Bulk reconstruction and non-isometry in the backwards-forwards holographic black hole map

Oliver DeWolfe 🆔 and Kenneth Higginbotham 🆔

Department of Physics and Center for Theory of Quantum Matter,
390 UCB University of Colorado Boulder,
CO 80309, U.S.A.

E-mail: oliver.dewolfe@colorado.edu, kenneth.higginbotham@colorado.edu

ABSTRACT: The backwards-forwards map, introduced as a generalization of the non-isometric holographic maps of the black hole interior of Akers, Engelhardt, Harlow, Penington, and Vardhan to include non-trivial dynamics in the effective description, has two possible formulations differing in when the post-selection is performed. While these two forms are equivalent on the set of dynamically generated states — states formed from unitary time evolution acting on well-defined initial configurations of infalling matter — they differ on the generic set of states necessary to describe the apparent world of the infalling observer. We show that while both versions successfully reproduce the Page curve, the version involving post-selection as the final step, dubbed the backwards-forwards-post-selection (BFP) map, has the desirable properties of being non-isometric but isometric on average and providing state-dependent reconstruction of bulk operators, while the other version does not. Thus the BFP map is a suitable non-isometric code describing the black hole interior including interior interactions.

Contents

1 Introduction

Recently, non-isometric codes have become a fascinating avenue for exploring how black hole interiors might be described by a theory of quantum gravity. The non-isometric holographic maps proposed by Akers, Engelhardt, Harlow, Penington, and Vardhan (PHEVA[1]) use post-selection to remove the extra degrees of freedom found in a semiclassical description of the interior [1]. These codes have been successfully used to derive the quantum extremal surface (QES) formula [2–6], obtain the Page curve [7], realize a construction of the Python's lunch [8–10], and provide a state-dependent reconstruction of interior operators [11], in a way consistent with expectations for the role of computational complexity [12]. The non-isometric property of these maps has also been observed in the context of two-dimensional conformal field theories and random tensor networks [13], further motivating their utility in understanding how semiclassical properties of black holes should be encoded in quantum gravity.

So far, realizations of non-isometric holographic maps — which take a black hole state in the *effective* (interior observer) description and map it to a *fundamental* (exterior observer) description — have been built on qudit toy models of black holes. Despite the inherent assumptions and simplifications in these qudit models, it is encouraging that non-isometric codes have been so successful even in these simplified contexts. The most detailed constructions

---

[1]We will continue to make use of this user-friendly permutation of the authors' initials.

involve building non-isometric maps out of the dynamics of the black hole itself. PHEVA constructed the first such dynamical holographic map out of the black hole dynamics in the fundamental description [1]. In doing so, they assumed that the effective dynamics were trivial — once inside the interior, nothing would happen to an infalling observer. Since an observer inside a large black hole should be able to do all sorts of physics experiments, it is natural to generalize this construction to include non-trivial interactions in the effective description.

The first step in this direction was taken by Kim and Preskill [14]. Concerned by issues surrounding post-selection that arose in the context of the final state proposal [15, 16], namely the possibility that the post-selection involved in the map could lead to violations of unitarity or to superpolynomial computational speedup violating the extended Church-Turing thesis [17–19], they included interactions between an infalling robot and radiation modes (both inside and outside the horizon) in the effective description and applied PHEVA's dynamical non-isometric map. Kim and Preskill found that the new interactions combined with the post-selection in the non-isometric map to create small violations to the unitarity of black hole evaporation, but that any computational speedups are limited by the size of the infaller's Hilbert space.

In a previous work [20], we took a further step by proposing that nontrivial interior dynamics should be reflected in the holographic map itself. We took the perspective that a holographic map from the effective to the fundamental description at a moment in time could be regarded as a composition of backwards time dynamics in the effective description until all infalling modes have been brought back outside the black hole, followed by forwards time dynamics in the fundamental description bringing them back in again. In the perspective of this "backwards-forwards" map, post-selection arises because Hawking pairs annihilate and return to the vacuum during backwards effective time evolution. Because the map incorporates effective dynamics directly, it restores unitarity of evaporation and as well as avoiding computational speedup problems typically associated with post-selection.

Our previous work [20] focused on the action of the backwards-forwards map on *dynamically generated states*: black hole states formed by the unitary evolution of initial configurations of non-singular infalling matter. These states form a subspace of the full effective Hilbert space. On this subspace, we showed that the backwards-forwards map can be transformed to a "post-selection version" which is equivalent to PHEVA's dynamical map when effective interactions are removed. All of the results in [20] are appropriate for this dynamically generated subspace.

However, PHEVA's non-isometric maps were applied to more than just the dynamically generated subspace. In particular, their calculations [1] showing that non-isometric holographic maps act isometrically on subexponentially complex states, and that their maps provide state-dependent reconstruction of bulk operators, required the maps' actions on generic states — general linear combinations of the dynamically generated states and the null states annihilated by the holographic map. These states cannot be reached by any dynamical perturbation of a black hole state formed from unitary collapse — no observer or robot could fall into the black hole and perform a unitary operation to take the black hole out of the dynamically generated subspace. However, they are states in the effective description Hilbert space, and a well-defined holographic map should have an action on all such states. Moreover, they can be perceived by an interior observer's measurement, considered as involving a unitary operation that entangles an observer and their apparatus with the system they are

measuring. This entangling unitary is dynamical and cannot take the black hole state out of the dynamically generated subspace alone; however, the observer could (purposefully or naively) choose a basis for their apparatus that involves null states outside of the dynamically generated subspace. Thus when they observe their apparatus, the observer could see the wavefunction collapse to a state inaccessible by dynamics alone.

Thus to demonstrate that the backwards-forwards holographic map is an appropriate interacting generalization of the maps of [1], and can successfully reproduce properties like isometry on subexponential states and bulk reconstruction, we must also investigate its action on generic states. This action must be treated carefully; for a generic state of the effective description, backwards time evolution will not return Hawking modes to the expected maximally entangled states before they annihilate. Such a mismatch could lead to singularities forming in the past.

We will see that when acting on generic states, the two presentations of the backwards-forwards map that coincided on dynamically generated states are no longer equivalent. The original formulation of the map involves post-selecting on the Hawking states as soon as they return to the vacuum, while the other version saves post-selection for the end, more closely analogous to the PHEVA map. The aim of this work is to determine which (if either) of the two versions of the backwards-forwards map can work well on generic states. This will involve making sure both maps are defined so as to avoid past singularities, and verifying whether they satisfy two criteria established by PHEVA in [1]: that the holographic map should act isometrically on generic states for Haar-typical fundamental dynamics, and that it should provide a reconstruction of interior unitary operators in the effective description up to exponentially small errors, in a fashion consistent with entanglement wedge reconstruction. We will find that the first formulation of the backwards-forwards map — which we refer to as the BF map — does not satisfy these criteria, while the alternate version — referred to here as the backwards-forwards-post-selection (BFP) map — does. Reproducing a number of calculations of [1], we see that while the derivation of the Page curve succeeds for both maps, isometry on average and bulk reconstruction follow only for the BFP map, in a way analogous to the original PHEVA calculations with generally small corrections.

This paper will be organized as follows. Section 2 will review the definitions of the holographic maps proposed in both [1] and [20] and establish the notational conventions taken here. We will also discuss how the backwards-forwards and BFP maps can avoid the creation of a past singularity during the backwards evolution. For completeness, we will demonstrate here that our maps can be used to obtain the same QES formula and Page curve as PHEVA's dynamical maps. Section 3 will review our results from [20] for the action of our maps on the dynamically generated subspace and provide new clarifying details. Section 4 will check the two criteria above for both the backwards-forwards and BFP maps. Finally, we will conclude in section 5 where we will discuss the assumptions made in this work and opportunities for future studies.

2 The backwards-forwards holographic maps

We begin by defining the black hole model, its effective description and fundamental description time dynamics, and the backwards-forwards holographic maps that will be used throughout the paper.

2.1 Effective and fundamental descriptions and dynamics

The two descriptions of the black hole system are taken to be described by systems of qudits [1] and thus are represented as finite-dimensional Hilbert spaces. The black hole degrees of freedom in the effective description are divided into left-moving (radially ingoing) modes ℓ and right-moving (radially outgoing) modes r. In the fundamental description, internal black hole degrees of freedom are labeled by B. Both descriptions share the same exterior: R_{in} describes infalling modes, and R_{out} contains the Hawking radiation. Thus we get the following factorizations of the effective and fundamental Hilbert spaces,

$$\mathcal{H}_{\mathrm{eff}} = \mathcal{H}_\ell \otimes \mathcal{H}_r \otimes \mathcal{H}_{R_{\mathrm{in}}} \otimes \mathcal{H}_{R_{\mathrm{out}}}, \qquad \mathcal{H}_{\mathrm{fun}} = \mathcal{H}_B \otimes \mathcal{H}_{R_{\mathrm{in}}} \otimes \mathcal{H}_{R_{\mathrm{out}}}. \tag{2.1}$$

When needed, we will include a reference system $\mathcal{H}_L$ in the effective description that keeps track of (and purifies) the ℓ modes that fell into the black hole from R_{in}. When referring to different Hilbert space factors we will often drop the $\mathcal{H}$ and just refer to them as ℓ, r and so forth.

The fundamental description characterizes the point of view of the external observer, who perceives the black hole as an object that accepts infalling degrees of freedom, scrambles them, and re-radiates them as Hawking radiation. This is pictured on the left-hand-side of figure 1; at time zero a set of modes falling in from R_{in} form the black hole, and are joined by a set of fixed degrees of freedom $|\psi\rangle_f$ coming down from high energies during the initial collapse. These are together processed by a unitary transformation U_0 into black hole modes B. At each succeeding time step, another unitary transformation U_t processes the modes in B and one new infalling degree of freedom from R_{in}, outputting two Hawking modes into R_{out} while the rest remain in B. If m_0 infallers from R_{in} are joined by $n_0 - m_0$ fixed degrees of freedom f, the unitaries describing the fundamental dynamics are maps,

$$U_t : \begin{cases} R_{\mathrm{in}}^{(m_0)} f^{(n_0 - m_0)} \to B^{(n_0)}, & t = 0 \\ R_{\mathrm{in}}^{(1)} B^{(n_0 + 1 - t)} \to B^{(n_0 - t)} R_{\mathrm{out}}^{(2)}, & t > 0 \end{cases} \tag{2.2}$$

where parenthetical superscripts have been used to denote the number of qudits from each factor. In [1], it was assumed that each U_t was drawn from the Haar measure. Recent work by Kim and Preskill [14] suggests that pseudorandom unitaries could offer a more realistic description of each U_t.

The effective description describes the semiclassical gravity picture seen by an observer behind the black hole horizon. In this description, modes fall into the black hole and are joined at each time step by the creation of Hawking pairs, one particle in the pair remaining in r inside the horizon and the other escaping into R_{out} as Hawking radiation. These Hawking pairs are created in the maximally entangled state,

$$|\mathrm{MAX}\rangle_{r,R_{\mathrm{out}}} = \frac{1}{\sqrt{q}} \sum_{j=1}^q |j\rangle_r |j\rangle_{R_{\mathrm{out}}}, \tag{2.3}$$

where q is the dimension of the qudits, reflecting the assumed smoothness of spacetime at the horizon. Dealing with this pair creation backwards will be a subtlety for backwards
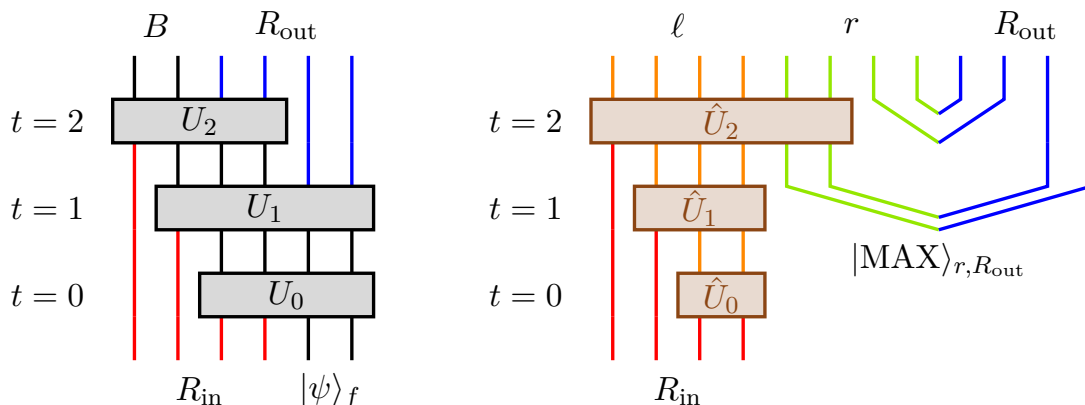
Figure 1. An example of fundamental dynamics U_t (left) and effective dynamics $\hat{U}_t$ (right) until $t = 2$ for $n_0 = 4$ and $m_0 = 2$. Lines are colored by qudit type: red for $R_{\rm in}$, blue for $R_{\rm out}$, black for B and f, orange for ℓ, and green for r.

time evolution. In addition, our previous work [20] introduced new non-trivial dynamics $\hat{U} \equiv \hat{U}_t \hat{U}_{t-1} \ldots \hat{U}_0$ for the interior of the black hole in the effective description. Each $\hat{U}_t$ acts on effective degrees of freedom as

$$
\hat{U}_t \; : \; \begin{cases} R_{\rm in}^{(m_0)} \to \ell^{(m_0)}, & t = 0 \\ R_{\rm in}^{(1)} \ell^{(m_0-1+t)} r^{(2(t-1))} \to \ell^{(m_0+t)} r^{(2(t-1))}, & t > 0 \,. \end{cases} \tag{2.4}
$$

An example of these dynamics is shown in the right panel of figure 1. The $\hat{U}_t$ represent the semiclassical gravity interactions perceived by the infalling observer, but for this toy model we have so far made no restrictions on their precise form; later on we will see that some assumptions about their average properties will be useful.

