi(x).$$

So, the interesting commutator that we want to investigate is

$$[\Pi^a(x), \phi(y)]_{\mathrm{D.B.}} = -\int dz_1 dz_2\{\Pi^a(x), \phi_3(z_1)\} \times D^{32}(z_1, z_2)\{\phi_2(z_2), \Phi(y)\} = \theta(x^a - y^a) + A(\hat{x},\hat{y}).$$

This is a simple and universal result. However, notice that the choice of the function $A(\hat{x},\hat{y})$ gives us some freedom in choosing the exact commutator, as we point out below.

Now, we apply all this to the case of the AdS black brane. In the region outside the brane, we choose the gauge $A_z = 0$. Recall that this is the gauge that must be chosen close to the boundary in any case to get the usual relationship between bulk and boundary correlators.

We choose the function $A = -1$, and this leads to the commutators

$$[\Pi^z(t, x_1), \phi(t, x_2)]_{\text{D.B.}} = -\theta(z_1 - z_2)\delta^{d-1}(x_1 - x_2).$$

The physical interpretation of this commutator in terms of Wilson lines is simple. We think of the field ϕ as being attached to a Wilson line that goes all the way to the boundary at $z = 0$, along a path of constant spatial coordinates. So, if the electric field operator is placed at a *smaller* value of z (closer to the boundary), it intersects the Wilson line, leading to the nonzero commutator above.

Now consider the region behind the horizon. In this region, z becomes a timelike coordinate and t becomes a spatial coordinate. We now choose the gauge $A_t = 0$. Now, we find the commutator

$$[\Pi^t(z, x_1), \phi(z, x_2)]_{\text{D.B.}} = -\theta(t_1 - t_2)\delta^{d-1}(x_1 - x_2),$$

where the last δ^{d-1} excludes the δ function in z, of course, which is now playing the role of a time coordinate.

This formula suggests an amusing interpretation in terms of Wilson lines. The operators behind the horizon have Wilson lines that extend deeper into the black hole, and eventually reemerge near the boundary through a wormhole. So, their charge can be measured at infinity, but not their position.

2. Mirror operators below the Hawking-Page temperature?

We now discuss another issue that has sometimes been raised. We expect the tilde operators to exist in any thermal state, including one where the temperature is low enough that the dual state is represented by a gas of gravitons, rather than a black hole. What is the significance of mirror operators below the Hawking-Page temperature?

In fact, the issue of gauge invariance helps us here. The mirror operators cannot be used in any region that is connected to the boundary. First, note that the form of the equal-time commutator between a conserved current and a charged local operator is fixed by locality. We must have

$$[j^0(t, \Omega), \mathcal{O}(t, \Omega')] = q\mathcal{O}(t, \Omega)\delta^{d-1}(\Omega - \Omega'),$$

where the delta function is understood to be correctly normalized on the sphere. Clearly, the commutation relations that we have imposed in (3.19) are not of this form. So, the $\tilde{\mathcal{O}}$ operators *cannot* be understood to be local operators on the boundary.

Now, this also implies that they cannot appear in fields, in a region that is not causally separated from the boundary. We could, otherwise, take a limit as these operators tend to the boundary, and the commutator would have the wrong form for a local field. This fact, by itself, implies that the mirror operators $\tilde{\mathcal{O}}$ do not appear in expressions for local bulk fields below the Hawking-Page temperature.

To conclude, the existence of the $\tilde{\mathcal{O}}$ operators is necessary to construct fields behind the boundary. However, just because they can be defined in a state in the CFT does not mean that they appear in the expressions for bulk fields.

3. Constructing the other side?

For the same reason, it is clear that the mirror operators do not really represent a region III of the black hole. Note that, in the eternal black hole geometry if $\hat{Q}$ is the charge that we are measuring near the boundary of the first CFT, and $\tilde{\mathcal{O}}^i(t, \Omega)$ are operators in the second CFT, then we would have $[\hat{Q}, \tilde{\mathcal{O}}^i(t, \Omega)] = 0$. The commutation relations (3.19), which allow us to measure the charge of the $\tilde{\mathcal{O}}$ operators from the boundary, tell us that boundary of the CFT always covers the region in which the $\tilde{\mathcal{O}}$ operators live. This seems to suggest that we cannot really reconstruct region III from our construction.

APPENDIX C: THE “MEASUREMENT” ARGUMENT

The first AMPS paper contained the statement “We can therefore construct operators acting on the early radiation, whose action [2] is equal to that of a projection operator onto any given subspace of the late radiation.”

This statement by itself does not lead to any paradox. The apparent paradox appears when we also consider the radiation behind the horizon and this leads to a seeming violation of the strong subadditivity of entropy. We have already discussed this issue in Sec. IV A. There, we also discussed in detail how it was not possible, while remaining within a framework that is described by semiclassical spacetime, for the observer to distill the part that is entangled with the late radiation.

However, even if the authors of [2] did not intend this, the statement above has led to a misunderstanding that the existence of these operators acting on the early radiation, that can project the late radiation onto a given state, *ipso facto*, implies that the horizon may have a firewall.

