Wormholes and factorization in exact effective theory

Published for SISSA by Springer

Received: *August 6, 2024*
Revised: *December 18, 2024*
Accepted: *March 3, 2025*
Published: *May 5, 2025*

Wormholes and factorization in exact effective theory

Sergio Hernández-Cuenca

Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

E-mail: sergiohc@mit.edu

Abstract: We study the general framework of effective theories obtained via exact path integration of a complete theory over some sector of its degrees of freedom. Theories constructed this way contain multi-integrals which couple fields arbitrarily far apart, and in certain settings even on path-disconnected components of the space. These are not just entanglement, but genuine non-local interactions that we dub quantum wormholes.

Any state the path integral of such an effective theory prepares is shown to be a partial trace of a state of the complete theory over the integrated-out sector. The resulting reduced density operator is generally mixed due to bra-ket wormholes. An infinite family of ensembles of pure states of the complete theory giving the same effective state is identified. These allow one to equivalently interpret any effective state as being prepared by an ensemble of theories. When computing entropic quantities, bra-ket wormholes give rise to replica wormholes. This causes replica path integrals for the effective theory to not factorize even when the underlying manifold does, as expected from mixing. In contrast, effective theories obtained by derivative expansions have no quantum wormholes and prepare pure states. There exist operators in the algebra of effective theories which can distinguish mixed from pure states, implying a breakdown of non-exact effective theories for sufficiently complex observables.

This framework unifies and provides new insights into much of the phenomena observed in quantum gravity, including the interplay between wormholes and unitarity, the breakdown of bulk effective theory, the factorization puzzle, state ensembles, theory ensembles, quantum error correction, and baby universes. Some interesting lessons are drawn accounting also for characteristic aspects of gravity concerning IR/UV mixing and Kaluza-Klein reductions.

Keywords: AdS-CFT Correspondence, Effective Field Theories, Gauge-Gravity Correspondence, Nonperturbative Effects

ArXiv ePrint: 2404.10035

https://doi.org/10.1007/JHEP05(2025)024

Contents

1 Introduction

A longstanding problem in quantum gravity, particularly in the context of holography, concerns whether or not certain topologies should be included in a gravitational path integral formulation. If one treats the gravitational path integral as indiscriminately allowing for all possible manifold topologies consistent with boundary conditions, a puzzle immediately arises: à-priori independent boundary theories set up on disjoint spaces seem to get coupled by wormholes in the bulk. In particular, the boundary partition function should factorize across connected components, whereas the bulk geometries seem to break factorization.

JHEP05(2025)024

Such a phenomenon would clash with the AdS/CFT correspondence as originally formulated and generally understood [1–4]. The puzzle becomes particularly sharp given the existence of wormhole geometries which are classical solutions, and thus saddle points expected to dominate in a controlled semiclassical regime [5–13].[1] One may raise concerns about the subtle stability properties of some such solutions, or speculate that perhaps other saddles or non-perturbative contributions could cancel them out. However, lower-dimensional realizations of gravity continue to reaffirm that gravitational wormholes, if allowed to contribute, genuinely introduce statistical correlations between disjoint boundary regions. The various realizations in 3-dimensional gravity show how the effect of saddle wormholes cannot be reproduced by any single boundary CFT, but only by averages over CFT ensembles or similar randomized constructs [14–22]. In 2-dimensional Jackiw-Teitelboim (JT) theories of gravity [23–25], calculations can be taken beyond the semiclassical regime, leading to even more robust results. A full path integral over the moduli space of wormhole geometries of arbitrary topology can be performed in JT and precisely matched by the topological expansion of double-scale matrix integrals [26–28]. From the perspective of random matrix theory, JT gravity theories genuinely describe the spectral statistics of an ensemble of theories with a random matrix Hamiltonian, where factorization is drastically broken.

Faced with this factorization problem, one may be tempted to just throw away wormholes and somehow restrict the gravitational path integral to not include them. However, there are independent compelling reasons that they should in fact be included. A prominent one is the recent computations involving evaporating black holes [29–34], where wormholes are needed to resolve a sharp version of the black hole information paradox [35–40].[2] In this particular context, the relevant geometries are replica wormholes whose inclusion in the gravitational path integral is crucial to recover a unitary Page curve for the entropy between a black hole and its radiation. Another setting where wormholes are needed is to reproduce the universal late-time behavior of the spectral form factor in chaotic quantum systems [6, 26, 41–47]. Given these successes, giving up wormhole contributions to the gravitational path integral altogether would seem to be too naïve a route.