2.2 Non-isometric codes for the black hole interior

At any moment in time t, we expect there to be a holographic map V_t:

$$
V_t : \mathcal{H}_{\rm eff} \to \mathcal{H}_{\rm fun} \tag{2.5}
$$

carrying a configuration in the effective description to a configuration in the fundamental description, and acting as the identity on $R_{\rm in} \otimes R_{\rm out}$. The generic non-isometric map $V : \ell r \to B$ defined in [1] takes the form

$$
V = \sqrt{|P|} \langle \phi |_P U_{\rm H} | \psi \rangle_f \,, \tag{2.6}
$$

where $U_{\rm H}$ is a typical block unitary drawn from the Haar measure with no additional structure, $|\psi\rangle_f$ is the insertion of fixed degrees of freedom, and $\langle \phi |_P$ represents post-selecting a subsystem P (the complement of B in the output of $U_{\rm H}$) on some specified state. The prefactor $\sqrt{|P|}$ is included to ensure correct normalization of the state after post-selection.

The authors of [1] further provided a realization of a map of the form (2.6) constructed out of the fundamental description dynamics: modes in ℓ are fed into fundamental time
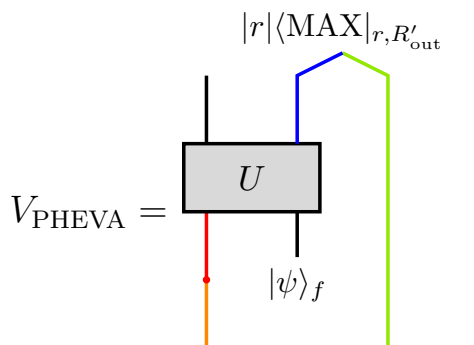
Figure 2. A schematic representation of the PHEVA dynamical holographic map for the black hole interior [1]. Unlike in figure 1, each line here represents all qudits in that factor of the Hilbert space.

evolution, while modes in r are post-selected together with the R_{out} modes coming out of the fundamental dynamics, giving a map of the form

$$V_{\text{PHEVA}} = |r| \langle \text{MAX}|_{r,R'_{\text{out}}} U |\psi\rangle_f . \tag{2.7}$$

Here there are two copies of the Hawking radiation R_{out}, and we label the one that is output of the fundamental dynamics as R'_{out}. When the input rR_{out} modes are in the maximally entangled state $|\text{MAX}\rangle_{r,R'_{\text{out}}}$, the map functions as a teleportation protocol, transferring the state of the degrees of freedom in R'_{out} to those in R_{out}. Figure 2 gives a circuit representation of PHEVA's dynamical holographic map.

The PHEVA map was possible for a case with no effective description dynamics, where each mode in ℓ can be identified uniquely with a mode in R_{in}; but in general, it is unnatural to use the state of the ℓ modes existing well beyond the horizon as an input to fundamental dynamics, which expects modes just crossing the horizon. The "backwards-forwards" map of [20] was designed to generalize V_{PHEVA} and resolve this issue, functioning as a holographic map taking into account effective description dynamics.

2.3 The backwards-forwards (BF) map

The philosophy of the backwards-forwards (BF) map is that the holographic map is a composition of backwards time dynamics in the effective description, followed by forward time dynamics in the fundamental description. Beginning with a state on $\ell r R_{\text{out}}$ at some time step t, the map first evolves these effective degrees of freedom backwards in the effective description using $\hat{U}^\dagger$. At each time step in the backwards evolution, one qudit from ℓ leaves the black hole and rejoins R_{in}. Furthermore, at each time-step two Hawking pairs in the maximally entangled state disappear. Once the backwards evolution is complete and there are no more black hole degrees of freedom, the map performs forwards fundamental time evolution U on R_{in} to bring the black hole into the fundamental description. Thus the map is aware of both kinds of dynamics.

Since the forward time evolution of the effective description is $\hat{U}|\text{MAX}\rangle_{r,R_{\text{out}}}$ where we have concatenated the $\hat{U}_t$ at each time step to a single $\hat{U}$, and the forward time evolution of the fundamental description is $U|\psi\rangle_f$ with the U_t concatenated into U, we obtain the backwards-forwards (BF) map [20],

$$V_{\text{BF}} = U|\psi\rangle_f \langle \text{MAX}|_{r,R_{\text{out}}} \hat{U}^\dagger . \tag{2.8}$$
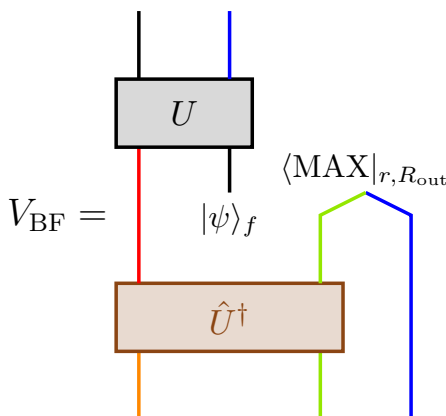
Figure 3. A schematic representation of the backwards-forwards (BF) map defined in equation (2.8). Unlike in figure 1, each line here represents all qudits in that factor of the Hilbert space.
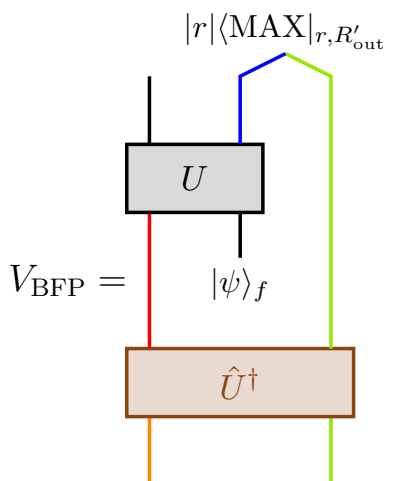

Figure 4. A schematic representation of the backwards-forwards-post-selection (BFP) map defined in equation (2.9). Unlike in figure 1, each line here represents all qudits in that factor of the Hilbert space.

Inverting the effective dynamics in this way has produced a post-selection $\langle \mathrm{MAX}|_{r,R_{\mathrm{out}}}$ that removes the Hawking pairs as we evolve back to their moment of creation. Equivalently, we can imagine decoupling each pair from the dynamics as we get back to its time of creation, and then post-selecting on all of them at $t = 0$. The backwards-forwards map is presented in figure 3.

2.4 The backwards-forwards-post-selection (BFP) map

The backwards-forwards map presented in equation (2.8) and figure 3 is not obviously a generalization of the PHEVA map V_{PHEVA} (2.7), which involves post-selection on the output of the fundamental dynamics. However, when the state acted on by the BF map (2.8) is a *dynamically generated state* — a state that evolved from non-singular initial conditions of matter infalling into the black hole — the BF map can be recast in another form which is of the PHEVA type. For such a dynamically generated state, the backwards time evolution simply follows the state's original generation in reverse, and each Hawking pair is guaranteed to be in the maximally entangled state at the time-step it disappears. Thus if we think of

decoupling and freezing each Hawking pair at the appropriate time-step, when we get to $t = 0$ the rR_{out} factor of the Hilbert space will indeed be in the state $|\text{MAX}\rangle_{r,R_{\text{out}}}$. Rather than post-selecting on it, we can keep it hanging around (we can imagine storing it in a "quantum memory") while the forwards time evolution of the fundamental description takes place. Since the fundamental dynamics also has output Hawking radiation, we will have two factors of R_{out}, and we label the one coming from the fundamental dynamics as R'_{out}. Finally after the fundamental dynamics is complete, we post-select on r which has remained in our quantum memory with R'_{out}, rather than with R_{out}. As with the PHEVA map (2.7), this functions as a teleportation protocol, transferring the state of the degrees of freedom in R'_{out} to those in R_{out}. This defines the backwards-forwards-post-selection (BFP) version of the map,

$$V_{\text{BFP}} = |r|\langle\text{MAX}|_{r,R'_{\text{out}}} U|\psi\rangle_f \hat{U}^\dagger, \tag{2.9}$$

where again the factor of $|r|$ is included for normalization, as illustrated in figure 4. It is of the general PHEVA form (2.6), and when the effective dynamics is trivial $\hat{U} = \mathbb{1}$, it reduces to the dynamical holographic map V_{PHEVA} (2.7).

2.5 Action of the maps on non-dynamically generated states

When we are imagining acting the maps V_{BF} and V_{BFP} on states that were created by forward time evolution from non-singular initial configurations of matter, the backwards time evolution acts in a straightforward way that just reverses this evolution, and the two maps are equivalent. However, it is well-known that the Hilbert space of the effective description after the Page time will obtain many more degrees of freedom than exist in the fundamental description. Many of these states are null states, expected to be annihilated by the holographic map. A generic state will be a linear combination of dynamically generated states and null states. We need to be somewhat careful about the definition of the backwards evolution on one of these generic states, due to the Hawking modes coming into and out of the spectrum.

One way to think about the generation of the Hawking pairs is the following. The interior of a black hole grows as it evolves in time, and as it does, spacetime near the horizon expands. Similar to adiabatically increasing the width of a square well, the energy levels of quantum fields on this expanding spacetime will be red-shifted, bringing field modes down from high energies. These new modes may be occupied, which we interpret as the generation of Hawking radiation.[2]

During the backwards evolution of a black hole state in the effective description, radiation modes will be blue-shifted back to high energy scales. We assume that if these radiation modes are occupied in the particular way that they are generated (i.e. maximally entangled) they will disappear nicely at high energies, exiting smoothly from the effective theory and no longer participating in interactions with modes remaining at lower energies. However, if we begin with a generic configuration in the effective description, radiation modes may approach high energy scales in the wrong state during the backwards evolution. One may

---

[2]In principle, a well-defined theory of semiclassical gravity requires an ultraviolet cutoff to keep fluctuations of the stress-energy tensor (and therefore the metric) finite when including backreaction [21]. We will not need an explicit ultraviolet cutoff here, but having such a cutoff in mind may be conceptually useful when considering these Hawking modes descending from high energies.

anticipate this could cause some catastrophe in the UV, creating a singularity (or white hole) in the past [1, 21, 22].

Let us consider how to properly implement the two maps to avoid such singularities. The BF map is more straightforward: it involves post-selection on each Hawking pair being in the correct maximally entangled state that it would be generated in. Thus if a particular set of Hawking modes is in the "wrong" state as they are lifted back up out of the spectrum, post-selection simply annihilates the state.

For the BFP map, however, we do not immediately post-select on the Hawking modes as they leave the spectrum; we wish to keep them around to post-select after the fundamental dynamics. Thus, we define the BFP map such that as each set of Hawking modes is lifted out of the spectrum, it is frozen (stored in our "quantum memory") and undergoes no further evolution. We may think of this memory as a record of the occupation of the various modes as they leave the effective theory. Since the modes are frozen at this point, they do not evolve to arbitrary high energies and no singularity will form. We keep these degrees of freedom until they are needed for the post-selection at the end of the BFP map. Since we found in [20] that exterior interactions drop out of the backwards-forwards map, we can also freeze out modes that leave the black hole when they are converted from ℓ back to R_{in} during the backwards evolution.

The backwards effective evolution portion of the BFP map is completed once all modes have been frozen as either infalling matter R_{in} or radiation rR_{out}. At this point, we feed R_{in} into the fundamental dynamics at the appropriate time-steps. The fundamental dynamics should be performed for a length of time given by the size of radiation modes frozen out during the backwards evolution, $t = (1/2)\log_q |R_{\text{out}}|$. The fundamental dynamics will produce black hole degrees of freedom B and radiation R'_{out}. The backwards-forwards map then performs post-selection on the stored r qudits and the radiation output of the fundamental dynamics R'_{out}.

Once we are considering the action of the BF and BFP maps on generic states, it is not obvious that they are equivalent. Indeed, we will find that they are not. In the next section, we consider the action of the maps on dynamically generated states in more detail, and show that they both can be used to derive the Page curve. In the section following the next, we consider their action on generic states and show that only the BFP map successfully reproduces isometry on average and state-dependent reconstruction of bulk states.

Let us make a few remarks before turning to the next section. We note that this implementation of freezing out the Hawking modes during the holographic map is suitable for this kind of toy model of the black hole, but a deeper description would involve understanding the effective description degrees of freedom as modes in semiclassical gravity and understanding the role of backreaction on the geometry. Additionally, we emphasize that these holographic maps are just maps between two different descriptions of the black hole at a moment in time. While they are constructed using time evolution unitaries, they do not represent the dynamics or experiences of any observer. In particular, post-selection is not part of the effective dynamics of the black hole.

The color scheme for labeling different factors of the Hilbert space in our circuit diagrams is summarized in figure 5. In addition, we will denote density matrices and pure states in the fundamental description with an uppercase Ψ_i, $|\Psi_i\rangle$, with the index i labeling an

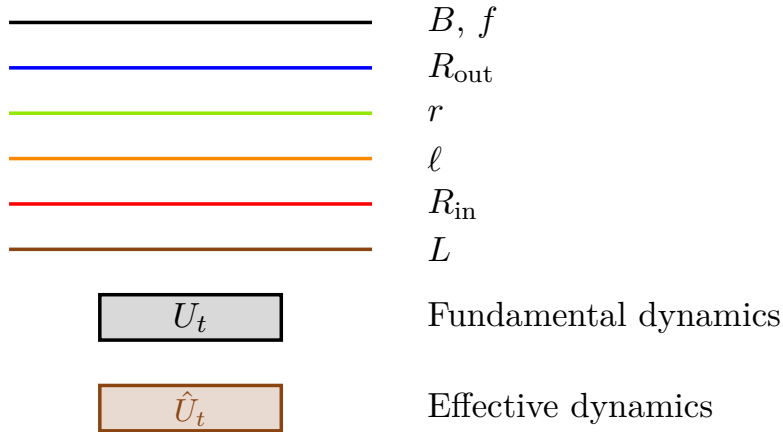**Figure 5.** The color scheme used to identify qudits and dynamics in circuit diagrams.

orthonormal basis $\{|\Psi_i\rangle\}$ of $\mathcal{H}_{\text{fun}}$,

$$\langle\Psi_j|\Psi_i\rangle = \delta_{ij}. \tag{2.10}$$

States in the effective description will be distinguished with a hat: $\hat{\Psi}_i$, $|\hat{\Psi}_i\rangle$. Again, i labels an orthonormal basis $\{|\hat{\Psi}_i\rangle\}$ of $\mathcal{H}_{\text{eff}}$. We will leave hats off of states of R_{in} since they are shared by both descriptions.