In this appendix, we clarify some of these basic issues.

“Acting” with operators.—The basic fact to realize is that even operators that are localized to a region “acting” on a state can produce unusual effects far away from the region. In fact, the Reeh-Schlieder theorem tells us the following. Consider a local quantum field theory, and the set of local operators acting on some open set M, which we denote by $\Phi(M)$, and *any state* of finite energy $|\Omega\rangle$, which may even be the vacuum. Then, the set $\Phi(M)|\Omega\rangle$ is dense in the full Hilbert space $\mathcal{H}$.

For details and references about this theorem see [31]. This theorem has to do with the fact that even the vacuum state has long-range entanglement. So, by delicately

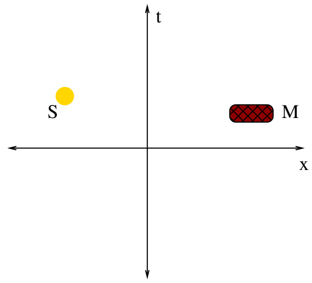

FIG. 11 (color online). Implication of the Reeh-Schlieder theorem. A spacetime diagram shows how operators in the bounded region M can create the sun S at large spacelike separation.

manipulating even a localized region of this state, we can create whatever we want even in the causal complement of the region. The physical implication of this theorem in shown in Fig 11.

This theorem *does not* imply any violation of locality. In fact, from a physical perspective, we can modify the Hamiltonian and cause the state to undergo unitary evolution, but we cannot "act" with some arbitrary operators on a state.

Complicated measurements.—Now, it is also well known that very complicated "measurements" do fall into the class of actions that can be obtained through unitary evolution. We could turn on a Hamiltonian that would entangle a local quantum field theory with another much larger system and permit us to measure some quantity with great accuracy. The next simple point we want to make is that if we actually perform such a measurement, it can disturb the system and again lead to funny effects.

As we have already explained in Sec. IVA, if we try to measure the early radiation to distill the part that is entangled with the late radiation, there is no reason to expect that this operation will have a simple semiclassical interpretation because it will involve insertions of operators with energies that scale with $\mathcal{N}$.

However, even some measurements that may have spacetime interpretations can disturb the system enough to create a firewall. This is particularly true of measurements that are extremely sharp i.e. measurements where the associated projection operators project onto an extremely low-dimensional subspace of the Hilbert space.

Here, we point out how this phenomenon can be seen even in the flat space *Minkowski vacuum.* We will show how, by a very special measurement in the Minkowski vacuum, we can create a firewall at the Rindler horizon. Consider quantizing a massless scalar field $\Box\phi = 0$, in $d+1$-dimensional spacetime, with metric

$$ds^2 = -dt^2 + dz^2 + d\boldsymbol{x}^2,$$

where $\boldsymbol{x}$ is a $(d-1)$-dimensional vector. In region I (as shown in Fig. 12), we transform to the coordinates

$$t = \sigma \sinh\tau, \qquad z = \sigma\cosh\tau,$$

so that the metric becomes

$$ds^2 = -\sigma^2 d\tau^2 + d\sigma^2 + d\boldsymbol{x}^2.$$

We can quantize the field in region I (as shown in the figure) using the expansion

$$\phi(\tau,\sigma,\boldsymbol{x}) = \int_{\omega>0} \frac{d\omega d^{d-1}\boldsymbol{k}}{(2\pi)^d} \times \left[\frac{1}{\sqrt{2\omega}} a_{\omega,\boldsymbol{k}} e^{-i\omega\tau + i\boldsymbol{k}\boldsymbol{x}} \frac{2K_{i\omega}(|\boldsymbol{k}|\sigma)}{|\Gamma(i\omega)|} + \text{H.c.}\right].$$

For region III, we use the coordinate transformation $t = -\sigma\sinh\tau$, $z = -\sigma\cosh\tau$ and expand the field as

$$\phi(\tau,\sigma,\boldsymbol{x}) = \int_{\omega>0} \frac{d\omega d^{d-1}\boldsymbol{k}}{(2\pi)^d} \times \left[\frac{1}{\sqrt{2\omega}} \tilde{a}_{\omega,\boldsymbol{k}} e^{i\omega\tau - i\boldsymbol{k}\boldsymbol{x}} \frac{2K_{i\omega}(|\boldsymbol{k}|\sigma)}{|\Gamma(i\omega)|} + \text{H.c.}\right].$$

Now, it is well known, that in this expansion, even the Minkowski vacuum appears as an entangled state,

$$|\Omega\rangle_{\text{Mink}} = \sum_E e^{-\pi E} |E\rangle_I \otimes |E\rangle_{III},$$

where the sum over E runs over the entire Fock space and E is the energy of the state in this Fock space.