This paper is at its core motivated by the factorization problem in quantum gravity, and in particular by the quest to understand how the AdS/CFT correspondence could possibly be reconciled with the ensembles that wormholes in effective gravity give rise to. A heuristic viewpoint that resonates across the community is the idea that simple models of gravity, by ignoring microscopic details about their parent, UV-complete theory, should only be able to provide statistical answers. Such statistics would capture the ambiguity in which complete theory gave rise to the effective theory at hand, and reproduce results consistent with the ensemble of all possible such UV completions. While this intuition seems consistent with the observed phenomena, a precise mechanism for how non-factorization and ensembles came to be in the first place is still lacking. The central goal of this paper is to present a very general framework where not only do ensembles and non-factorization naturally arise, but also many other puzzling aspects of quantum gravity can be thoroughly addressed. In particular, as will be seen, features often understood as exclusively quantum gravitational turn out to also arise in field theory, including bra-ket wormholes, replica wormholes, and the breakdown of effective theory with complexity.

[1] This discussion concerns spacetime wormholes, and not spatial wormholes such as Einstein-Rosen bridges.
[2] The mechanism by which information dynamically escapes the black hole remains largely unknown.

1.1 Overview

Our discussion starts by supposing we have a theory which provides a complete description of some system in terms of a path integral. We then consider the construction of a general effective theory for some sector of interest as obtained by path-integrating over any other degrees of freedom of the theory. Note that this broad notion of effective theory is completely independent of renormalization and does not rely on any assumption of energy scale separation. Importantly, even when the starting point is a local theory, exact functional integration generally leads to an effective theory whose action involves multi-integrals over the underlying space. These non-localities cause the states the path integral of the exact effective theory prepares to generally be mixed due to correlations with the sector that was integrated out.

The non-local correlations in such a theory may be long-range, thus disallowing for a local derivative expansion and possible truncation thereof. However, even when parametrically suppressed, we show that a derivative expansion would lead to results inconsistent with the original theory. This becomes manifest in calculations where no matter how short-range, non-localities end up straddling across disjoint spaces. For obvious reasons, we refer to such non-local couplings across connected components of a space as quantum wormholes.

These quantum wormholes are seen to be instrumental in reproducing the fact that states prepared by an exact effective theory are mixed, in which case they take the form of bra-ket wormholes. When it comes to replica calculations they give rise to replica wormholes, which are required for consistency of the framework to reproduce valid results for states with mixing. At a more operational level, quantum wormholes also provide a mechanism for the breakdown of any effective theory where approximations have been made. In particular, no such approximate theory can possibly be able to prepare states capturing the expectation values of sufficiently complex observables in the truly mixed state of the exact effective theory.

We emphasize that there is nothing radical about much of these phenomena, which are in fact well understood in condensed matter theory as applied to open systems. In that context, the Schwinger-Keldysh formalism [48–52] is used to prepare a density operator, integrate out the environment, and obtain a reduced density operator for the system of interest. Then the evolution of this system is given by a Feynman-Vernon-like influence functional, which describes the non-equilibrium dynamics with the variables intrinsic to the system (cf. an effective theory) [53–55].[3] In such a setting, it is rather natural to expect information loss dissipatively to the bath (cf. non-unitarity of the evolution), and for the intrinsic predictions of the system of interest to only be statistical (cf. ensembles averages over the environment) [63]. The situation in effective theories is entirely analogous and, in our case, simplified by the fact that we will consider state preparation without Lorentzian evolution. As a result, for us the influence functional is nothing but the Euclidean effective path integral, and what our quantum wormholes capture is precisely the mixing of reduced density operators.[4]

[3]An influence functional is an operator describing Lorentzian time evolution of a reduced density operator, and can be obtained from a path integral by integrating out the environment degrees of freedom. In an operator formulation, making the usual Born-Markov approximations, such evolution of an open quantum system is given by a master equation in Lindblad form [56–59]. Similar techniques are also used in cosmology [60, 61], and have been applied to holographic systems [52]. For a recent discussion of the Lindbladian formalism in the context of exact renormalization, see [62].

[4]I thank Mukund Rangamani for illuminating discussions on these analogies and useful references.