3 Action on the dynamically generated subspace

Now that we have defined our holographic maps, we begin with the action of V_{BF} and V_{BFP} on dynamically generated states. Here we will review essential results from our previous work in [20] and present new results to help clarify other properties of these maps.

We will show that both the V_{BF} and V_{BFP} act isometrically on dynamically generated states. This fact has important consequences. Work by Kim and Preskill [14] brought forward concerns that the post-selection present in the dynamical holographic map proposed by [1] might lead to superpolynomial computational complexity, violating the quantum extended Church-Turing thesis [17–19]. Post-selection has also been shown to lead to undesirable artifacts such as superluminal signaling and speedups to Grover's algorithm [23]. By including non-trivial effective dynamics in the backwards-forwards map directly, we have shown that these maps do act isometrically on dynamically generated states despite the presence of post-selection in the map. Therefore, both backwards-forwards maps avoid these issues typically associated with post-selection.

While both maps V_{BF} and V_{BFP} act isometrically on dynamically generated states of the effective description, only V_{BF} acts unitarily. We show that $V_{\text{BFP}}^\dagger$ in general maps dynamically generated states of the fundamental description to generic states of the effective description, and thus is not itself an isometry. Projecting the result of $V_{\text{BFP}}^\dagger$ on the dynamically generated subspace restores unitarity, however. We also show that both V_{BF} and V_{BFP} are equivariant, commuting with time evolution for the two descriptions. We conclude the section with a demonstration that both maps successfully reproduce the Page curve.

In this section we will always assume that black hole states are generated from unitary dynamics and are a part of the dynamically generated subspace. In the effective description, we define a dynamically generated state using the effective dynamics,

$$|\hat{\Psi}_i(t)\rangle_{\ell r R_{\mathrm{out}}} \equiv \hat{U}|\Psi_i\rangle_{R_{\mathrm{in}}}|\mathrm{MAX}\rangle_{r,R_{\mathrm{out}}}, \tag{3.1}$$

where we've included (t) in the ket to indicate this state is an element of the dynamically generated subspace and to emphasize the time dependent nature of these states. Dynamically generated states in the effective description are characterized by the initialization of the Hawking pairs rR_{out} in the maximally entangled state. Thus $|\hat{\Psi}(t)\rangle$ lives in a subspace of dimension $|R_{\mathrm{in}}|$.

We also define a dynamically generated subspace of the fundamental description. We imagine that the fixed states f are completely determined by the conditions of the collapse that formed the black hole; therefore, they carry no additional information than what is already in R_{in}. We define a dynamically generated state of the fundamental description using the fundamental dynamics,

$$|\Psi_i(t)\rangle_{BR_{\mathrm{out}}} \equiv U|\Psi_i\rangle_{R_{\mathrm{in}}}|\psi\rangle_f. \tag{3.2}$$

The dynamically generated subspace of the fundamental description defined by these states has dimension $|R_{\mathrm{in}}|$, the same size as the dynamically generated subspace of the effective description.

It will be useful to define projectors onto these dynamically generated subspaces. For the effective description, we define

$$\hat{P}_t \equiv \sum_i |\hat{\Psi}_i(t)\rangle\langle\hat{\Psi}_i(t)|, \tag{3.3}$$

with the sum ranging over an orthonormal basis $\{|\hat{\Psi}_i(t)\rangle\}$ of the dynamically generated subspace. Similarly, we define a projector onto the dynamically generated subspace of the fundamental description,

$$P_t \equiv \sum_i |\Psi_i(t)\rangle\langle\Psi_i(t)|, \tag{3.4}$$

with the sum ranging over a similar orthonormal basis $\{|\Psi_i(t)\rangle\}$. We note that both of these projectors are time dependent since the dynamically generated states they are constructed from depend on the unitary dynamics U_t and $\hat{U}_t$.

3.1 The BF map acts unitarily

We begin with the action of the backwards-forwards map V_{BF} (2.8) on the dynamically generated states (3.1) of the effective description. Recall that the backwards-forwards map begins by performing backwards effective evolution. This backwards evolution includes the annihilation of Hawking pairs to the vacuum, which we interpret as post-selection on $\langle\mathrm{MAX}|_{r,R_{\mathrm{out}}}$. Thus acting on (3.1) with V_{BF}, we find

$$\begin{aligned} V_{\mathrm{BF}}|\hat{\Psi}_i(t)\rangle_{\ell r R_{\mathrm{out}}} &= \left(U|\psi\rangle_f\langle\mathrm{MAX}|_{r,R_{\mathrm{out}}}\hat{U}^\dagger\right)\hat{U}|\Psi_i\rangle_{R_{\mathrm{in}}}|\mathrm{MAX}\rangle_{r,R_{\mathrm{out}}} \\ &= U|\Psi_i\rangle_{R_{\mathrm{in}}}|\psi\rangle_f \\ &= |\Psi_i(t)\rangle_{BR_{\mathrm{out}}}. \end{aligned} \tag{3.5}$$

V_{BF} maps a dynamically generated state of the effective description to a dynamically generated state of the fundamental description. Similarly, we can ask how the adjoint of the backwards-forwards map,

$$V_{\mathrm{BF}}^{\dagger} = \hat{U}|\mathrm{MAX}\rangle_{r,R_{\mathrm{out}}}\langle\psi|_{f}U^{\dagger} \tag{3.6}$$

acts on dynamically generated states of the fundamental description. $V_{\mathrm{BF}}^{\dagger}$ acts on states in the fundamental description, first evolving them backwards using fundamental dynamics. Post-selection is performed on f being in the fixed state $\langle\psi|_{f}$ as they are blue-shifted to high energies in the backwards evolution. Maximally entangled radiation is then inserted on rR_{out} before forwards effective dynamics are applied. Acting on (3.2), the adjoint gives

$$\begin{aligned}
V_{\mathrm{BF}}^{\dagger}|\Psi_{i}(t)\rangle_{BR_{\mathrm{out}}} &= \left(\hat{U}|\mathrm{MAX}\rangle_{r,R_{\mathrm{out}}}\langle\psi|_{f}U^{\dagger}\right)U|\Psi_{i}\rangle_{R_{\mathrm{in}}}|\psi\rangle_{f}\\
&= \hat{U}|\Psi_{i}\rangle_{R_{\mathrm{in}}}|\mathrm{MAX}\rangle_{r,R_{\mathrm{out}}}\\
&= |\hat{\Psi}_{i}(t)\rangle_{\ell rR_{\mathrm{out}}}.
\end{aligned} \tag{3.7}$$

$V_{\mathrm{BF}}^{\dagger}$ then maps a dynamically generated state of the fundamental description to a dynamically generated state of the effective description. We see that the backwards-forwards map and its adjoint take states back and forth between the dynamically generated subspaces of both descriptions; it never maps a dynamically generated state to a generic state of the other Hilbert space. Furthermore, inner products between states in each dynamically generated subspace are preserved, and V_{BF} acts unitarily on the dynamically generated subspace of both descriptions. We express this fact as operator equations using the projectors (3.3) and (3.4),

$$\hat{P}_{t}V_{\mathrm{BF}}^{\dagger}V_{\mathrm{BF}}\hat{P}_{t} = \hat{P}_{t} \tag{3.8}$$

$$P_{t}V_{\mathrm{BF}}V_{\mathrm{BF}}^{\dagger}P_{t} = P_{t} \tag{3.9}$$

It is good that V_{BF} acts unitarily, since the backwards-forwards map is constructed directly from backwards and forwards unitary evolution. Furthermore, the above results are exact — no averaging over U or $\hat{U}$ was needed to find them.

We also note that the backwards-forwards map is equivariant: effective description time evolution followed by the holographic map is equivalent to the holographic map followed by fundamental description time evolution,

$$V_{t+1}\hat{U}_{t+1} = U_{t+1}V_{t}. \tag{3.10}$$

This was an important property of the dynamical non-isometric code first proposed by PHEVA in [1]. V_{BF} is manifestly equivariant, as can be seen by substituting (2.8) into (3.10).

3.2 The BFP map acts isometrically

We now turn to the action of the backwards-forwards-post-selection (BFP) map (2.9) on the dynamically generated subspace. The fact that the post-selection acts at the end of the map on rR_{out}', instead of earlier on rR_{out}, will have interesting consequences for the action of the BFP map on dynamically generated states.

First, we check the action of V_{BFP} on dynamically generated states of the effective description (3.1),

$$\begin{aligned}
V_{\mathrm{BFP}}|\hat{\Psi}_{i}(t)\rangle_{\ell rR_{\mathrm{out}}} &= |r|\left(\langle\mathrm{MAX}|_{r,R_{\mathrm{out}}'}U|\psi\rangle_{f}\hat{U}^{\dagger}\right)\hat{U}|\Psi_{i}\rangle_{R_{\mathrm{in}}}|\mathrm{MAX}\rangle_{r,R_{\mathrm{out}}}\\
&= |r|\langle\mathrm{MAX}|_{r,R_{\mathrm{out}}'}|\Psi_{i}(t)\rangle_{BR_{\mathrm{out}}'}|\mathrm{MAX}\rangle_{r,R_{\mathrm{out}}}.
\end{aligned} \tag{3.11}$$

Using a circuit representation,

$$V_{\text{BFP}}|\hat{\Psi}_i(t)\rangle_{\ell r R_{\text{out}}} = \quad\quad , \quad\quad (3.12)$$

we see that V_{BFP} maps a dynamically generated state of the effective description to a dynamically generated state of the fundamental description followed by a teleportation protocol acting on the radiation degrees of freedom. We can remove the teleportation by unbending the radiation lines, finding

$$V_{\text{BFP}}|\hat{\Psi}_i(t)\rangle_{\ell r R_{\text{out}}} = |\Psi_i(t)\rangle_{B R_{\text{out}}}. \quad\quad (3.13)$$

V_{BFP} indeed maps states in the dynamically generated subspace of the effective description to dynamically generated states of the fundamental description. Furthermore, inner products of these states are preserved. Taking an inner product of (3.12) with itself gives

$$\langle\hat{\Psi}_j(t)|V_{\text{BFP}}^{\dagger}V_{\text{BFP}}|\hat{\Psi}_i(t)\rangle = |r|^2 \quad\quad (3.14)$$

Straightening the green and blue lines adds a factor of $1/|r|^2$, which cancels with the factor of $|r|^2$ inserted to maintain normalization after post-selection. Thus, (3.14) reduces to

$$\left(\langle\Psi_j|_{R_{\text{in}}}\langle\psi|_f\right)U^{\dagger}U\left(|\Psi_i\rangle_{R_{\text{in}}}|\psi\rangle_f\right) = \delta_{ij}. \quad\quad (3.15)$$

Indeed, the BFP map acts isometrically on dynamically generated states in the effective description, as would be expected from a generic teleportation protocol. In fact, this result did not require any averaging over U or $\hat{U}$ — inner products of states given by (3.1) are preserved exactly.

We now turn to the action of the adjoint of the BFP map,

$$V_{\text{BFP}}^{\dagger} = |r|\hat{U}\langle\psi|_f U^{\dagger}|\text{MAX}\rangle_{r,R_{\text{out}}'} \quad\quad (3.16)$$

on dynamically generated states of the fundamental description (3.2). The adjoint of the BFP map differs from the adjoint of the backwards-forwards map in an important way: instead of using radiation R_{out} from the dynamically generated fundamental state in the backwards evolution of the fundamental description, $V_{\text{BFP}}^{\dagger}$ applies $U^{\dagger}$ to the maximally entangled radiation R'_{out} inserted at the beginning of the map. The remainder of the adjoint map works similarly, with post-selection on the backwards evolved fixed states $\langle\psi|_f$. Acting on (3.2), the adjoint gives

$$V_{\text{BFP}}^{\dagger}|\Psi_i(t)\rangle_{BR_{\text{out}}} = \left(|r|\hat{U}\langle\psi|_f U^{\dagger}|\text{MAX}\rangle_{r,R'_{\text{out}}}\right) U|\Psi_i\rangle_{R_{\text{in}}}|\psi\rangle_f = |r|$$
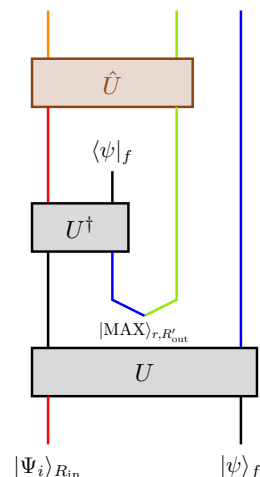
$$\tag{3.17}$$

We see that r will not be in the maximally entangled state with R_{out} before the application of $\hat{U}$ (as is required of a dynamically generated state) because it was inserted in the maximally entangled state with R'_{out} instead. Therefore, $V_{\text{BFP}}^{\dagger}$ does not produce a state with support only in the dynamically generated subspace,

$$V_{\text{BFP}}^{\dagger}|\Psi_i(t)\rangle_{BR_{\text{out}}} \neq |\hat{\Psi}_i(t)\rangle_{\ell r R_{\text{out}}}. \tag{3.18}$$