Now, consider an observer who lives in region I for a long time and makes an accurate measurement of the Rindler energy. At the end of this process, the observer is entangled with a superposition of states in the Fock space that have the specific energy corresponding to the result of his measurement. However, the stress-tensor in a state with a specific energy diverges at the Rindler horizon [40]. Hence, for the observer O shown in Fig. 12, this creates a firewall as he crosses the Rindler horizon. Obviously this

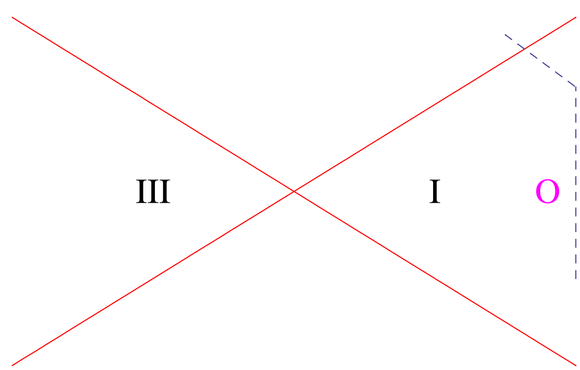

FIG. 12 (color online). An observer who measures the Rindler energy, and then tries to cross the Rindler horizon, encounters a "firewall."

does not mean that the Minkowski vacuum has such a firewall, but merely that a very sharp measurement,[15] which effectively involves entangling the system with an extremely large measurement "apparatus," can disturb the system enough to create unusual objects.

State dependence of the AMPS projection.—Finally, we would like to mention that the projection that AMPS consider [2] is state dependent. We are merely pointing out this fact, and do not attach any special significance to this issue, since we feel that state-dependent operators like the density matrix or the mirror operators that we have been considering are useful.

Note that AMPS would like to consider a measurement (in their notation) where the state of the black hole after the Page time is

$$|\Psi\rangle = \sum_i |\psi_i\rangle_E \otimes |i\rangle_L,$$

where E indexes the early radiation and L indexes the late radiation, since $\dim \mathcal{H}_E \gg \dim \mathcal{H}_L$. AMPS point out that for any state $|i\rangle_L$ of the late radiation and the corresponding projection operator $P^i = |i\rangle_L \langle i|_L$ we can define another projection operator $\hat{P}^i$ written in terms of the early radiation such that

$$\hat{P}^i |\Psi\rangle \approx P^i |\Psi\rangle = |\psi_i\rangle_E \otimes |i\rangle_L. \tag{C1}$$

We can write

$$\hat{P}^i = |\Psi_i\rangle_E \langle \Psi_i|_E \otimes I_L.$$

However, since the precise state $|\Psi_i\rangle$ that is correlated with $|i\rangle_L$ depends on the state $|\Psi\rangle$, this projector $\hat{P}^i$ must be correlated with the state $|\Psi\rangle$ to perform the action (C1).

APPENDIX D: TECHNICAL DETAILS OF THE TOMITA-TAKESAKI CONSTRUCTION

In this appendix we present some proofs of statements made in Sec. VI. For illustrative purposes we concentrate on the case of a finite-dimensional algebra acting on a finite-dimensional space $\mathcal{H}_\Psi$, because then the proofs are easy and we do not need to worry about issues of convergence. Of course, in the finite-dimensional case the quickest way to prove these statements is by working in an appropriate Schmidt basis, as was done in Sec. VI C. Here we provide an alternative presentation which may help the reader in following the more elaborate proofs for the infinite-dimensional case, which can be found in the mathematical literature [32].

Remember that we have the finite-dimensional algebra $\mathcal{A}$ acting on the Hilbert space $\mathcal{H}_\Psi$. We assume that if $A \in \mathcal{A}$ then $A^\dagger \in \mathcal{A}$. Clearly $\mathcal{A}$ is a von Neumann algebra. We also assume that states of the form $A|\Psi\rangle, A \in \mathcal{A}$ span the entire Hilbert space $\mathcal{H}_\Psi$, which means that the vector $|\Psi\rangle$ is *cyclic* for the algebra $\mathcal{A}$ and also that $A|\Psi\rangle = 0$ implies $A = 0$, which means $|\Psi\rangle$ is *separating*. Below we present the proofs of various technical statements that enter in the construction of the mirror operators in the language of the Tomita-Takesaki framework.

First we define the commutant $\mathcal{A}'$ (the set of operators acting on $\mathcal{H}_\Psi$ which commute with all elements of $\mathcal{A}$), which is also a von Neumann algebra. The von Neumann bicommutant theorem guarantees that

$$(\mathcal{A}')' = \mathcal{A}.$$

For a proof of this classic theorem we refer the reader to [32].

Then we show that if $|\Psi\rangle$ is cyclic and separating for $\mathcal{A}$ then it is also cyclic and separating for $\mathcal{A}'$.