JHEP05(2025)024

Our exposition will be field-theoretic and thus, by holography, allow us to address questions in quantum gravity from different complementary points of view. From a boundary perspective, the existence of quantum wormholes naturalizes non-factorization and strong non-localities as a general feature of effective theories. Taking the bulk as holographically defined, in accord with the mantra that the bulk geometrizes boundary quantum correlations, effective gravity wormholes could then be understood as an emergent gravitational description of some of the quantum wormholes of a dual effective theory.[5] In such a setting, the rules for the gravitational path integral should be dictated by the boundary theory, and because there are quantum wormholes in the latter, so there should be geometric wormholes in the former. Interestingly though, as will be seen, not always does effective field theory lead to quantum wormholes, suggesting that perhaps not always should geometric wormholes be allowed.

From a bulk perspective, were one to start from a non-perturbative theory of quantum gravity, it would be desirable to have an intrinsic description of geometric wormholes. Understanding precisely how an effective gravity theory in terms of a gravitational path integral over manifolds with dynamical metrics descends from an exact, UV-complete theory of quantum gravity remains a formidable challenge [65, 66]. Nonetheless, quantum wormholes in this case also demystify non-factorization and the emergence of ensembles, and additionally grant alternative plausible interpretations for gravitational wormholes.

A tantalizing possibility is that geometric wormholes are the quantum wormholes that arise upon integrating out whatever quantum gravitational degrees of freedom there may be more fundamental than metrics and gravitons.[6] This seems plausible given some of the distinguishing features quantum gravity exhibits which are usually extraneous to field theory, most noteworthy being diffeomorphism invariance and the phenomenon of IR/UV mixing. Together, if anything, these strongly suggest that effective gravity may exacerbate the strong non-localities that already arise in effective field theory. Another compelling alternative is to put gravitational wormholes on the same footing as quantum wormholes of quantum fields. Namely, for a field theory defined on gravitating backgrounds of different topology, integrating out gravitational wormholes gives rise to non-localities and quantum wormholes of the same flavor as those obtained by integrating out quantum fields. Hence it would seem conceivable that geometric wormholes may be equivalently describable as quantum wormholes of a field theory on a background of fixed topology. Without making any strong claims about either of these speculations, we believe the heuristics are sufficient to motivate further explorations.

1.2 Summary

Before providing a detailed outline of the contents of this paper, here we summarize some of the main takeaways and where a discussion thereof is presented. The reader is welcome to use this together with the outline that follows as a guide for a targeted reading of this paper that suits their interests.

Takeaways:

1. An effective theory obtained by exact functional integration from a local interacting theory generally has an action with arbitrarily many multi-integrals over the underlying space.

[5] See e.g. [64] for a related discussion of Euclidean wormholes in a holographic setting.

[6] A bottom-up realization of this idea via the renormalization of string effects recently appeared in [67].

JHEP05(2025)024

Such a theory provides an exact description of the system of interest, regardless of what environment degrees of freedom were integrated out. In certain cases one can simplify the multi-integrals down to a single integral by performing local derivative expansions. However, even if all terms in the resulting infinite series are kept or even resumed, any derivative expansion of this form already destroys quantum wormholes and gives a strictly different, approximate theory. Upon a possible truncation of higher derivative terms, one simply obtains an even coarser, truncated theory. In order, we refer to these different kinds of effective theories as exact, approximate, and truncated. The effective theories one typically studies in quantum field theory are of the latter type. See section 2.

2. The path integral of an exact effective theory prepares mixed states, even when one considers the full system. At the level of state preparation, this is because integrating out a sector to get an effective theory precisely corresponds to tracing out a density operator over that same sector. From the perspective of the exact effective theory, this is a consequence of bra-ket wormholes corresponding to the multi-integrals over the space on which the state is prepared. These mixed states capture exactly any observable intrinsic to the effective theory, as reduced density operators do. Approximate and truncated effective theories have no quantum wormholes and thus prepare pure states. No such pure state can possibly reproduce the expectation values of all observables on an exact effective state, indicating a breakdown of such descriptions. See section 3, particularly sections 3.1–3.3.

3. In quantum theory, mixed states are classical ensembles of pure states. The partial trace that constructs the reduced density operators of the exact effective theory via a path integral singles out a preferred ensemble of pure states for the system in terms of states of the integrated-out environment. This ensemble defines an orthogonal set of states in the complete theory, but is over-complete and gives rise to null states in the effective theory. The connection of this ensemble to the complete theory allows for an equivalent rewriting of the mixed states of the effective theory as being prepared by an ensemble of theories. The probability density functional that specifies the measure over the ensemble of states fixes the measure over the ensemble of theories. See section 3, particularly sections 3.4 and 3.5.