Instead, $V_{\text{BFP}}^{\dagger}$ maps dynamically generated states of the fundamental description to generic states of the effective Hilbert space. Furthermore, V_{BFP} does not even preserve the inner product of dynamically generated fundamental states. Taking the inner product of (3.17) with itself and averaging over U using (A.3), we find

$$\int dU \, \langle\Psi_j(t)|V_{\text{BFP}}V_{\text{BFP}}^{\dagger}|\Psi_i(t)\rangle = \frac{|r|^2}{|B|^2|r|^2-1}\left(|B|\cdot|\ell|\cdot|r|+|B|^2-1-\frac{|B|\cdot|\ell|}{|r|}\right)\delta_{ij}, \tag{3.19}$$

which generically deviates from unit norm. $V_{\text{BFP}}^{\dagger}$ does not act isometrically on dynamically generated states of the fundamental description, and therefore V_{BFP} cannot act as a unitary on the dynamically generated subspaces of both descriptions. This emphasizes the fact that the BFP map does not represent time dynamics in either description. As before, we can summarize these results as operator equations:

$$\hat{P}_t V_{\text{BFP}}^{\dagger} V_{\text{BFP}} \hat{P}_t = \hat{P}_t \tag{3.20}$$

$$\int dU \, P_t V_{\text{BFP}} V_{\text{BFP}}^{\dagger} P_t = \frac{|r|^2}{|B|^2|r|^2-1}\left(|B|\cdot|\ell|\cdot|r|+|B|^2-1-\frac{|B|\cdot|\ell|}{|r|}\right)P_t \tag{3.21}$$
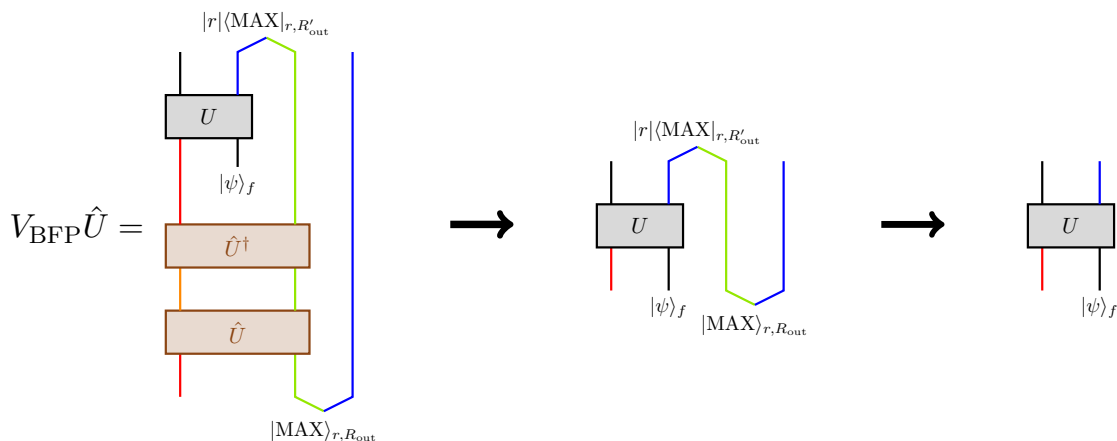
– 14 –

Figure 6. Demonstrating equivariance is satisfied by V_{BFP}. The left figure shows time evolution in the effective description composed with the BFP map.

While the result of $V_{\mathrm{BFP}}^{\dagger}$ is not a dynamically generated state, it does have non-zero overlap with the dynamically generated subspace. We could force the result of $V_{\mathrm{BFP}}^{\dagger}$ into the dynamically generated subspace of the effective description by applying the projector (3.3),

$$\hat{P}_t V_{\mathrm{BFP}}^{\dagger}|\Psi_i(t)\rangle_{BR_{\mathrm{out}}} = |\hat{\Psi}_i(t)\rangle_{\ell r R_{\mathrm{out}}}. \tag{3.22}$$

In fact, adding the projector restores preservation of the inner product without the need for averaging,

$$\langle\Psi_j(t)|V_{\mathrm{BFP}}\hat{P}_t V_{\mathrm{BFP}}^{\dagger}|\Psi_i(t)\rangle = \delta_{ij}, \tag{3.23}$$

where we used $\hat{P}_t^{\dagger} = \hat{P}_t^2 = \hat{P}_t$. Therefore $\hat{P}_t V_{\mathrm{BFP}}^{\dagger}$ acts isometrically, whereas $V_{\mathrm{BFP}}^{\dagger}$ alone does not,

$$P_t V_{\mathrm{BFP}}\hat{P}_t V_{\mathrm{BFP}}^{\dagger}P_t = P_t. \tag{3.24}$$

The BFP map is also equivariant, although demonstrating that it satisfies (3.10) is not as straightforward as it was for the backwards-forwards map. Figure 6 demonstrates that composing effective dynamics with V_{BFP} indeed gives fundamental dynamics, proving equivariance.[3] Our work in [20] went further to show that equivariance still holds for V_{BFP} in the presence of exterior interactions between R_{in} and R_{out}, even though these exterior interactions drop out of the holographic map.

3.3 Relationship between BF and BFP

Before turning to the derivation of the Page curve, we compare the two maps V_{BF} and V_{BFP} on dynamically generated states. One of the most important results in our previous work [20] demonstrated the equivalence of V_{BF} and V_{BFP} on dynamically generated states of the effective description. The transformations relating the two are reproduced in figure 7, and we can represent the result as an operator equation:

$$V_{\mathrm{BF}}\hat{P}_t = V_{\mathrm{BFP}}\hat{P}_t. \tag{3.25}$$

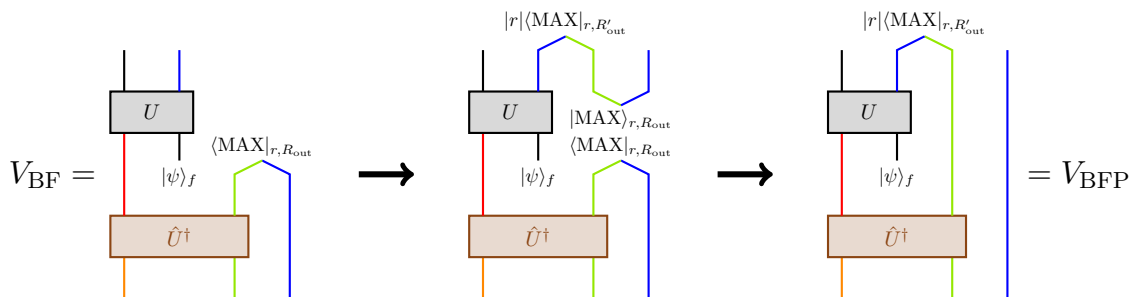[3]This result is essentially the same as the result of (3.12), but we repeat it here for clarity.

Figure 7. The transformations used in [20] to equate V_{BF} and V_{BFP} on the dynamically generated subspace of the effective description.
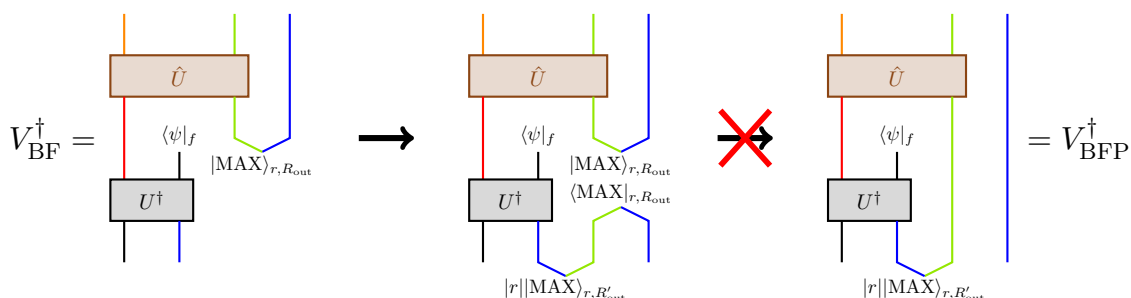

Figure 8. The transformations of figure 7 cannot relate the adjoints of the backwards-forwards map (left) and the BFP map (right) since fundamental dynamics do not guarantee that the Hawking pairs rR_{out} will be in the maximally entangled state.

We note that the second transformation in figure 7 — replacing the projector $|\mathrm{MAX}\rangle\langle\mathrm{MAX}|$ with the identity $\mathbb{1}$ — is only possible because we are restricting to the dynamically generated subspace. This restriction guarantees that the Hawking pairs rR_{out} will be returned to the maximally entangled state by the backwards effective dynamics $\hat{U}^{\dagger}$, on which the projector acts trivially.

We interpret the inverse transformations (going from right to left in figure 7) as the analog of the "straightening procedure" of [1] that relates PHEVA's dynamical holographic map to the fundamental dynamics. When interior interactions in the effective description were included by [14] without adjusting the map, this straightening procedure led to partially transposed unitaries in the fundamental dynamics that spoiled the unitarity of black hole evaporation. By including interior interactions in our maps directly, we have demonstrated that the isometric BFP map can be related to unitary backwards-forwards map in a way that does not lead to partially transposed unitaries. The unitarity of black hole evaporation is preserved.

However, it is not possible to relate the adjoints $V_{\mathrm{BF}}^{\dagger}$ and $V_{\mathrm{BFP}}^{\dagger}$ with similar transformations,

$$V_{\mathrm{BF}}^{\dagger}P_t \neq V_{\mathrm{BFP}}^{\dagger}P_t, \tag{3.26}$$

since (3.17) demonstrated that $V_{\mathrm{BFP}}^{\dagger}$ maps fundamental dynamically generated states outside of the dynamically generated subspace of the effective description. This is further demon-

strated by figure 8; the projector onto the maximally entangled state cannot be replaced with the identity since fundamental dynamics do not guarantee that the Hawking pairs rR_{out} will be in the maximally entangled state.

Instead, we found that it is $\hat{P}_t V_{\text{BFP}}^\dagger$ that acts isometrically on the dynamically generated subspace of the fundamental description. Therefore we expect

$$\hat{P}_t V_{\text{BF}}^\dagger P_t = \hat{P}_t V_{\text{BFP}}^\dagger P_t \tag{3.27}$$

to be true. Indeed, projecting (3.25) onto the dynamically generated subspace of the fundamental description and taking the overall adjoint gives exactly this relation between the adjoints of the two maps.

3.4 Page curve

The work of [1] goes further to derive the QES formula for the entropy of the radiation with their dynamical map, obtaining the Page curve. We show here that both the BF and BFP maps also reproduce the same QES formula and Page curve.

First, we note that the effective dynamics we presented in [20] don't involve a mechanism that allows information to escape the black hole. Because the horizon causally separates the interior and exterior of the black hole, we don't include any unitaries $\hat{U}_t$ in the effective description that connects ℓr modes with R_{out}. Thus each $\hat{U}_t$ only keeps information behind the horizon. A naive calculation of the radiation's entropy shows that it can only increase, giving the Hawking curve.

Holographic maps provide a proxy for these missing dynamics to calculate the correct entropy of the radiation. Applying either holographic map to the effective dynamics, we map to the fundamental description where the calculation of the radiation's entropy does give the Page curve. Calculating the fine-grained entropy $S_2(\hat{\Psi}_{R_{\text{out}}})$ of the radiation after application of the V_{BF} or V_{BFP} requires an average over U. We will perform this integration using the spin network method described in appendix A. Figure 9 shows the leading two contributions to the average over U of $e^{-S_2(\hat{\Psi}_{R_{\text{out}}})}$ using V_{BFP} for $n_0 = 4$ and $m_0 = 2$ up until $t = 2$. The result for V_{BF} is the same and can be found by straightening the bent radiation lines. Generalizing to arbitrary n_0, m_0, and t, these evaluate to

$$q^{-2t} \qquad \text{and} \qquad q^{-(n_0+t)} e^{-S_2(\hat{\Psi}_{R_{\text{in}}})} \tag{3.28}$$

respectively, where we've used the large q approximation. These terms will dominate at different times; taking a logarithm, we find the same Page curve as in [1],

$$S_2(\hat{\Psi}_{R_{\text{out}}}) = \min\left[2t \log q, \ (n_0 - t) \log q + S_2(\Psi_{R_{\text{in}}})\right]. \tag{3.29}$$

Assuming for the moment that $\Psi_{R_{\text{in}}}$ is pure, the first term in (3.29) dominates for $|R_{\text{out}}| < |B|$ and the second dominates for $|R_{\text{out}}| > |B|$. The dominance of the terms will switch when $|R_{\text{out}}| \sim |B|$, and we define this to be the Page time,

$$t_{\text{Page}} \sim \frac{n_0}{3}. \tag{3.30}$$

Using a mixed state for $\Psi_{R_{\text{in}}}$ will only push the Page time later. If we also consider a more general set of effective dynamics where $\mathcal{I}$ qudits fall in and $\mathcal{O}$ qudits are radiated at each
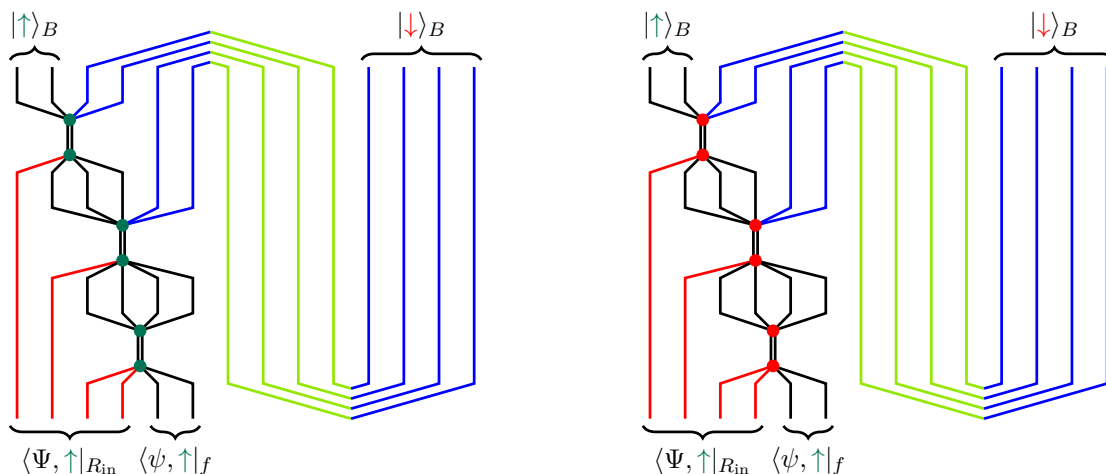
Figure 9. The two dominant contributions in the calculation of the fine-grained entropy of R_{out} using the BFP holographic map. Green dots indicate the insertion of a $\uparrow$ state; red dots indicated the insertion of a $\downarrow$ state.

time step, the Page time is given by[4]

$$t_{\text{Page}} \sim \frac{n_0 + S_2(\Psi_{R_{\text{in}}})}{2\mathcal{O} - \mathcal{I}}.$$

(3.31)

Choosing $\mathcal{I} = 1$ and $\mathcal{O} = 2$ gives the dynamics shown in figure 1 and reproduces (3.30) above. Had we instead chosen $\mathcal{I} = 0$ and $\mathcal{O} = 1$ (letting one qudit escape as radiation per time step and none fall in) we would've found the usual $t_{\text{Page}} \sim n_0/2$.