Proof: (i) First we will prove that the vector $|\Psi\rangle$ is separating for the algebra $\mathcal{A}'$. Suppose we have an operator $A' \in \mathcal{A}'$ such that

$$A'|\Psi\rangle = 0.$$

We will show that implies that $A' = 0$ as an operator. Consider any other vector $|A\rangle$ in $\mathcal{H}_\Psi$. Since (by assumption) $|\Psi\rangle$ is cyclic for the algebra $\mathcal{A}$, it means that we can find an element $A \in \mathcal{A}$ such that $|A\rangle = A|\Psi\rangle$. We have

$$A'|A\rangle = A'A|\Psi\rangle = AA'|\Psi\rangle = 0,$$

where we used that $[A, A'] = 0$ and the assumption that $A'|\Psi\rangle = 0$. From this equation we find that A' actually annihilates every vector in $\mathcal{H}_\Psi$, hence we find the operator equation

$$A' = 0.$$

(ii) Then, we will prove that the vector $|\Psi\rangle$ is cyclic for the algebra $\mathcal{A}'$, which means that for every vector $|\Psi'\rangle \in \mathcal{H}_\Psi$, there is $A' \in \mathcal{A}'$ such that $|\Psi'\rangle = A'|\Psi\rangle$. Define the space

$$\mathcal{H}'_\Psi = \mathcal{A}'|\Psi\rangle.$$

The space $\mathcal{H}'_\Psi$ is a subspace of $\mathcal{H}_\Psi$. What we need to prove is that actually $\mathcal{H}'_\Psi = \mathcal{H}_\Psi$. Define as P the projection operator on $\mathcal{H}_\Psi{}'$. It is clear that P commutes with all elements of $\mathcal{A}'$, hence $P \in (\mathcal{A}')' = \mathcal{A}$. This means that we have

$$(\mathcal{I} - P)|\Psi\rangle = 0,$$

and since $(\mathcal{I} - P) \in \mathcal{A}$ and since, by assumption, $|\Psi\rangle$ is separating for $\mathcal{A}$ we find that $P = \mathcal{I}$, or $\mathcal{H}'_\Psi = \mathcal{H}_\Psi$, which shows that $|\Psi\rangle$ is cyclic for the algebra $\mathcal{A}$.

[15]In this case, unlike the AMPS scenario, the measurement merely has to be sharp and not even fine-tuned.

Under the previous assumptions (i.e. that $\mathcal{A}$ is a von Neumann algebra acting on $\mathcal{H}_\Psi$ and that $|\Psi\rangle$ is cyclic and separating) we define the antilinear operator $S\colon \mathcal{H}_\Psi \to \mathcal{H}_\Psi$ by

$$SA|\Psi\rangle = A^\dagger|\Psi\rangle.$$

It is clear that

$$S^2 = 1,$$

and also

$$S|\Psi\rangle = |\Psi\rangle.$$

Since S is antilinear, the Hermitian conjugate operator is defined by

$$(|A\rangle, S^\dagger|B\rangle) = (|B\rangle, S|A\rangle).$$

We will now prove that for all $A' \in \mathcal{A}'$ we have

$$S^\dagger A'|\Psi\rangle = (A')^\dagger|\Psi\rangle. \tag{D1}$$

Proof: Consider any state $|B\rangle \in \mathcal{H}_\Psi$ and multiply both sides of (D1) with $\langle B|$. Since $|\Psi\rangle$ is cyclic we can write $|B\rangle = B|\Psi\rangle$ for $B \in \mathcal{A}$ and we have

$$\begin{aligned}\langle B|S^\dagger A'|\Psi\rangle &= (B|\Psi\rangle, S^\dagger A'|\Psi\rangle) = (A'|\Psi\rangle, SB|\Psi\rangle)\\ &= \langle\Psi|(A')^\dagger B^\dagger|\Psi\rangle = \langle\Psi|B^\dagger(A')^\dagger|\Psi\rangle\\ &= \langle B|(A')^\dagger|\Psi\rangle,\end{aligned}$$

which is indeed true.

From the previous item it follows that

$$S^\dagger|\Psi\rangle = |\Psi\rangle.$$

We define the linear operator $\Delta\colon \mathcal{H}_\Psi \to \mathcal{H}_\Psi$ by

$$\Delta = S^\dagger S.$$

From the previous results it is obvious that

$$\Delta|\Psi\rangle = |\Psi\rangle.$$

We show that Δ is Hermitian and positive (all eigenvalues strictly > 0).

Proof: Consider any two states $|A\rangle, |B\rangle$ of $\mathcal{H}_\Psi$ which can be written as $|A\rangle = A|\Psi\rangle$, $|B\rangle = B|\Psi\rangle$, with $A, B \in \mathcal{A}$. We have

$$\begin{aligned}\langle A|\Delta|B\rangle &= (A|\Psi\rangle, S^\dagger SB|\Psi\rangle) = (A|\Psi\rangle, S^\dagger B^\dagger|\Psi\rangle)\\ &= (B^\dagger|\Psi\rangle, SA|\Psi\rangle) = (B^\dagger|\Psi\rangle A^\dagger|\Psi\rangle),\end{aligned}$$

or to summarize

$$\langle A|\Delta|B\rangle = \langle\Psi|BA^\dagger|\Psi\rangle.$$

Similarly

$$\langle B|\Delta|A\rangle = \langle\Psi|AB^\dagger|\Psi\rangle.$$

From which we see that Δ is Hermitian. To prove that Δ is *strictly* positive we notice that any state in $\mathcal{H}_\Psi$ can be written as $|A\rangle = A|\Psi\rangle$ for some $A \neq 0$. We have

$$\langle A|\Delta|A\rangle = \langle\Psi|AA^\dagger|\Psi\rangle = ||A^\dagger|\Psi\rangle||^2 > 0,$$

since by assumption that $|\Psi\rangle$ is separating, $A^\dagger|\Psi\rangle \neq 0$.