4. The amount of mixing of exact effective states can be quantified by entropic quantities obtained from replica calculations. It is still useful to construct a replica manifold, but a direct evaluation of the full path integral of the exact effective theory on this manifold turns out to be naïve. The correct prescription requires restricting the action multi-integrals to run over replicas separately to prepare independent states. Nonetheless, bra-ket wormholes end up connecting consecutive replicas and giving rise to replica wormholes, in consistency with mixing. See figure 1 for an illustration. Multi-swap operators can be used to make the discussion of replica symmetry and its breaking more transparent. Under different assumptions of replica symmetry, replica wormhole calculations are seen to be equivalent to powers of partition functions ensemble-averaged over theories. See section 4.

5. There is no canonical association of operator algebras to subspaces in non-local effective theory. The algebra additively generated by local operators within a subspace is generally a strict subalgebra of the commutant of the additively generated algebra of the complementary

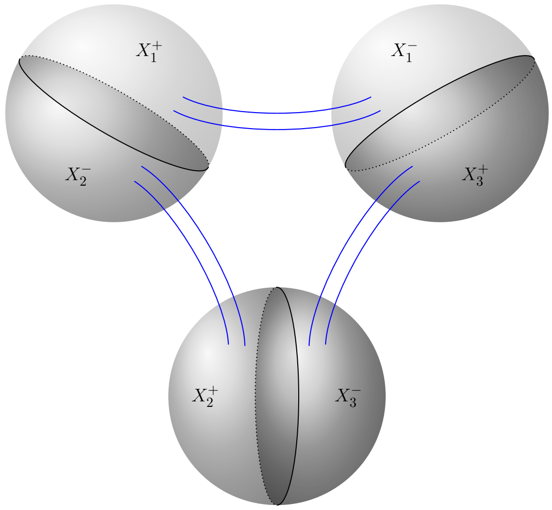

Figure 1. Representation of a 3-replica calculation. Three independent copies of a density operator are prepared by an exact effective theory path integral on $X_k^+ \sqcup X_k^-$ for $k \in \{1, 2, 3\}$, where boundary conditions on ∂X_k^+ and ∂X_k^- are left unspecified. The 'bra' and 'ket' of the k^{th} replica are respectively prepared on X_k^+ and X_k^-, which are coupled by bra-ket wormholes despite being disjoint spaces. Under a trace, boundaries get cyclically identified according to $\partial X_k^+ \sim \partial X_{k+1}^-$ with $k + 3 \sim k$. As a result, one obtains a nontrivial 3-replica wormhole. See section 4 for more details.

subspace. This is a violation of Haag-like duality caused by non-local operators which are only contained in the latter. In an exact effective theory, this violation seems to occur globally due to quantum wormholes. See section 5, particularly sections 5.1 and 5.2.

6. A violation of algebraic duality implies an incompleteness in the description of a state that can be attained by local observables, which cannot distinguish states differing only by expectation values for non-local operators. Relative entropy can be used as an operational measure of distinguishability between such states. When applied to a state prepared by an exact effective theory with respect to one prepared by an approximate or truncated effective theory, the relative entropy is formally infinite. This means that there exist observables which can perfectly distinguish such states, implying a breakdown of non-exact effective theories. For a truncated effective theory, this breakdown can be associated to energy scales that invalidate the neglect of higher derivatives. For an approximate effective theory, what naturally breaks down is the description of complex operators. See section 5, particularly section 5.3.

7. Effective field theory reproduces many features observed in holographic quantum gravity. An exact effective theory is only possibly consistent with the unitary evolution of the parent complete theory if replica wormholes are included in entropy calculations, which parallels the situation encountered in calculations of the Page curve for evaporating black holes. When it comes to partition functions, however, exact effective field theory factorizes. If geometric wormholes are the gravitational avatar of quantum wormholes, this suggests that wormholes should not be included in such calculations. Interestingly, under assumptions of replica symmetry, certain replica calculations precisely yield ensemble averages over partition

JHEP05(2025)024

functions. These allow one to reinterpret bulk calculations of multi-boundary partition functions with wormholes included as replica calculations consistent with a single complete theory on the boundary. Effective theories also exhibit the structure of operator algebra quantum error correction and ensembles of baby universes. See section 6.1.

8. The lessons of effective field theory are also applicable to quantum gravity outside the framework of holography. Diffeomorphism invariance and IR/UV mixing suggest that non-localities in effective gravity may be even stronger than in field theory. However, a gravity theory obtained by Kaluza-Klein reduction is more akin to a truncated rather than an exact effective theory. Performing a path integral after such a truncation has catastrophically non-local consequences of the same form as the wormholes one obtains from integrating α-parameters. See section 6.2.