4 Action on generic states

Now that we have reviewed the successes of the backwards-forwards and BFP maps on the dynamically generated subspace, we turn to generic states of the effective Hilbert space. The first dynamical holographic map proposed by PHEVA was shown to satisfy two important criteria demonstrating its effectiveness on generic states: (1) it acts isometrically on average with exponentially small fluctuations, and (2) it provides a good state-dependent reconstruction of unitary operators in the effective description on BR_{out} in the fundamental description. We show here that the BFP map also satisfies these criteria, continuing to demonstrate that it is a proper generalization of PHEVA's work to include non-trivial interior interactions, while the BF map does not.

For the remainder of this work, we drop the assumption that black hole states are in the dynamically generated subspaces of either description. In particular, we will not assume r and R_{out} will return to a maximally entangled state after backwards evolution in the effective description. Similarly, we will not assume that a state in the fundamental description was obtained by fundamental dynamics applied to a particular fixed state on f. Any state on the full effective or fundamental Hilbert spaces can be used as an input to the backwards-forwards or BFP maps in what follows.

---

[4]In [1], the Page time was identified as the time when $|\ell| \cdot |r| \sim |B|$; at this time, the holographic map transitions from an approximate isometry to non-isometric. For general $\mathcal{I}$ and $\mathcal{O}$, this time is $t \sim (n_0 - m_0)/(2\mathcal{O})$. Since both Page times are linear in n_0, we take their difference to be insignificant.

4.1 Failure of the BF map

We begin by showing that the backwards-forwards map does not even satisfy the first criterion for acting on generic states: it does not act isometrically on average. Performing an average over U and $\hat{U}$ of $V_{\mathrm{BF}}^\dagger V_{\mathrm{BF}}$ acting on generic states, we find

$$\int d\hat{U} dU \, \langle\hat{\Psi}_j|_{\ell r R_{\mathrm{out}}} V_{\mathrm{BF}}^\dagger V_{\mathrm{BF}} |\hat{\Psi}_i\rangle_{\ell r R_{\mathrm{out}}} = \int d\hat{U} dU \quad\quad\quad = \int d\hat{U} \tag{4.1}$$

where we used the unitarity of U to remove it in the second equality without a need for averaging. Averaging over $\hat{U}$ using (A.2) gives a factor of $1/|r|$, and straightening the now connected radiation lines gives a second factor of $1/|r|$. Together, we find

$$\int d\hat{U} \langle\hat{\Psi}_j|_{\ell r R_{\mathrm{out}}} V_{\mathrm{BF}}^\dagger V_{\mathrm{BF}} |\hat{\Psi}_i\rangle_{\ell r R_{\mathrm{out}}} = \frac{1}{|r|^2}\delta_{ij}. \tag{4.2}$$

Therefore V_{BF} does not preserve the inner product of generic effective description states on average. Instead, it tends to decrease the inner product, indicating that the backwards-forwards map is annihilating too many generic states.

We can intuitively understand this failure of the backwards-forwards map as follows. Consider a state $|\Psi\rangle_{R_{\mathrm{in}}}$ that collapses unitarily into a black hole and begins evaporating by the effective dynamics. A naive measurement of the interior will perturb the state in a way that cannot be undone by the backwards evolution in the backwards-forwards map. Therefore, the state that comes out of the black hole after backwards evolution could be in some state $|\Psi'\rangle_{R_{\mathrm{in}}}$ not equal to the original. Instead, the original state $|\Psi\rangle$ would now be shared between the R_{in} and radiation r modes stored after the backwards evolution. If we were to follow the backwards-forwards map and post-select on rR_{out} at this point, the information now stored in the radiation would be lost. In this way, the backwards-forwards map annihilates too many generic states.

4.2 Deviation of BFP from isometry

Following the failure of the backwards-forwards map on generic states, we test the BFP map against the two criteria introduced in [1]. First, we check the action of $V_{\mathrm{BFP}}^\dagger V_{\mathrm{BFP}}$ on

generic states of the effective Hilbert space,

$$\int d\hat{U} dU \, \langle\hat{\Psi}_j|_{\ell r} V_{\text{BFP}}^\dagger V_{\text{BFP}} |\hat{\Psi}_i\rangle_{\ell r}$$

$$= \int d\hat{U} dU \, |r|^2 \quad \hspace{3cm} = \int d\hat{U} \, |r|^2 \frac{|B|}{|B| \cdot |R'_{\text{out}}|} \hspace{2cm}, \qquad (4.3)$$

where we've temporarily dropped the R_{out} factor since V_{BFP} acts trivially on the outgoing radiation. Since we aren't assuming that rR_{out} began in the maximally entangled state, we cannot simply get rid of post-selection by straightening lines as we did in (3.14). Instead, we integrate U over the Haar measure using equation (A.2); the result of this integration is shown in the second equality of (4.3). We can now straighten the connected radiation lines, contributing another factor of $1/|r|$; this allows the effective dynamics to reduce to the identity without any need for integration. Recalling that $|R'_{\text{out}}| = |R_{\text{out}}| = |r|$, we find

$$\int dU \, \langle\hat{\Psi}_j|_{\ell r} V_{\text{BFP}}^\dagger V_{\text{BFP}} |\hat{\Psi}_i\rangle_{\ell r} = \delta_{ij}. \qquad (4.4)$$

For Haar average fundamental dynamics U, the BFP map acts isometrically on generic states of the effective Hilbert space.

It is not enough to only act isometrically on average; PHEVA demonstrated that holographic maps for the black hole interior must have fluctuations around this average that are exponentially small in the black hole's entropy. In calculating the fluctuations, it will be useful to consider overlaps between non-orthogonal states; for this we define

$$|\hat{\Psi}_\alpha\rangle_{\ell r R_{\text{out}}} = \sum_i \alpha_i |\hat{\Psi}_i\rangle_{\ell r R_{\text{out}}}, \qquad (4.5)$$

such that $0 \le \langle\hat{\Psi}_\beta|\hat{\Psi}_\alpha\rangle \le 1$. As in [1], we quantify the fluctuations of V_{BFP} about isometry using

$$\int d\hat{U} dU \left| \langle\hat{\Psi}_\beta|_{\ell r R_{\text{out}}} V_{\text{BFP}}^\dagger V_{\text{BFP}} |\hat{\Psi}_\alpha\rangle_{\ell r R_{\text{out}}} - \langle\hat{\Psi}_\beta|\hat{\Psi}_\alpha\rangle_{\ell r R_{\text{out}}} \right|^2$$

$$= \int d\hat{U} dU \, \langle\hat{\Psi}_\beta|\langle\hat{\Psi}_\alpha| (V_{\text{BFP}}^\dagger V_{\text{BFP}})^{\otimes 2} |\hat{\Psi}_\alpha\rangle|\hat{\Psi}_\beta\rangle - \left| \langle\hat{\Psi}_\beta|\hat{\Psi}_\alpha\rangle \right|^2, \qquad (4.6)$$

where we have dropped the $\ell r R_{\text{out}}$ subscripts in the second line for ease of reading. Equation (A.3) is required to integrate over the two copies of $V_{\text{BFP}}^\dagger V_{\text{BFP}}$ in the first term. Integrating over U and straightening any bent radiation lines gives

$$= \int d\hat{U} \, \frac{|r|^2}{|B|^2 |r|^2 - 1} \left[|B|^2 |\langle \hat{\Psi}_\alpha | \hat{\Psi}_\beta \rangle_{\ell r R_{\text{out}}}|^2 + |B| \operatorname{tr} \left((\hat{\Psi}_\alpha)_{R_{\text{out}}} (\hat{\Psi}_\beta)_{R_{\text{out}}} \right) \right]$$

(4.7)

where $(\hat{\Psi}_\alpha)_{R_{\text{out}}} = \operatorname{tr}_{\ell r} |\hat{\Psi}_\alpha\rangle\langle\hat{\Psi}_\alpha|_{\ell r R_{\text{out}}}$. Integrating the last two terms of (4.7) over $\hat{U}$ using (A.3) and plugging the result in (4.6), we find the fluctuations of the BFP map around isometry:

$$\int d\hat{U} dU \left| \langle \hat{\Psi}_\beta |_{\ell r R_{\text{out}}} V_{\text{BFP}}^\dagger V_{\text{BFP}} | \hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} - \langle \hat{\Psi}_\beta | \hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} \right|^2$$

$$= \frac{(|B| \cdot |\ell| \cdot |r| - 1)(|r|^2 - 1)}{(|B|^2 |r|^2 - 1)(|\ell|^2 |r|^2 - 1)} \left[|\ell| \cdot |r| \operatorname{tr} \left((\hat{\Psi}_\alpha)_{R_{\text{out}}} (\hat{\Psi}_\beta)_{R_{\text{out}}} \right) - \left| \langle \hat{\Psi}_\beta | \hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} \right|^2 \right].$$

(4.8)

To determine the strength of these fluctuations, we find an upper bound of (4.8). We can maximize over the choice of $\hat{\Psi}_\alpha$ and $\hat{\Psi}_\beta$ by setting

$$\operatorname{tr} \left((\hat{\Psi}_\alpha)_{R_{\text{out}}} (\hat{\Psi}_\beta)_{R_{\text{out}}} \right) = 1, \qquad \left| \langle \hat{\Psi}_\beta | \hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} \right|^2 = 0.$$

(4.9)

Applying the inequality

$$\frac{1}{x-1} \leq \frac{2}{x}, \quad x \geq 2, \tag{4.10}$$

on the overall prefactor twice gives an upper bound on the fluctuations of $V_{\text{BFP}}^\dagger V_{\text{BFP}}$ as

$$\int d\hat{U} dU \left| \langle \hat{\Psi}_\beta |_{\ell r R_{\text{out}}} V_{\text{BFP}}^\dagger V_{\text{BFP}} |\hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} - \langle \hat{\Psi}_\beta |\hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} \right|^2 \leq \frac{4}{|B|}. \tag{4.11}$$

For average U and $\hat{U}$, the fluctuations of the BFP map around isometry are indeed exponentially small in the black hole's entropy, thanks to the factor of $|B|$ in the denominator. Therefore, the backwards-forwards map preserves the inner product of all generic states on the effective Hilbert space with exponentially small errors. Furthermore, this upper bound differs from the upper bound on fluctuations of PHEVA's dynamical model by only a factor of 2; for convenience, we reproduce their exact result and bound here,[5]

$$\int d\hat{U} dU \left| \langle \hat{\Psi}_\beta |_{\ell r R_{\text{out}}} V_{\text{PHEVA}}^\dagger V_{\text{PHEVA}} |\hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} - \langle \hat{\Psi}_\beta |\hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} \right|^2$$

$$= \frac{|r|-1}{|B|^2 |r|^2 - 1} \left[|B| \cdot |r| \operatorname{tr} \left((\hat{\Psi}_\alpha)_{R_{\text{out}}} (\hat{\Psi}_\beta)_{R_{\text{out}}} \right) - \left| \langle \hat{\Psi}_\beta |\hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} \right|^2 \right] \leq \frac{2}{|B|}. \tag{4.12}$$

Figure 10 compares the exact fluctuations for the BFP map in (4.8) with those for PHEVA's dynamical map in (4.12) as a function of time using

$$|B| = q^{n_0 - t}, \quad |r| = q^{2t}, \quad |\ell| = q^{m_0 + t}. \tag{4.13}$$

The parameters $q = 2$, $n_0 = 10$, $m_0 = 1$, and $\operatorname{tr}((\hat{\Psi}_\alpha)_{R_{\text{out}}} (\hat{\Psi}_\beta)_{R_{\text{out}}}) = 1$ were used for both plots. We see that fluctuations of both models behave similarly, getting larger at late times. With these choices of parameters, the fluctuations of the BFP map are slightly larger than PHEVA's dynamical map, but their agreement strengthens as the black hole continues to evaporate. We note that the BFP map does have smaller fluctuations than PHEVA's map for other choices of parameters. Agreement between the two strengthens as q is increased since both expressions have the same large q approximation,

$$\int d\hat{U} dU \left| \langle \hat{\Psi}_\beta |_{\ell r R_{\text{out}}} V^\dagger V |\hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} - \langle \hat{\Psi}_\beta |\hat{\Psi}_\alpha \rangle_{\ell r R_{\text{out}}} \right|^2$$

$$\approx q^{-n_0 + t} \operatorname{tr} \left((\hat{\Psi}_\alpha)_{R_{\text{out}}} (\hat{\Psi}_\beta)_{R_{\text{out}}} \right), \quad \text{large } q. \tag{4.14}$$

We've now seen that the BFP map satisfies the first criterion for a holographic map acting on generic states: it preserves the inner product of generic states on average with exponentially suppressed fluctuations. Furthermore, it does so in a way comparable to the dynamical holographic maps proposed by PHEVA in [1]. We note that unlike previous results, this did require averaging $\hat{U}$ over the Haar measure. We do not expect each $\hat{U}_t$, which are determined by the dynamics of semiclassical gravity, to be strongly scrambling or complex enough to be well modeled by Haar random unitaries. Instead, we expect unitary k-designs could be used to achieve the above integrals without making the dynamics too scrambling or complex.