Since Δ is Hermitian and strictly positive, it means we can define the inverse Δ^{-1} and all other powers Δ^z, $z \in \mathbf{C}$.

Now we show that for any $A \in \mathcal{A}$ and any $A' \in \mathcal{A}'$ we have

$$SAS \in \mathcal{A}', \tag{D2}$$

$$S^\dagger A' S^\dagger \in \mathcal{A}. \tag{D3}$$

Proof: Since $|\Psi\rangle$ is cyclic for $\mathcal{A}$, any vector $|B\rangle$ in $\mathcal{H}_\Psi$ can be written as $|B\rangle = B|\Psi\rangle$ with $B \in \mathcal{A}$. Consider any operator $C \in \mathcal{A}$. We have

$$(SAS)C|B\rangle = SAB^\dagger C^\dagger|\Psi\rangle = CBA^\dagger|\Psi\rangle,$$

and also

$$C(SAS)|B\rangle = CSAB^\dagger|\Psi\rangle = CBA^\dagger|\Psi\rangle.$$

Hence

$$[SAS, C] = 0,$$

as an operator, for all $C \in \mathcal{A}$, or $SAS \in \mathcal{A}'$. Similarly we prove $S^\dagger A' S^\dagger \in \mathcal{A}$.

We consider the polar decomposition of S as

$$S = J\Delta^{1/2}. \tag{D4}$$

Here we define $\Delta^{1/2}$ to have positive eigenvalues. We will now prove that

$$J\Delta^{1/2} = \Delta^{-1/2}J. \tag{D5}$$

Proof: Since we defined $J = S\Delta^{-1/2}$ this can also be written as

$$\Delta^{1/2}S = S\Delta^{-1/2}. \tag{D6}$$

Multiplying both sides from the left by $S^\dagger$ and using $S^\dagger S = \Delta$ we find that we have to prove equivalently

$$S^\dagger\Delta^{1/2}S = \Delta^{1/2}.$$

We argued before that the right-hand side of this equation is a positive operator. We briefly show that the left-hand side is also positive. For any nonvanishing state $|A\rangle = A|\Psi\rangle$ we notice that $S|A\rangle = A^\dagger|\Psi\rangle$ is also nonvanishing. Hence

$$(|A\rangle, S^\dagger\Delta^{1/2}S|A\rangle) = (A|\Psi\rangle, S^\dagger\Delta^{1/2}A^\dagger|\Psi\rangle)$$
$$= (\Delta^{1/2}A^\dagger|\Psi\rangle, SA|\Psi\rangle) = \langle\Psi|A\Delta^{1/2}A^\dagger|\Psi\rangle > 0,$$

since $\Delta^{1/2}$ is a strictly positive operator. This demonstrates that both sides of Eq. (D6) are strictly positive. Hence to prove that equation, we can just check that the square of the equation is true. The square of the left-hand side is

$$S^\dagger\Delta^{1/2}SS^\dagger\Delta^{1/2}S = S^\dagger\Delta^{1/2}\Delta^{-1}\Delta^{1/2}S = S^\dagger S = \Delta,$$

which is the square of the right-hand side, as we wanted to prove.

Now we prove that for any $A \in \mathcal{A}$ we have

$$\Delta A\Delta^{-1} \in \mathcal{A}.$$

Proof: We have

$$\Delta A\Delta^{-1} = S^\dagger SASS^\dagger = S^\dagger(SAS)S^\dagger.$$

From the relation (D2) we find that $SAS = A'$ for some $A' \in \mathcal{A}'$. But then from (D3) $S^\dagger A' S^\dagger \in \mathcal{A}$, as we wanted to prove.

By induction we can prove that

$$\Delta^m A\Delta^{-m} \in \mathcal{A}, \quad m = 0, 1, 2, \ldots$$

Actually we will now prove that

$$\Delta^z A\Delta^{-z} \in \mathcal{A}, \quad z \in \mathbf{C}.$$

Proof: To do this, we will show that for any $z \in \mathbf{C}$ the operator $\Delta^z A\Delta^{-z}$ commutes with all elements of $\mathcal{A}'$ and hence it belongs to $(\mathcal{A}')' = \mathcal{A}$. Consider any elements $A' \in \mathcal{A}'$. We will prove that the commutator $[\Delta^z A\Delta^{-z}, A']$ vanishes. Notice, we have already proved that it vanishes when $z =$ positive integer. Consider the matrix elements of this commutator on any two states $|\Psi_1\rangle, |\Psi_2\rangle$. We define the function

$$f(z) = \frac{1}{||\Delta||^{2z}}\langle\Psi_1|[\Delta^z A\Delta^{-z}, A']|\Psi_2\rangle.$$

Here we defined the norm $||\Delta||$ of the operator. Since Δ is a finite-dimensional (positive) matrix, the function $f(z)$ is a holomorphic function of z. It is zero at $z = 0, 1, 2, \ldots$ and does not grow too fast at infinity. Then by Carlson's theorem it is identically equal to zero. Hence for any z and any $A' \in \mathcal{A}'$ we have $[\Delta^z A\Delta^{-z}, A'] = 0$ and hence $\Delta^z A\Delta^{-z} \in (\mathcal{A}')' = \mathcal{A}$, as we wanted to prove.