Outline:

In section 2, we begin by introducing some basic notation and notions of locality that will be employed throughout. Section 2.2 defines exact effective theories as generally obtained by exact path integration over some sector of a given theory, and addresses how such theories contain non-local interactions which can lead to quantum wormholes. Section 2.3 clarifies where such a non-local effective theory stands relative to standard effective field theory treatments where through derivative expansions and truncations one ends up with a local theory. Section 2.4 touches for the first time on factorization at the level of partition functions, explaining how quantum wormholes generally do not arise in this case. Section 2.5 presents some examples of theories where exact path integration can be performed and illustrates the kind of non-localities that are important for this paper.

In section 3, we delve into the formalism of state preparation using the path integral, in particular as it applies to the construction of functional density operators. Section 3.1 does so first using the path integral of a local, complete theory. Section 3.2 then consider the preparation of states using the path integral of an exact effective theory, and shows how the effective state precisely corresponds to the reduced density operator of the complete state upon partially tracing over the pertinent sector via path integration. Section 3.3 explains how the generally mixed state that results from this procedure is inequivalent to the states one would obtain were one to perform a derivative expansion of the non-local theory. The inequivalence is due to quantum wormholes: these are the sole sources of their mixing, and any local derivative expansion would destroy them, thereby eliminating any information about the existence of a parent, complete theory. To ask fundamentally fine-grained questions about the theory, this should not be done. Section 3.4 shows how the path integral singles out a family of ensembles of pure states of the complete theory, all of which realize the same mixed, effective state upon partial tracing. Section 3.5 then explains how these mixed states can equivalently be understood as being prepared by an ensemble of theories.

In section 4, we undertake the exploration of replica path integrals. Section 4.1 motivates entropies as measures of mixing for effective states, and introduces the replica trick to calculate von Neumann entropies. Section 4.2 briefly describes how replica calculations are implemented in a path integral formulation of density operators, and applies it to the simpler case of local, complete theories. Section 4.3 then proceeds to explain the more subtle case of

JHEP05(2025)024

performing replica path integrals for states prepared by an effective theory. The specific case in which the replica calculation takes place on a space consisting of multiple components is thoroughly explained, as it constitutes a key illustration of how quantum wormholes break factorization even in effective field theory. Section 4.4 shows how the von Neumann entropy of the effective states can be written in terms of the classical Shannon entropy of an ensemble of pure states of the parent theory. Section 4.5 concludes with alternative representations of the replica path integrals in terms of field projections, which give a more transparent account of how complete and effective theory replicas differ, and what the role of replica symmetry is.

In section 5, we approach the observations throughout the paper in terms of the operator algebras of the different theories at play. Section 5.1 considers the association of algebras of operators to subspaces of a theory, and introduces the key notion of duality that allows for a canonical assignment. Section 5.2 explores the algebraic structure of effective theories, finding and characterizing the ways in which they violate duality. Quantum wormholes are identified as responsible for a particularly interesting violation of duality on a global scale of the theory. Section 5.3 then employs relative entropy as a natural operational measure to quantify the contributions of quantum wormholes to effective states. This provides not only a way of distinguishing between states of an effective theory, but also to distinguish states prepared by different effective theories. We observe that approximate and truncated effective theories cannot possibly capture observables of arbitrary complexity, which provides a mechanism for the breakdown of local effective theory.

In section 6, we conclude by exploring ways in which the field-theoretic framework of exact effective theory can also be applied to the realm of quantum gravity.

Section 6.2 approaches this from a holographic viewpoint by applying the lessons of this paper to the boundary theory. This provides an interpretation of the gravitational path integral as the effective gravity realization of a boundary effective theory. The dual to the quantum wormholes in the latter are naturally identified with the geometric wormholes in the former. Section 6.1.1 illustrates this in the specific setting of black hole evaporation, where the relevant effective boundary theory is the one obtained by integrating out the black hole states. From a purely effective theory perspective, a truncation of quantum wormholes would provide a description inconsistent with unitarity of the parent theory. The bulk results parallel this: replica wormholes are needed for consistency with unitarity, and ignoring them gives information loss. Section 6.1.2 considers the possibility that certain gravitational wormholes in holography may be unrelated to quantum wormholes on the boundary and thus unphysical. Section 6.1.3 provides an example of an ambiguity in holography between factorizing and non-factorizing quantities which seemingly would be computed by the same gravitational path integral according to standard rules. Section 6.1.4 shows how the algebraic structure of effective theories reproduces all the features of operator algebra quantum error correction in holography. Section 6.1.5 concludes the holographic discussion by identifying the baby universe Hilbert space and the α-sectors in an effective theory.