Before continuing to reconstruction using the BFP map, we comment on the relationship between the BFP map and the backwards-forwards map. As described in section 3.3,

---

[5]PHEVA found fluctuations in equation (2.7) of [1] for their generic holographic map defined in equation (2.6) of this work. The result quoted here applies to their dynamical holographic map defined here in (2.7).
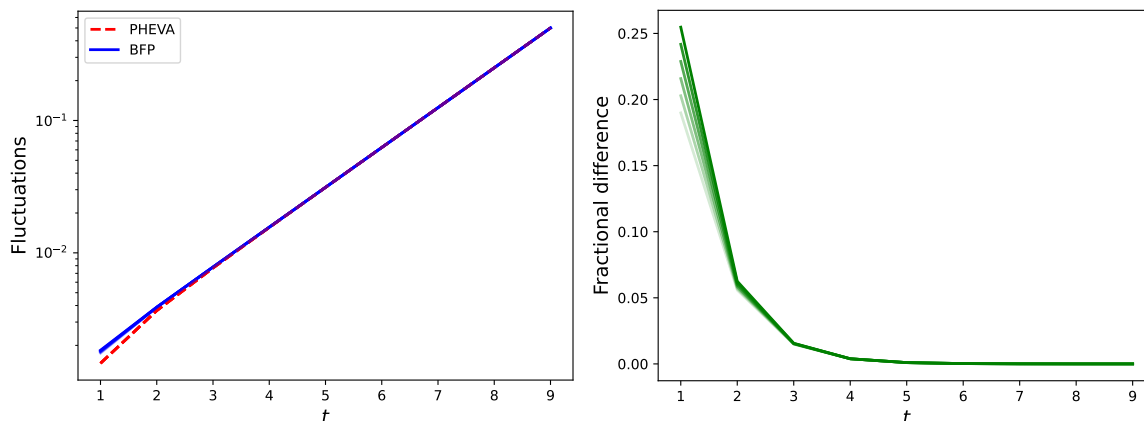
Figure 10. Plots comparing the fluctuations around isometry as a function of time for V_{BFP} and V_{PHEVA} using $q = 2$, $n_0 = 10$, and $m_0 = 1$. The left panel plots the exact expressions for the BFP map (4.8) in blue and PHEVA's dynamical map (4.12) in dashed red. The right panel shows the fractional difference between the two, (BFP − PHEVA)/PHEVA. $\text{tr}((\hat{\Psi}_\alpha)_{R_{\text{out}}}(\hat{\Psi}_\beta)_{R_{\text{out}}})$ was set to 1 in all plots; the color shade indicates the value of $|\langle\hat{\Psi}_\beta|\hat{\Psi}_\alpha\rangle_{\ell r R_{\text{out}}}|^2$, with 0 being the darkest, and 1 being the lightest.

the transformations relating the two allowed us to argue that the unitarity of black hole evaporation was preserved by the map, since they connect the BFP map to the unitarity of the backwards-forwards map on dynamically generated states. However, these transformations are no longer valid for generic states. We can no longer guarantee that radiation modes in the effective description will backwards evolve to the maximally entangled state, so we cannot replace the projector $|\text{MAX}\rangle\langle\text{MAX}|_{r,R_{\text{out}}}$ with the identity in the second step of figure 7. Even so, we've seen that the backwards-forwards map no longer acts isometrically on generic states, let alone unitarily. Therefore, we cannot claim unitarity of black hole evaporation for the full effective Hilbert space.

However, there is no issue here — generic states can only be reached by a measurement performed by an interior observer. Since an observer cannot force themselves to forget the result of their measurement, there is no reason to expect unitarity from their perspective. A second observer who hasn't learned the result of the measurement would see no violation of unitarity; all that's occurred from their perspective is a dynamical unitary entangling the first observer with the black hole interior. This keeps the interior state in the dynamically generated subspace, where both the transformations of figure 7 and unitarity still hold.

4.3 Reconstruction using BFP

We now check the second criterion for a holographic map to work properly on generic states: the BFP map must be able to reconstruct non-dynamical effective description unitary operators in the fundamental description. It was shown in [1] that PHEVA's holographic maps provided a state-dependent reconstruction somewhere on BR_{out} at all times. Furthermore, reconstruction on B and R_{out} separately implied the appropriate bounds for entanglement wedge reconstruction:

$$\text{Reconstruction on } B \implies S(\hat{\Psi}_{R_{\text{out}}}) \ll S(\hat{\Psi}_L) + \log|B| \tag{4.15}$$

$$\text{Reconstruction on } R_{\text{out}} \implies S(\hat{\Psi}_{R_{\text{out}}}) \gg S(\hat{\Psi}_L) + \log|B|. \tag{4.16}$$

The ability of their holographic maps to reconstruct measurements performed by an observer in the black hole interior then follows from their ability to reconstruct unitary operators entangling the observer and their measurement apparatus with the interior state.

Here, we demonstrate that the BFP map provides an equally good state-dependent reconstruction of interior unitaries W in the fundamental description. To do so, we follow the ideas laid out by theorem 3.7 of [11] and theorem 5.2 of [1]. In our current notation, these state that if the action of $W \in \mathcal{H}_{\text{eff}}$ on the effective Hilbert space decouples from a reference system L,

$$\left\| \text{tr}_{BR_{\text{out}}} \left(V_{\text{BFP}} W |\hat{\Psi}\rangle\langle\hat{\Psi}| W^\dagger V_{\text{BFP}}^\dagger \right) - \text{tr}_{BR_{\text{out}}} \left(V_{\text{BFP}} |\hat{\Psi}\rangle\langle\hat{\Psi}| V_{\text{BFP}}^\dagger \right) \right\|_1 \leq \epsilon_2, \qquad (4.17)$$

then there exists some unitary operator $\tilde{W} \in \mathcal{H}_{\text{fun}}$ acting on the complement of L in the fundamental Hilbert space such that

$$\left\| \tilde{W} V_{\text{BFP}} |\hat{\Psi}\rangle - V_{\text{BFP}} W |\hat{\Psi}\rangle \right\| \leq \epsilon_1 \qquad (4.18)$$

with $\epsilon_1 \leq \sqrt{\epsilon_2}$. In other words, if the reference system L cannot be used to discern whether or not W was applied inside the interior, then there exists some unitary acting somewhere on BR_{out} in the fundamental Hilbert space that serves as the reconstruction of W. We will consider this reconstruction to be "good" if ϵ_2 (and therefore ϵ_1) is exponentially small in the black hole's entropy.

We emphasize that this state-dependent reconstruction is only needed for non-dynamical W. Because the BFP map acts isometrically on dynamically generated states, reconstruction in the dynamical subspace is exact and state-independent. Therefore, if W were a dynamical unitary, it would be undone by the backwards evolution and there would exist a state-independent reconstruction in the fundamental description. However, these nice properties do not apply to non-dynamical unitaries since they cannot be undone by backwards dynamics. In this case, V_{BFP} acts only isometrically on average; therefore no state-independent reconstruction is guaranteed to exist. (See theorem 5.1 of [1] for more details.) In what follows, we will always assume W is a non-dynamical unitary.

We begin by checking the reconstruction of $W_{\ell r R_{\text{out}}}$ on BR_{out} through decoupling from L. This calculation is essentially a repeat of that done by PHEVA in [1] for their generic holographic map, here applied to the BFP map. As in [1], we first define a W-dependent state in the fundamental description for notational convenience:

$$\Psi_{LBR_{\text{out}}}(W) \equiv (V_{\text{BFP}} \otimes \mathbb{1}_L) W |\hat{\Psi}\rangle\langle\hat{\Psi}|_{L\ell r R_{\text{out}}} W^\dagger (V_{\text{BFP}}^\dagger \otimes \mathbb{1}_L) \qquad (4.19)$$

Introducing Haar integration over fundamental dynamics U and effective dynamics $\hat{U}$, the left hand side of the decoupling inequality (4.17) becomes

$$\int d\hat{U} dU \left\| \Psi_L(W_{\ell r R_{\text{out}}}) - \Psi_L(\mathbb{1}) \right\|_1 \leq \sqrt{|L| \int d\hat{U} dU \left\| \Psi_L(W_{\ell r R_{\text{out}}}) - \Psi_L(\mathbb{1}) \right\|_2^2}, \qquad (4.20)$$

where the right hand side of (4.20) is the upper bound of the L_1 norm given (as in [1]) by the L_2 norm and Jensen's inequality. Our task is then to compute the radicand of (4.20),

$$\int d\hat{U} dU \left\| \Psi_L(W_{\ell r R_{\text{out}}}) - \Psi_L(\mathbb{1}) \right\|_2^2$$
$$= \int d\hat{U} dU \ \text{tr} \left[\Psi_L(W_{\ell r R_{\text{out}}})^2 - 2\Psi_L(W_{\ell r R_{\text{out}}})\Psi_L(\mathbb{1}) + \Psi_L(\mathbb{1})^2 \right]. \qquad (4.21)$$

All three terms in the right hand side of (4.21) involve essentially the same integration, differing only by whether or not W is acting on $\ell r R_{\text{out}}$. We will demonstrate the calculation of the first term explicitly; the remaining two will follow naturally from the first. Represented as a circuit diagram, this term is

$$\int d\hat{U}dU \ \text{tr} \ \Psi_L(W_{\ell r R_{\text{out}}})^2 = |r|^4 \qquad \text{(4.22)}$$

We first integrate over the four factors of fundamental dynamics U using equation (A.3) from appendix A. This generates four terms,

$$\text{(4.23)}$$

Thanks to the cyclic property of the trace, $\hat{U}$ disappears from the first and second terms of (4.23) without the need for integration. W then also disappears in the first term to give $\operatorname{tr}\hat{\Psi}_L^2$, while the second reduces to $\operatorname{tr}(\operatorname{tr}_{R_{\text{out}}} W|\hat{\Psi}\rangle\langle\hat{\Psi}|W^\dagger)^2$. However, we are not so fortunate with the third and fourth terms, and we must integrate over the effective dynamics using (A.3) from appendix A. The resulting integrations also give contributions of $\operatorname{tr}\hat{\Psi}_L^2$ and $\operatorname{tr}(\operatorname{tr}_{R_{\text{out}}} W|\hat{\Psi}\rangle\langle\hat{\Psi}|W^\dagger)^2$, but with additional factors of $|\ell|$ and $|r|$; combining with the first two, we find

$$
\int d\hat{U} dU \, \operatorname{tr}\Psi_L(W_{\ell r R_{\text{out}}})^2
$$
$$
= \left[\frac{|B|^2|r|^2}{|B|^2|r|^2-1} - \frac{|B|\cdot|\ell|\cdot|r|(|r|^2-1)+|r|^2(|\ell|^2-1)}{(|B|^2|r|^2-1)(|\ell|^2|r|^2-1)}\right]\operatorname{tr}\hat{\Psi}_L^2
$$
$$
+ \left[\frac{|B|\cdot|r|^2}{|B|^2|r|^2-1} - \frac{|B|\cdot|r|^2(|\ell|^2-1)+|\ell|\cdot|r|(|r|^2-1)}{(|B|^2|r|^2-1)(|\ell|^2|r|^2-1)}\right]\operatorname{tr}\left(\operatorname{tr}_{R_{\text{out}}} W|\hat{\Psi}\rangle\langle\hat{\Psi}|W^\dagger\right)^2. \tag{4.24}
$$

This completes the integration over the first term in the right hand side of (4.21). The result for the second and third terms are similar — we only need to replace $\operatorname{tr}(\operatorname{tr}_{R_{\text{out}}} W|\hat{\Psi}\rangle\langle\hat{\Psi}|W^\dagger)^2$ with

$$
\operatorname{tr}(\operatorname{tr}_{R_{\text{out}}} W|\hat{\Psi}\rangle\langle\hat{\Psi}|W^\dagger)(\operatorname{tr}_{R_{\text{out}}}|\hat{\Psi}\rangle\langle\hat{\Psi}|) \qquad \text{and} \qquad \operatorname{tr}(\operatorname{tr}_{R_{\text{out}}}|\hat{\Psi}\rangle\langle\hat{\Psi}|)^2, \tag{4.25}
$$

respectively. The $\operatorname{tr}\hat{\Psi}_L^2$ term in (4.24) has no W dependence and is the same for all three terms in (4.21); it therefore cancels in the sum. The final result for the integration over the squared L_2 norm is then given by

$$
\int d\hat{U} dU \left|\left|\Psi_L(W_{\ell r R_{\text{out}}}) - \Psi_L(\mathbb{1})\right|\right|_2^2
$$
$$
= \left[\frac{|B|\cdot|r|^2}{|B|^2|r|^2-1} - \frac{|B|\cdot|r|^2(|\ell|^2-1)+|\ell|\cdot|r|(|r|^2-1)}{(|B|^2|r|^2-1)(|\ell|^2|r|^2-1)}\right]
$$
$$
\times \operatorname{tr}\left[\left(\operatorname{tr}_{R_{\text{out}}} W|\hat{\Psi}\rangle\langle\hat{\Psi}|W^\dagger\right)^2 - 2\left(\operatorname{tr}_{R_{\text{out}}} W|\hat{\Psi}\rangle\langle\hat{\Psi}|W^\dagger\right)\left(\operatorname{tr}_{R_{\text{out}}}|\hat{\Psi}\rangle\langle\hat{\Psi}|\right) + \left(\operatorname{tr}_{R_{\text{out}}}|\hat{\Psi}\rangle\langle\hat{\Psi}|\right)^2\right]. \tag{4.26}
$$

To check that the action of W decouples to exponential precision in the black hole's entropy, we upper bound this function. We first maximize over choices of the initial state $|\hat{\Psi}\rangle$ in the effective description; since $\operatorname{tr}(\rho)^2 \leq 1$ for any density matrix ρ, the state dependent portion of (4.26) takes a maximum value of 2. Applying (4.10) multiple times, we find

$$
\int d\hat{U} dU \left|\left|\Psi_L(W_{\ell r R_{\text{out}}}) - \Psi_L(\mathbb{1})\right|\right|_2^2 \leq \frac{4}{|B|}\left(1 + \frac{2}{|\ell|^2|r|^2} + \frac{2}{|B|\cdot|\ell|\cdot|r|^3}\right). \tag{4.27}
$$

Plugging into (4.20), we find the upper bound for the average decoupling from L,

$$
\int d\hat{U} dU \left|\left|\Psi_L(W_{\ell r R_{\text{out}}}) - \Psi_L(\mathbb{1})\right|\right|_1 \leq 2\sqrt{\frac{|L|}{|B|}\left(1 + \frac{2}{|\ell|^2|r|^2} + \frac{2}{|B|\cdot|\ell|\cdot|r|^3}\right)}. \tag{4.28}
$$

Indeed, the factor of $1/\sqrt{|B|}$ ensures that the information about W decouples from L to a precision exponentially small in the black hole's entropy. According to the theorems discussed

earlier, this implies that the BFP map does indeed provide a good reconstruction of interior operators W in the effective description on BR_{out} in the fundamental description. In fact, the first term in the upper bound (4.28) is exactly the same upper bound found by PHEVA in [1] for their generic holographic map (2.6). The second two terms arise from the inclusion of non-trivial effective dynamics. They are small compared to the first term (particularly at late times) and we consider them to be "corrections" to the PHEVA result.