This shows in particular that

$$\Delta^{1/2}A\Delta^{-1/2} \in \mathcal{A}. \tag{D7}$$

If we remember equations (D4), (D5), we can write $J = \Delta^{1/2}S = S\Delta^{-1/2}$. Hence

$$JAJ = \Delta^{1/2}SAS\Delta^{-1/2} \in \mathcal{A}.$$

by combining (D2) and (D7)

$$JAJ \in \mathcal{A}'$$

and similarly

$$JA'J \in \mathcal{A}.$$

So if we define the mirror operators as

$$\tilde{A} = JAJ,$$

then we see that they commute with the original operators.

Moreover we can easily show that any element in $\mathcal{A}'$ can be written as JAJ for some $A \in \mathcal{A}$ hence the previous inclusions are actually equalities

$$J\mathcal{A}J = \mathcal{A}', \quad J\mathcal{A}'J = \mathcal{A}.$$

Let us also summarize the other important result we derived above

$$\Delta^z\mathcal{A}\Delta^{-z} = \mathcal{A}, \quad \Delta^z\mathcal{A}'\Delta^{-z} = \mathcal{A}'.$$

The latter equations can be interpreted as follows. If we write $\Delta = e^{-K}$ these equations show that

$$e^{iKt}\mathcal{A}e^{-iKt} = \mathcal{A}, \qquad e^{iKt}\mathcal{A}'e^{-iKt} = \mathcal{A}',$$

so the two algebras $\mathcal{A}, \mathcal{A}'$ are closed under "time evolution" using the modular Hamiltonian K.

APPENDIX E: NUMERICALLY COMPUTING THE MIRROR OPERATORS IN THE SPIN CHAIN

In the main text, we have proved that the mirror operators exist under the appropriate conditions. Nevertheless, it is still fun to see this in an explicit numerical computation. The spin chain provides us with a nice toy model, in which we numerically compute the mirror operators and examine their matrix elements.

We include a MATHEMATICA program "spinchaintildes.nb" with the Supplemental Material available with this paper that performs this computation [41]. Here, we provide a few comments to help the reader understand this program.

The numbers involved.—In the text, we have proved that these mirror operators exist in the spin chain provided that we take the number of insertions K, so that we have

$$\mathcal{D}_{\mathcal{A}} = \sum_{j=0}^{K} \binom{\mathcal{N}}{j} 3^j \leq 2^{\mathcal{N}}. \tag{E1}$$

Note, that, in the text, we have always taken the dimension of the set $\mathcal{A}$ to be much smaller than that of the Hilbert space, $\mathcal{D}_{\mathcal{A}} \ll \mathcal{D}_{\mathcal{H}}$, to avoid issues with edge effects. However, as we see here, we actually need a much weaker condition, and in the case of the spin chain the precise condition is specified in (E1).

It is interesting to examine the numbers that are involved here. Even for $K = 2$, the first value of $\mathcal{N}$ for which $\mathcal{D}_{\mathcal{A}} \leq 2^{\mathcal{N}}$ is $\mathcal{N} = 9$. For $\mathcal{N} = 9$ and $K = 2$, we have $\mathcal{D}_{\mathcal{A}} = 277$ compared to $2^9 = 512$.

If we want to take $K = 3$, we see that we must take $\mathcal{N} \geq 14$. With $\mathcal{N} = 14$, we have $\mathcal{D}_{\mathcal{A}} = 10690$ compared to $2^{14} = 16384$. Since the algorithm below involves the inversion of a $\mathcal{D}_{\mathcal{A}} \times \mathcal{D}_{\mathcal{A}}$ matrix, we see that it rapidly becomes expensive to find the tildes for higher values of K.

Algorithm.—Now, we briefly describe the algorithm used to do the numerical computation.

First, we compute the set of all possible products of $\tilde{\mathbf{s}}_a^i$ operators, up to K operators. These products are put together in an array, that we can call ℓ_α here. It is clear that the index α ranges from $1 \ldots \mathcal{D}_{\mathcal{A}}$.

Now, we take a state

$$|\Psi\rangle = \sum_B \alpha_B |B\rangle,$$

where the α_B are chosen to be arbitrary complex coefficients, satisfying $|\alpha_B|^2 = 1$.

Then, we generate the set of vectors

$$|v_\alpha\rangle = \ell_\alpha |\Psi\rangle$$

Now, we have two choices. We can either compute the antilinear map S, and then generate the mirror operators, or else just solve Eqs. (3.35) and (3.36). Computing S might seem a little more efficient, because we need to compute this antilinear map only once, and then evaluate (6.3). (For the spin chain, $\Delta = 1$, and so $S = J$.) However, since we need to consider $S\mathbf{s}_a^i S \ell_i$, we see that we need to compute the action of S on a product of $K + 1$ spin operators. As we pointed out above, increasing K is expensive, so in this program we simply compute the mirror operators for each a, i separately.