Section 6.2 finally makes some remarks on how the lessons from the effective theory framework may extrapolate to an effective gravity theory as potentially obtained from within a UV-complete theory of quantum gravity. Section 6.2.1 argues that the phenomenon of IR/UV mixing may be a mechanism by which non-localities in effective gravity may be even more

JHEP05(2025)024

dramatic than in field theory. Within quantum gravity, the plausible expectation seems to be that geometric wormholes are the effective manifestation of quantum wormholes associated to gravitational degrees of freedom. Section 6.2.2 briefly reflects on the way manipulations in effective gravity generally involve drastic truncations of the theory rather than integrations over sectors. Section 6.2.3 concludes with an example of how a naïve mode truncation inside a path integral can lead to catastrophically non-local results.

2 Theory framework

Let I be the Euclidean action of a non-gravitational theory defined on a Riemannian manifold M, and use ϕ to collectively denote all quantum fields of the theory. The partition function of the theory with action I is defined by a Euclidean path integral over the phase space of continuous field configurations of ϕ, and given by

$$Z \equiv \int_M \mathcal{D}\phi \, e^{-I[\phi]}. \tag{2.1}$$

The subscript M on the path integral indicates that the quantum fields ϕ are maps on M. If this theory provides a complete framework for the description of the physics of the system of interest, we refer to it as a *complete theory*.

2.1 Non-local terminology

It will be useful throughout to characterize theories in terms of their local or non-local properties. A theory will be said to be *local* if $I[\phi]$ involves a single integral over M, and the highest derivative of any field ϕ is of finite order. Otherwise, the theory will be said to be *non-local*. It will be important in what follows to distinguish between different types of non-localities. If $I[\phi]$ involves a single integral over M, but includes derivatives of arbitrarily high order for some field, the theory will be qualified to be *weakly non-local*. In contrast, if $I[\phi]$ contains more than one integral over M, the theory will be referred to as *strongly non-local*.[7] An interesting possibility in a strongly non-local theory is that, if the path integral is evaluated on a space consisting of more than one connected component, the multi-integrals in the action may couple fields on disconnected spaces. Such strongly non-local couplings across distinct connected components of the space will be referred to as *quantum wormholes*. The terminology introduced here for different types of non-locality will be used ubiquitously and is for convenience summarized in table 1.

2.2 Exact quantum effective theory

Hereon we will reserve the symbols I and Z respectively to refer to the action and partition function of a complete theory. Generally, we will be interested in complete theories which are local, so it shall be assumed that the action $I[\phi]$ is local in what follows. If one is only interested in the physics of a specific sector of quantum fields of the theory, a useful procedure

[7]As will be explained in section 2.3, it is sometimes possible to perform a derivative expansion to reduce a strongly non-local theory to a weakly non-local one. Certain non-local contributions are always lost in this process, so in general strong non-locality implies weak non-locality, but the converse is strictly false.

JHEP05(2025)024

Terminology	Description
Weak non-locality	Infinite derivative couplings with fields at the same point
Strong non-locality	Multi-integral coupling with fields at different points
Quantum wormhole	Strong non-locality coupling fields on disjoint spaces

Table 1. Summary of terminology for key notions of non-locality.

is to construct an effective theory for those degrees of freedom. This section makes precise the very general framework of effective theory that is relevant to this paper.[8]

Consider a splitting of the quantum fields of a complete theory into two sectors by letting $\phi = (\varphi, \Phi)$. This splitting may be completely arbitrary. Some natural examples correspond to identifying a system of interest and an environment, separating fields into bosonic and fermionic species, or splitting degrees of freedom by energy scales. For instance, these examples are respectively realized when studying open quantum systems, exploring the bosonic sector of a supersymmetric theory, or doing Wilsonian renormalization. Our treatment is completely agnostic as to what distinguishes φ and Φ.