We also compare our reconstruction with the BFP map against reconstruction using PHEVA's dynamical map. The exact result for their squared L_2 norm is given by,

$$\int dU \left|\left|\Psi_L^{\text{PHEVA}}(W_{\ell r R_{\text{out}}}) - \Psi_L^{\text{PHEVA}}(\mathbb{1})\right|\right|_2^2$$
$$= \frac{|r|^2}{|B|^2|r|^2 - 1}\left[|B|D_{R_{\text{out}}} - \frac{|B|}{|r|}D_{r R_{\text{out}}} - \frac{1}{|r|}D_{\ell R_{\text{out}}}\right], \tag{4.29}$$

where

$$D_A \equiv \text{tr}\left[\left(\text{tr}_A W|\hat{\Psi}\rangle\langle\hat{\Psi}|W^\dagger\right)^2 - 2\left(\text{tr}_A W|\hat{\Psi}\rangle\langle\hat{\Psi}|W^\dagger\right)\left(\text{tr}_A |\hat{\Psi}\rangle\langle\hat{\Psi}|\right) + \left(\text{tr}_A |\hat{\Psi}\rangle\langle\hat{\Psi}|\right)^2\right]. \tag{4.30}$$

We note that our result for the BFP map (4.26) also depends on $D_{R_{\text{out}}}$, but $D_{r R_{\text{out}}}$ and $D_{\ell R_{\text{out}}}$ contribute to the decoupling for PHEVA's dynamical map only. The upper bound on the L_1 norm for PHEVA's dynamical map is then

$$\int dU \left|\left|\Psi_L^{\text{PHEVA}}(W_{\ell r R_{\text{out}}}) - \Psi_L^{\text{PHEVA}}(\mathbb{1})\right|\right|_1 \leq 2\sqrt{\frac{|L|}{|B|}\left(1 - \frac{1}{|r|} - \frac{1}{|B|\cdot|r|}\right)}, \tag{4.31}$$

which also has the same leading term as the generic map accompanied by small corrections from the fundamental dynamics alone. We note that the corrections appearing in (4.28) involve additional factors of $|\ell|$ and $|r|$ thanks to the effective dynamics.

Figure 11 compares the exact expressions for the squared L_2 decoupling using the BFP map (4.26) and PHEVA's dynamical map (4.29) as a function of time during evaporation. Common parameters were set to $q = 2$, $n_0 = 10$, $m_0 = 1$, and $D_{R_{\text{out}}} = 2$. The full range of (4.29) from the maximum at $D_{r R_{\text{out}}} = D_{\ell R_{\text{out}}} = 0$ to the minimum at $D_{r R_{\text{out}}} = D_{\ell R_{\text{out}}} = 2$ is shaded in both plots. We see that the BFP map always falls within the range of possible values for PHEVA's dynamical map, performing better for certain choices of $|\hat{\Psi}\rangle$. Stronger agreement can be seen at later times.

Reconstruction on B. Now that we have verified that the BFP map can be used to reconstruct W anywhere on BR_{out}, we check that the reconstructions on B and R_{out} separately satisfy the requirements of entanglement wedge reconstruction. Many of the details involved in these calculations are similar to those above for the reconstruction on BR_{out}, so we will suppress them and focus on the relevant differences.

Generalizing the above theorems, our ability to use V_{BFP} to reconstruct $W_{\ell r}$ on B depends on the decoupling of $W_{\ell r}$ from both L and R_{out},

$$\int d\hat{U}dU \left\|\Psi_{LR_{\text{out}}}(W_{\ell r}) - \Psi_{LR_{\text{out}}}(\mathbb{1})\right\|_1$$
$$\leq \sqrt{|L|\cdot|R_{\text{out}}|\int d\hat{U}dU \left\|\Psi_{LR_{\text{out}}}(W_{\ell r}) - \Psi_{LR_{\text{out}}}(\mathbb{1})\right\|_2^2}. \tag{4.32}$$
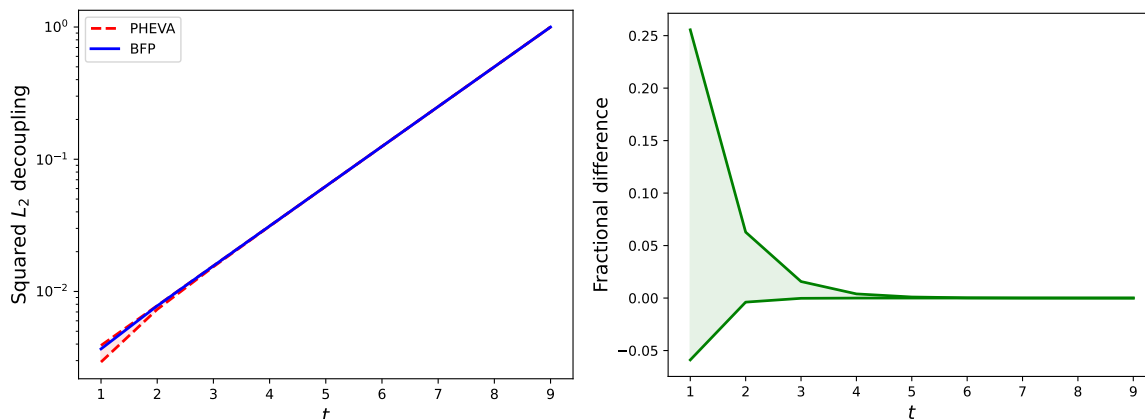
Figure 11. Plots comparing the squared L_2 decoupling from the reference system L for the BFP map (4.26) in blue and PHEVA's dynamical map (4.29) in dashed red. The right panel shows the fractional difference between the two, (BFP − PHEVA)/PHEVA. $D_{R_{\text{out}}}$ was set to 2 in all plots, with the full range of possible values with this choice for PHEVA's dynamical map shaded.

We again bound the L_1 decoupling using the L_2 decoupling, here given by

$$
\int d\hat{U} dU \left|\left| \Psi_{LR_{\text{out}}}(W_{\ell r}) - \Psi_{LR_{\text{out}}}(\mathbb{1}) \right|\right|_2^2
$$
$$
= \left[\frac{|B| \cdot |r|^2}{|B|^2 |r|^2 - 1} - \frac{|B| \cdot |r|^2 (|\ell|^2 - 1) + |\ell| \cdot |r|(|r|^2 - 1)}{(|B|^2 |r|^2 - 1)(|\ell|^2 |r|^2 - 1)} \right]
$$
$$
\times \left[2 - 2 \operatorname{tr} \left(W_{\ell r} |\hat{\Psi}\rangle\langle\hat{\Psi}| W_{\ell r}^\dagger |\hat{\Psi}\rangle\langle\hat{\Psi}| \right) \right]. \tag{4.33}
$$

To find an upper bound on the L_2 decoupling, we maximize over choice of $|\hat{\Psi}\rangle$ by setting

$$
\operatorname{tr} \left(W_{\ell r} |\hat{\Psi}\rangle\langle\hat{\Psi}| W_{\ell r}^\dagger |\hat{\Psi}\rangle\langle\hat{\Psi}| \right) = 0. \tag{4.34}
$$

Applying (4.10) and substituting into (4.32), we find

$$
\int dU d\hat{U} \left|\left| \Psi_{LR_{\text{out}}}(W_{\ell r}) - \Psi_{LR_{\text{out}}}(\mathbb{1}) \right|\right|_1 \le 2 \sqrt{ \frac{|L| \cdot |R_{\text{out}}|}{|B|} \left(1 + \frac{2}{|\ell|^2 |r|^2} + \frac{2}{|B| \cdot |\ell| \cdot |r|^3} \right) }. \tag{4.35}
$$

Again, the first term is exactly what was found by PHEVA in [1] for their upper bound for reconstruction on B. The second and third terms are the same corrections as we found in (4.28) for the reconstruction on BR_{out}. Requiring that the upper bound on the L_1 decoupling be exponentially small in the black hole's entropy implies

$$
|L| \cdot |R_{\text{out}}| \left(1 + \frac{2}{|\ell|^2 |r|^2} + \frac{2}{|B| \cdot |\ell| \cdot |r|^3} \right) \ll |B| \qquad \Longrightarrow \qquad |L| \cdot |R_{\text{out}}| \ll |B|, \tag{4.36}
$$

which is true for times earlier than the Page time. Therefore interior operators $W_{\ell r}$ can be reconstructed on B to exponential precision before the Page time, reminiscent of entanglement wedge reconstruction. Indeed, taking a logarithm of (4.36) gives the condition for entanglement wedge reconstruction on B,

$$
S_2(\hat{\Psi}_{R_{\text{out}}}) \ll \log |B| + S_2(\hat{\Psi}_L), \tag{4.37}
$$

just as was found for PHEVA's generic holographic map in [1].

Reconstruction on R_{out}. Similarly, reconstruction of $W_{\ell r}$ on R_{out} depends on the decoupling of both L and B,

$$\int d\hat{U} dU \left\| \Psi_{LB}(W_{\ell r}) - \Psi_{LB}(\mathbb{1}) \right\|_1 \leq \sqrt{|L| \cdot |B| \int d\hat{U} dU \left\| \Psi_{LB}(W_{\ell r}) - \Psi_{LB}(\mathbb{1}) \right\|_2^2}, \quad (4.38)$$

which is upper bounded by the L_2 norm,

$$\int d\hat{U} dU \left\| \Psi_{LB}(W_{\ell r}) - \Psi_{LB}(\mathbb{1}) \right\|_2^2$$

$$= \left[\frac{|B|^2 |r|^2}{|B|^2 |r|^2 - 1} - \frac{|r|^2(|\ell|^2 - 1) + |B| \cdot |\ell| \cdot |r|(|r|^2 - 1)}{(|B|^2 |r|^2 - 1)(|\ell|^2 |r|^2 - 1)} \right]$$

$$\times \left[2e^{-S_2(\hat{\Psi}_{R_{\text{out}}})} - 2 \operatorname{tr} \left(W_{\ell r} \hat{\Psi}_{L\ell r} W_{\ell r}^\dagger \hat{\Psi}_{L\ell r} \right) \right]. \quad (4.39)$$

This L_2 norm is maximized when

$$\operatorname{tr} \left(W_{\ell r} \hat{\Psi}_{L\ell r} W_{\ell r}^\dagger \hat{\Psi}_{L\ell r} \right) = 0, \quad (4.40)$$

which bounds the L_1 decoupling of LB as

$$\int dU d\hat{U} \left\| \Psi_{LB}(W_{\ell r}) - \Psi_{LB}(\mathbb{1}) \right\|_1 \leq 2 \sqrt{\frac{|L| \cdot |B|}{e^{S_2(\hat{\Psi}_{R_{\text{out}}})}} \left(1 + \frac{2}{|B| \cdot |\ell| \cdot |r|^3} + \frac{2}{|B|^2 |\ell|^2 |r|^2} \right)}. \quad (4.41)$$

As before, the first term agrees exactly with [1]. The second term appears as a correction in both the reconstruction on BR_{out} from (4.28) and the reconstruction on B from (4.35), but the third is a new correction further suppressed by an additional factor of $1/|B|^2$. Requiring the right hand side of (4.41) to be small implies

$$|L| \cdot |B| \left(1 + \frac{2}{|B| \cdot |\ell| \cdot |r|^3} + \frac{2}{|B|^2 |\ell|^2 |r|^2} \right) \ll e^{S_2(\hat{\Psi}_{R_{\text{out}}})} \quad \implies \quad |L| \cdot |B| \ll e^{S_2(\hat{\Psi}_{R_{\text{out}}})}. \quad (4.42)$$

Again, taking a log gives the condition for entanglement wedge reconstruction on R_{out},

$$S_2(\hat{\Psi}_{R_{\text{out}}}) \gg \log |B| + S_2(\hat{\Psi}_L). \quad (4.43)$$

Thus $W_{\ell r}$ can be reconstructed on R_{out} with exponential precision at times much later than the Page time.

5 Conclusion

In this work, we have further investigated properties of the backwards-forwards and BFP maps for encoding black hole interiors with non-trivial interior dynamics first proposed in [20]. Restricting to the subspace of dynamically generated states, V_{BF} acts unitarily, V_{BFP} acts isometrically, and both reproduce the same QES formula and Page curve as found by PHEVA in [1]. On generic states of the effective Hilbert space, we have shown that only the BFP map satisfies all of the same criteria as PHEVA's dynamical map to be a "good" holographic map: it acts isometrically on average (with exponentially small fluctuations) and provides a exponentially precise state-dependent reconstruction on BR_{out} that implies entanglement wedge reconstruction.

Furthermore, the BFP map acts as an exact isometry on dynamically generated states. Thus the set of dynamically generated states forms a unique subspace of the effective Hilbert space. This has two consequences. First, reconstruction of dynamical unitaries is exact and state-independent on this subspace. Second, dynamically generated states must be subexponentially complex according to the measure concentration results of [1]. If we take the reference state to be the state of R_{in} as it crosses the horizon, this is ensured by the limited lifetime of a black hole — only a number of gates polynomial in n_0 can be applied before the black hole evaporates. This choice of reference state is supported by the independence of the BFP map from external interactions (as demonstrated in [20]) since any external unitary of exponential depth will drop out of the BFP map.