Consider some particular i_0, a_0. So, we need to solve the equations

$$\tilde{\mathbf{s}}_{a_0}^{i_0} v_i = -\ell_i \mathbf{s}_{a_0}^{i_0} |\Psi\rangle \equiv |u_i\rangle.$$

Now, precisely as in (3.8) we consider the "metric" defined by

$$g_{ij} = \langle v_i | v_j \rangle,$$

and "invert" this metric to get g^{jk} satisfying

$$g^{jk} g_{ki} = \delta_i^j.$$

This is the numerically expensive step because this matrix is $\mathcal{D}_A$ dimensional. In terms of this metric, precisely as in (3.10), now we have simply

$$\tilde{\mathbf{s}}_{a_0}^{i_0} = g^{jk} |u_j\rangle \langle v_k|.$$

The reader can experiment with this program. These explicit numerical computations show, for example, that the commutator of the mirror operators with the ordinary operators can be a rather complicated matrix, but precisely annihilates the state and its descendants produced by acting with some number of ordinary operators.

[1] K. Papadodimas and S. Raju, J. High Energy Phys. 10 (2013) 212.

[2] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, J. High Energy Phys. 02 (2013) 062.

[3] R. Bousso, Phys. Rev. D **87**, 124023 (2013); Y. Nomura, J. Varela, and S. J. Weinberg, J. High Energy Phys. 03 (2013) 059; S. D. Mathur and D. Turton, J. High Energy Phys. 01 (2014) 034; B. D. Chowdhury and A. Puhm, Phys. Rev. D **88**, 063509 (2013); L. Susskind, arXiv:1208.3445; I. Bena, A. Puhm, and B. Vercnocke, J. High Energy Phys. 12 (2012) 014; A. Giveon and N. Itzhaki, J. High Energy Phys. 12 (2012) 094; T. Banks and W. Fischler, arXiv:1208.4757; A. Ori, arXiv:1208.6480; R. Brustein, arXiv:1209.2686; L. Susskind, arXiv:1210.2098; D. Marolf and A. C. Wall, Classical Quantum Gravity **30**, 025001 (2013); S. Hossenfelder, arXiv:1210.5317; Y. Nomura, J. Varela, and S. J. Weinberg, Phys. Rev. D **87**, 084050 (2013); D.-i. Hwang, B.-H. Lee, and D.-h. Yeom, J. Cosmol. Astropart. Phys. 01 (2013) 005; S. G. Avery, B. D. Chowdhury, and A. Puhm, J. High Energy Phys. 09 (2013) 012; K. Larjo, D. A. Lowe, and L. Thorlacius, Phys. Rev. D **87**, 104018 (2013); S. Braunstein, Phys. Rev. Lett. **110**, 101301 (2013); S. K. Rama,

arXiv:1211.5645; D. N. Page, J. Cosmol. Astropart. Phys. 04 (2013) 037; Y. Nomura and J. Varela, J. High Energy Phys. 07 (2013) 124; M. Saravani, N. Afshordi, and R. B. Mann, arXiv:1212.4176; T. Jacobson, arXiv:1212.6944; L. Susskind, arXiv:1301.4505; W. Kim, B.-H. Lee, and D.-h. Yeom, J. High Energy Phys. 05 (2013) 060; I. Park, arXiv:1301.6320; S. D. Hsu, arXiv:1302.0451; S. B. Giddings, Phys. Rev. D **88**, 024018 (2013); B.-H. Lee and D.-h. Yeom, Nucl. Phys. B, Proc. Suppl. **246–247**, 178 (2014); S. G. Avery and B. D. Chowdhury, arXiv:1302.5428; R. Brustein and A. Medved, J. High Energy Phys. 09 (2013) 108; B. Kang, arXiv:1305.2797; R. Brustein and A. Medved, J. High Energy Phys. 09 (2013) 015; B. D. Chowdhury, Int. J. Mod. Phys. D **22**, 1342011 (2013); J. Maldacena and L. Susskind, arXiv:1306.0533; D. N. Page, arXiv:1306.0562; M. Axenides, E. Floratos, and S. Nicolis, arXiv:1306.5670; M. Gary, arXiv:1307.4972; B. D. Chowdhury, J. High Energy Phys. 10 (2013) 034; A. de la Fuente and R. Sundrum, arXiv:1307.7738; A. Almheiri and J. Sully, arXiv:1307.8149; J. L. F. Barbon and E. Rabinovici, J. High Energy Phys. 12 (2013) 023; Y. Nomura, J. Varela, and S. J. Weinberg, Phys. Rev. D **88**, 084052 (2013); S. Lloyd and J. Preskill, arXiv:1308.4209; S. D. H. Hsu, arXiv:1308.5686; D. N. Page, J. Cosmol. Astropart. Phys. 09 (2013) 028; S. B. Giddings and Y. Shi, arXiv:1310.5700 [Phys. Rev. D (to be published)]; R. Brustein and A. Medved, arXiv:1310.5861; S. D. Mathur, arXiv:1309.6583.