An effective theory for the φ fields can be obtained by integrating out the Φ fields first,[9]

$$e^{-\widetilde{I}[\varphi]} \equiv \int\limits_M \mathcal{D}\Phi \, e^{-I[\varphi,\Phi]}, \tag{2.2}$$

a path integral which defines the *quantum effective action* $\widetilde{I}$ for φ. In general, this process gives rise to an action $\widetilde{I}[\varphi]$ which involves infinite series of increasingly many multi-integrals of the φ fields over M.[10] In other words, an effective theory obtained this way is a strongly non-local theory. Some simple examples of this are provided in section 2.5. By linearity of the path integral though, functional integration is transitive, so further performing the integral over φ fields yields back the same partition function as in eq. (2.1). Namely, the partition function of the effective theory,

$$\widetilde{Z} \equiv \int\limits_M \mathcal{D}\varphi \, e^{-\widetilde{I}[\varphi]}, \tag{2.3}$$

matches $Z = \widetilde{Z}$ in the absence of insertions. Crucially though, Z and $\widetilde{Z}$ are to be understood as mathematically different objects: Z defines a functional integral over $\phi = (\varphi, \Phi)$ fields on M, whereas $\widetilde{Z}$ defines a functional integral only over φ fields on M. In the latter, the path integral over Φ was already performed in eq. (2.2) to obtain the functional form of $\widetilde{I}$. The theory that $\widetilde{Z}$ defines only has access to the dynamics of the φ fields.

[8]The term "effective" is a loaded one in physics, and typically triggers a mental association to Wilsonian renormalization. We emphasize here and in multiple other parts of this paper that our construction of effective theories is a lot more general and independent of the considerations of renormalization. In particular, we do not require making any distinction between energy scales or need to assume any form of decoupling.

[9]If the path integral at hand involves insertions of the Φ fields, the resulting quantum effective action will also depend on these nontrivially. From the viewpoint of state preparation (see section 3), the effective theory for the φ fields may be different depending on the state being prepared by the path integral over Φ fields.

[10]This is a general phenomenon which in the example of Wilsonian renormalization follows as an exact functional statement from Polchinski's flow equation [68–71].

JHEP05(2025)024

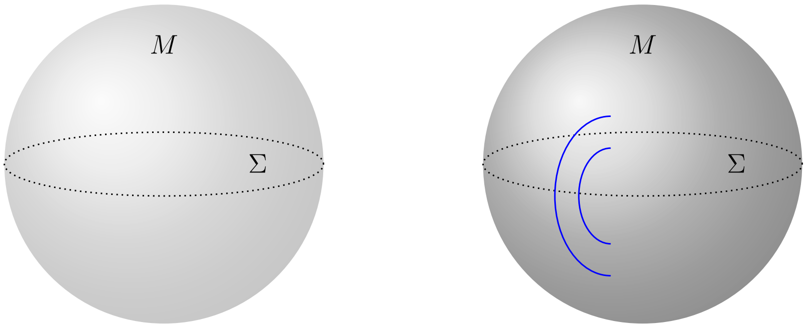

Figure 2. Representation of the path integral computation of partition functions on a spherical space M. The equatorial hypersurface Σ splits M into two dual halves. On the left, a complete theory with local action I is used to evaluate the partition function Z on M as defined in eq. (2.1). On the right, an exact QET with strongly non-local action $\widetilde{I}$ given by eq. (2.2) is used to evaluate the partition function $\widetilde{Z}$ defined in eq. (2.3). The latter is a strongly non-local theory involving arbitrarily many multi-local couplings. This is depicted in the case of a bilocal interaction with lines straddling across Σ between points on the top and bottom hemispheres at arbitrary separation. Numerically $\widetilde{Z} = Z$.

A theory with partition function $\widetilde{Z}$ given by eq. (2.3) and quantum effective action defined by the path integral in eq. (2.2) will be referred to as a *quantum effective theory (QET)*.[11] In particular, if eq. (2.2) is evaluated by exact functional integration without any approximations, the resulting theory will be said to be an *exact QET*. Tilded symbols like $\widetilde{I}$ and $\widetilde{Z}$ will always be used to refer to exact QET objects. The path integrals on M that give the partition function Z for a complete theory and $\widetilde{Z}$ for an exact QET are illustrated in figure 2.

This choice of terminology is easily motivated. The general quantum field theoretic concept of an effective action refers to the generating functional of one-particle irreducible correlation functions. Explicitly, introducing a classical current J sourcing a quantum field ϕ, one first defines the generating functional of connected correlation functions $W[J]$ by

$$e^{-W[J]} \equiv \int_M \mathcal{D}\phi\, e^{-I[\phi]-\int J\phi}. \tag{2.4}$$

The effective action is then the Legendre transform of $W[J]$, in terms of which one obtains an explicit form for the sourced partition functional of the theory with no path integrals left to be done. The way we define a QET through eq. (2.2) is entirely analogous, with a couple of important qualitative differences. In general the action in eq. (2.2) takes the form

$$I[\varphi, \Phi] = I_\varphi[\varphi] + I_\Phi[\Phi] + I_{\text{int}}[\varphi, \Phi], \tag{2.5}$$

with independent free actions I_φ and I_Φ for each sector, and an interaction term I_{int} that couples them. Comparing with eq. (2.4), we see that only Φ is integrated out in our case, with φ left untouched. Hence we have φ fields playing the role of currents, which source the Φ fields through the I_{int} term. This is why we call the action $\widetilde{I}[\varphi]$ effective.[12] Additionally, our

[11] In line with footnote 8, the decision to not refer to these objects by the initialism 'EFT' usually used for effective field theory has been deliberate. An EFT obtained by renormalization is just an example of a QET.