It is important to note that many of the calculations in this work depend on averaging the effective dynamics $\hat{U}$ over the Haar measure. We find it unlikely that the interactions present in semiclassical gravity will be strongly scrambling or complex enough to be truly modeled by Haar random unitaries. Instead, we might opt to model the effective dynamics with a unitary k-design that can reproduce the above integrals without being too strongly scrambling or complex. The BFP map would then involve unitary k-designs for the effective dynamics and pseudorandom unitaries for the fundamental dynamics (following the suggestions of [14]). We further assume that the combination of these is random enough for the measure concentration results of [1] to apply. So long as they do, we may continue claiming that the BFP map acts isometrically on subexponentially complex generic states of the effective description, just as PHEVA's dynamical map does.

Future work is also needed to better understand how to generalize the toy models of black holes considered here to more realistic scenarios. For example, the treatment of Hawking modes entering and exiting the model relied on a model of modes coming down from high energy to motivate the "freezing out" of radiation modes as they are blue-shifted towards some ultraviolet scale during the backwards evolution in the effective description. It would be interesting to place this within a more realistic model of effective field theory with a cutoff interacting within, and backreacting upon, semiclassical gravity. We also did not need to assume very much about the unitaries $\hat{U}$ used to model interior interactions in the effective description beyond that they are well-modeled by a unitary k-design, but one might hope that a more realistic model incorporating more of bulk locality would constrain these further; some progress could potentially be made using well understood models such as tensor networks and JT gravity. There is still much work to be done in this area.

A Integration over the Haar measure

Here we review some techniques for integrating over the Haar measure. We will only cover those formulas that are needed for the computations performed in this work; for a more general review, see for example [24].

Haar measure integration of a quantum circuit allows us to understand the behavior of the circuit for "typical" unitaries. The Haar measure dU is normalized such that

$$\int dU = 1. \tag{A.1}$$

There are generic formulas for integrals over arbitrary combinations of unitaries, but we only need two specific integrations for this work. The first involves a unitary U and its
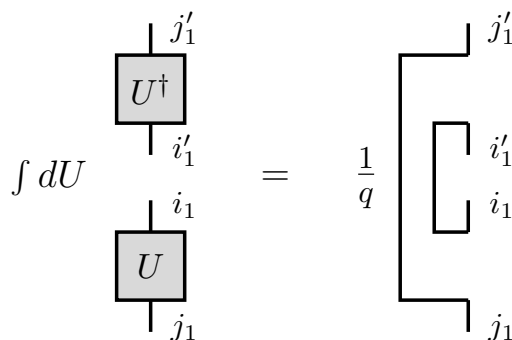
Figure 12. A pictorial representation of equation (A.2) for the Haar integration over U involving a single copy of U and its Hermitian conjugate.

Hermitian conjugate,

$$\int dU\, U^{\dagger}_{j'_1 i'_1} U_{i_1 j_1} = \frac{1}{q} \delta_{i_1 i'_1} \delta_{j_1 j'_1}, \tag{A.2}$$

where q is the dimension of each index. A pictorial representation of (A.2) is shown in figure 12. The legs i_1, i'_1, j_1, j'_1 can be connected in any way by inserting delta functions and summing over contracted indices. For example, if we connect legs i_1 and i'_1 with $\delta_{i_1 i'_1}$, the completed loop gives a contribution of q and the integral reduces to the identity $\mathbb{1}_{j'_1 j_1}$.

The second integral involves two copies of a unitary and its Hermitian conjugate,

$$\int dU\, U^{\dagger}_{j'_1 i'_1} U^{\dagger}_{j'_2 i'_2} U_{i_1 j_1} U_{i_2 j_2} = \frac{1}{q^2 - 1}\Big[\delta_{i_1 i'_1}\delta_{i_2 i'_2}\delta_{j_1 j'_1}\delta_{j_2 j'_2} + \delta_{i_1 i'_2}\delta_{i_2 i'_1}\delta_{j_1 j'_2}\delta_{j_2 j'_1}\Big]$$
$$- \frac{1}{q(q^2 - 1)}\Big[\delta_{i_1 i'_1}\delta_{i_2 i'_2}\delta_{j_1 j'_2}\delta_{j_2 j'_1} + \delta_{i_1 i'_2}\delta_{i_2 i'_1}\delta_{j_1 j'_1}\delta_{j_2 j'_2}\Big]. \tag{A.3}$$

A pictorial representation of (A.3) is shown in figure 13. We note that connecting i_1, i'_1 via $\delta_{i_1 i'_1}$ and i_2, i'_2 via $\delta_{i_2 i'_2}$ again reduces the integral to the identity.

It is sometimes convenient to re-express (A.3) as a partition function over a spin system [1, 25]. Notice that all of the delta functions in (A.3) connect i to i' (we will refer to these as "internal" indices) and j to j' (referred to as "external" indices). There are only two ways to connect each set of i and j indices, leading to the four terms in (A.3). The first connects indices of the same number, such as $\delta_{i_1 i'_1}\delta_{i_2 i'_2}$. These are the vertically contracted indices in figure 13. We will denote such contractions using a spin in the $\uparrow$ state; $|\uparrow\rangle$ for the internal i indices and $\langle\uparrow|$ for the external j indices. The second class of delta functions connects indices of different number, such as $\delta_{i_1 i'_2}\delta_{i_2 i'_1}$, shown by swapped contractions in figure 13. These contractions will be denoted by a spin in the $\downarrow$ state; again $|\downarrow\rangle$ for internal i indices and $\langle\downarrow|$ for external j indices.

We use the delta functions to define inner products between the "spin" basis $|\uparrow, \downarrow\rangle$ and the "index" basis $|i_1 i'_1 i_2 i'_2\rangle$ as

$$\langle i_1 i'_1 i_2 i'_2 | \uparrow\rangle = \frac{1}{q}\delta_{i_1 i'_1}\delta_{i_2 i'_2}, \quad \langle i_1 i'_1 i_2 i'_2 | \downarrow\rangle = \frac{1}{q}\delta_{i_1 i'_2}\delta_{i_2 i'_1}. \tag{A.4}$$

We can then rewrite (A.3) using the spin basis as

$$\int dU\, U^{\dagger} \otimes U^{\dagger} \otimes U \otimes U = \frac{q^2}{q^2 - 1}\Big(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|\Big) - \frac{q}{q^2 - 1}\Big(|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|\Big). \tag{A.5}$$

– 31 –

Figure 13. A pictorial representation of equation (A.3) for the Haar integration over U involving two copies of U and its Hermitian conjugate.

Taking the expectation value of (A.5) with states in the index basis reproduces (A.3). In the spirit of (A.5), we associate two spins with the unitary U we are integrating over. The first $|\uparrow, \downarrow\rangle$ indicates how to contract the internal indices i of U and its Hermitian conjugate; the second $\langle\uparrow, \downarrow|$ indicates how to contract the outer indices j. The full integration is given by a sum over all states on these two spins, analogous to a partition function.

This technique is most effective when integrating over many different unitaries U. Contracting indices between different unitaries requires an inner product on the spin basis,

$$\langle\uparrow \mid \downarrow\rangle = \frac{1}{q} \tag{A.6}$$

which can be found by inserting an identity as a complete set of states in the index basis and using (A.4). Multiple integrations can then be performed by taking products of the right hand side of (A.5) for each unitary being averaged, with the "innermost" unitary being the leftmost product.

This permits a useful graphical interpretation for integrating over multiple unitaries in two copies of a quantum circuit and its hermitian conjugate. We may construct a spin network by replacing every unitary in the circuit with two spins: one connected to all outgoing indices, the other connected to all ingoing. Then the integration is found by summing over all configurations of these spins. The weights of each term in the sum are seen in (A.5) and are found as follows. A unitary with k indices and two matching spins gives a contribution of

$$= \frac{q^{2k}}{q^{2k} - 1} \tag{A.7}$$

while the same unitary with two mismatching spins gives

$$\langle\downarrow| \qquad |\uparrow\rangle \qquad = \qquad \langle\uparrow| \qquad |\downarrow\rangle \qquad = \qquad \frac{q^k}{q^{2k}-1} \tag{A.8}$$

These simplify in a large q approximation to 1 and $1/q$, respectively. Connecting legs give a contribution according to the inner product (A.6): a leg connecting two matching spins gives a contribution of 1, while a leg connecting two mismatching spins gives $1/q$.

Dangling legs can be given a single $\uparrow$ or $\downarrow$ spin to perform certain computations. As an example relevant to section 3.4, consider computing the average second Renyi entropy of subsystem A for a density matrix ρ_{AB} after processing by a Haar random unitary U. We can take the logarithm to perform the average:

$$\int dU \ \mathrm{tr} \left(\mathrm{tr}_B \, U \rho_{AB} U^\dagger \right)^2 = \int dU \qquad \tag{A.9}$$

Instead of the circuit diagram on the right hand side of (A.9), we can use a spin state representation to integrate over the two copies of U and its Hermitian conjugate:

$$\int dU \ \mathrm{tr} \left(\mathrm{tr}_B \, U \rho_{AB} U^\dagger \right)^2 = \sum_{s_1,s_2=\uparrow,\downarrow} \tag{A.10}$$

The B subsystem is traced over by a vertical contraction in (A.9), so we fix the upper right leg in the $\uparrow$ state. The A subsystem is left take the product $(\rho_A)^2$ and is then traced over, represented by a swapped contraction in (A.9), so we fix the A spin in the $\downarrow$ state. Since the state ρ connects the internal legs of U and $U^\dagger$, we've inserted it into the bottom of the spin network in the $\uparrow$ state as $\langle\rho,\uparrow|$. Here, the contractions give

$$\langle\rho,\uparrow|\uparrow\rangle = \mathrm{tr}\,\rho, \qquad \langle\rho,\downarrow|\downarrow\rangle = \mathrm{tr}\,\rho^2 = -\ln S_2(\rho). \tag{A.11}$$

The full integration is then performed by summing over the two spins s_1 and s_2 for the unitary U and using the rules described above.

Acknowledgments

References

[1] C. Akers et al., *The black hole interior from non-isometric codes and complexity*, [arXiv:2207.06536](https://arxiv.org/abs/2207.06536) [INSPIRE].

[2] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [hep-th/0603001] [INSPIRE].

[3] T. Faulkner, A. Lewkowycz and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **11** (2013) 074 [arXiv:1307.2892] [INSPIRE].

[4] N. Engelhardt and A.C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073 [arXiv:1408.3203] [INSPIRE].

[5] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, *JHEP* **09** (2020) 002 [arXiv:1905.08255] [INSPIRE].

[6] A. Almheiri, N. Engelhardt, D. Marolf and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *JHEP* **12** (2019) 063 [arXiv:1905.08762] [INSPIRE].

[7] D.N. Page, *Information in black hole radiation*, *Phys. Rev. Lett.* **71** (1993) 3743 [hep-th/9306083] [INSPIRE].

[8] A.R. Brown, H. Gharibyan, G. Penington and L. Susskind, *The Python's Lunch: geometric obstructions to decoding Hawking radiation*, *JHEP* **08** (2020) 121 [arXiv:1912.00228] [INSPIRE].

[9] N. Engelhardt, G. Penington and A. Shahbazi-Moghaddam, *A world without pythons would be so simple*, *Class. Quant. Grav.* **38** (2021) 234001 [arXiv:2102.07774] [INSPIRE].

[10] N. Engelhardt, G. Penington and A. Shahbazi-Moghaddam, *Finding pythons in unexpected places*, *Class. Quant. Grav.* **39** (2022) 094002 [arXiv:2105.09316] [INSPIRE].

[11] C. Akers and G. Penington, *Quantum minimal surfaces from quantum error correction*, *SciPost Phys.* **12** (2022) 157 [arXiv:2109.14618] [INSPIRE].

[12] D. Harlow and P. Hayden, *Quantum Computation vs. Firewalls*, *JHEP* **06** (2013) 085 [arXiv:1301.4504] [INSPIRE].

[13] J. Chandra and T. Hartman, *Toward random tensor networks and holographic codes in CFT*, *JHEP* **05** (2023) 109 [arXiv:2302.02446] [INSPIRE].

[14] I.H. Kim and J. Preskill, *Complementarity and the unitarity of the black hole S-matrix*, *JHEP* **02** (2023) 233 [arXiv:2212.00194] [INSPIRE].

[15] G.T. Horowitz and J.M. Maldacena, *The black hole final state*, *JHEP* **02** (2004) 008 [hep-th/0310281] [INSPIRE].

[16] S. Lloyd and J. Preskill, *Unitarity of black hole evaporation in final-state projection models*, *JHEP* **08** (2014) 126 [`arXiv:1308.4209`] [INSPIRE].

[17] S. Aaronson, *Quantum computing, postselection, and probabilistic polynomial-time*, *Proc. R. Soc. A* **461** (2005) 3473.

[18] D. Deutsch and R. Penrose, *Quantum theory, the Church-Turing principle and the universal quantum computer*, *Proc. Roy. Soc. Lond. A* **400** (1997) 97.

[19] L. Susskind, *Horizons Protect Church-Turing*, `arXiv:2003.01807` [INSPIRE].

[20] O. DeWolfe and K. Higginbotham, *Non-isometric codes for the black hole interior from fundamental and effective dynamics*, *JHEP* **09** (2023) 068 [`arXiv:2304.12345`] [INSPIRE].

[21] Y. Kiem, H.L. Verlinde and E.P. Verlinde, *Black hole horizons and complementarity*, *Phys. Rev. D* **52** (1995) 7053 [`hep-th/9502074`] [INSPIRE].

[22] G. 't Hooft, *On the Quantum Structure of a Black Hole*, *Nucl. Phys. B* **256** (1985) 727 [INSPIRE].

[23] N. Bao, A. Bouland and S.P. Jordan, *Grover search and the no-signaling principle*, *Phys. Rev. Lett.* **117** (2016) 120501 [`arXiv:1511.00657`] [INSPIRE].

[24] A.A. Mele, *Introduction to Haar Measure Tools in Quantum Information: A Beginner's Tutorial*, *Quantum* **8** (2024) 1340 [`arXiv:2307.08956`] [INSPIRE].

[25] A. Nahum, S. Vijay and J. Haah, *Operator Spreading in Random Unitary Circuits*, *Phys. Rev. X* **8** (2018) 021014 [`arXiv:1705.08975`] [INSPIRE].