[4] E. Verlinde and H. Verlinde, arXiv:1306.0516; arXiv:1306.0515; J. High Energy Phys. 10 (2013) 107.

[5] D. Harlow and P. Hayden, J. High Energy Phys. 06 (2013) 085.

[6] S. D. Mathur and D. Turton, arXiv:1306.5488.

[7] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, J. High Energy Phys. 09 (2013) 018.

[8] M. Van Raamsdonk, arXiv:1307.1796.

[9] D. Marolf and J. Polchinski, Phys. Rev. Lett. **111**, 171301 (2013).

[10] S. D. Mathur, Classical Quantum Gravity **26**, 224001 (2009).

[11] R. Bousso, Phys. Rev. Lett. **112**, 041102 (2014).

[12] K. Papadodimas and S. Raju, Phys. Rev. Lett. **112**, 051301 (2014).

[13] T. Banks, M. R. Douglas, G. T. Horowitz, and E. J. Martinec, arXiv:hep-th/9808016; V. Balasubramanian, S. B. Giddings, and A. E. Lawrence, J. High Energy Phys. 03 (1999) 001; I. Bena, Phys. Rev. D **62**, 066007 (2000); A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Phys. Rev. D **75**, 106001 (2007); **74**066009 (2006).

[14] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Phys. Rev. D **73**, 086003 (2006).

[15] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, arXiv:0710.4334; M. Van Raamsdonk, arXiv:0907.2939; Gen. Relativ. Gravit. **42**, 2323 (2010); Classical Quantum Gravity **28**, 065002 (2011); B. Czech, J. L. Karczmarek, F. Nogueira, and M. Van Raamsdonk, Classical Quantum Gravity **29**, 155009 (2012).

[16] I. Heemskerk, D. Marolf, J. Polchinski, and J. Sully, J. High Energy Phys. 10 (2012) 165.

[17] R. Bousso, B. Freivogel, S. Leichenauer, V. Rosenhaus, and C. Zukowski, Phys. Rev. D **88**, 064057 (2013); S. Leichenauer and V. Rosenhaus, Phys. Rev. D **88**, 026003 (2013).

[18] S. Hawking, Commun. Math. Phys. **43**, 199 (1975).

[19] J. Maldacena, arXiv:1105.5632.

[20] J. M. Maldacena, J. High Energy Phys. 04 (2003) 021.

[21] P. Candelas, Phys. Rev. D **21**, 2185 (1980).

[22] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**171 (1957).

[23] Y. Takahashi and H. Umezawa, Int. J. Mod. Phys. B **10**, 1755 (1996).

[24] D. N. Page, Phys. Rev. Lett. **71**, 1291 (1993).

[25] O. Aharony, S. Minwalla, and T. Wiseman, Classical Quantum Gravity **23**, 2171 (2006).

[26] N. Birrell and P. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, New York, 1986).

[27] S. D. Mathur, arXiv:1108.0302.

[28] M. Srednicki, J. Phys. A **32**, 1163 (1999); M. Srednicki, Phys. Rev. E **50**, 888 (1994).

[29] S. H. Shenker and D. Stanford, arXiv:1306.0622.

[30] N. Bogoliubov and D. Shirkov, *Introduction to the Theory of Quantized Fields*, translated by G. M. Vokoff (Interscience, New York, 1959).

[31] R. Haag, *Local Quantum Physics: Fields, Particles, Algebras* (Springer, New York, 1992), 2nd ed.

[32] O. Bratteli and D. Robinson, *C* and W* Algebras, Symmetry Groups, Decomposition of States*, Operator Algebras and Quantum Statistical Mechanics (Springer, New York, 1979), Vol. 1.

[33] O. Bratteli and D. Robinson, *Equilibrium States. Models in Quantum Statistical Mechanics*, Operator Algebras and Quantum Statistical Mechanics (Springer, Berlin, 1996), Vol. 2.

[34] I. Ojima, Ann. Phys. (N.Y.) **137**, 1 (1981); N. Landsman and C. van Weert, Phys. Rep. **145**, 141 (1987).

[35] D. Marolf and J. Polchinski, "Fuzz, Fire or Complementarity" workshop, KITP, 2013.

[36] S. B. Giddings, D. Marolf, and J. B. Hartle, Phys. Rev. D **74**, 064018 (2006).

[37] D. Kabat and G. Lifschytz, Phys. Rev. D **87**, 086004 (2013).

[38] I. Heemskerk, J. High Energy Phys. 09 (2012) 106.

[39] P. A. M. Dirac, *Lectures on Quantum Mechanics* (Dover, New York, 2001). Originally published by Belfer Graduate School of Science in 1964.

[40] P. Candelas and D. Deutsch, Proc. R. Soc. A **362**, 251 (1978).

[41] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevD.89.086010 for the MATHEMATICA program "spinchaintildes.nb" that automates the calculations in this paper.