[12] Note however that this source-type term is generally nontrivial, so the resulting effective action no longer has any of the convexity properties of a Legendre transform.

JHEP05(2025)024

current-like objects φ are not classical, but actual quantum fields that we still want to treat dynamically. This is why we call the action $\widetilde{I}[\varphi]$ quantum. Finally, while in quantum field theory the effective action may often be obtained perturbatively or approximated to leading order by the classical action, note that its definition is given non-perturbatively by functional integration. This is why we refer to the resulting QET that $\widetilde{I}[\varphi]$ defines as exact.

2.3 Approximations and truncations

Although in quantum field theory one most often works with some effective theory in the above sense, rarely does one encounter a non-local action, let alone a strongly non-local one. This is because for most purposes an exact QET is an overkill, and truncations of it often suffice for describing the physics of interest to the desired degree of accuracy, but not always. Any such truncation gives rise to a coarser theory which may give good approximations for certain observables but is bound to fail completely for others. Here we briefly review this procedure in the context of renormalization, where a breakdown of truncated theories can easily be seen to be associated to energy scales. However, later in sections 3 and 5 we will see that truncated theories also fail at preparing states in concrete ways, which implies that such theories will also fail to reproduce the expectation values of sufficiently complex observables.

As explained below eq. (2.2), even in Wilsonian renormalization the process of integrating out some heavy Φ fields generates a strongly non-local action for light φ fields with infinitely many terms involving arbitrary multi-integrals over M. All of these non-localities account for the effects the now virtual Φ fields have in mediating interactions among φ fields (cf. figure 4). However, provided the Φ fields have a smallest positive mass m_Φ, such interactions will decay exponentially in distance scales of order m_Φ^{-1} and, additionally, higher integrals will exhibit a polynomial suppression in powers of m_Φ^{-1}. This decoupling phenomenon guarantees that the physics of φ fields behave approximately as local at energy scales $\Lambda \ll m_\Phi$.

At a mathematical level, this can be seen as follows. Under certain assumptions, small energies $\Lambda \ll m_\Phi$ allow one to expand all integrands in the strongly non-local QET action about any choice of reference point in M. Except for the reference point, all other points now appear outside φ fields, and their corresponding integrals over M can be explicitly performed. This way one obtains an action which contains infinitely many derivatives of the φ fields coming from the expansion about a reference point, but involves a single integral of that reference point over M. In other words, the resulting QET action is only weakly non-local. Now, this expansion is under control provided the energy scale of φ field derivatives is small compared to the mass scale of the Φ fields, i.e., so long as one probes the QET at energies $\Lambda \ll m_\Phi$, as noted above. Because higher derivative terms are suppressed, at small enough Λ energies one may recover an apparently local theory by truncating the expansion at some sufficiently high order. Hence, working at suitably small energies, one can often reduce a strongly non-local theory down to a weakly non-local one, and eventually to a local one by expanding and truncating. This procedure is exemplified in section 2.5. Of course, while the original theory is exact, the others are only progressively coarser approximations. In particular, these approximate theories will deviate significantly from the exact theory or even break down when probed at energies Λ comparable the mass scale m_Φ of the Φ fields that were integrated out.

JHEP05(2025)024

More abstractly, it may also be possible to apply these approximations to an exact QET obtained outside the framework of renormalization whenever one is only interested in some limit of the theory. The terminology that follows will be used in this broader sense. Let us first emphasize that, starting from an exact QET that is strongly non-local, any local derivative expansion already leads to an inequivalent theory. This is simply because such an expansion destroys quantum wormholes, whether or not the infinite series of derivatives is kept or even resummed after.[13] Indeed, if convergent, such a series can at most reproduce the action only within a connected component of the space, and not across connected components. The immediate goal of performing a derivative expansion is always to be able to evaluate all but one of the multi-integrals in the action. The resulting theory thus has infinitely many derivative terms but a single integral in its action, and is therefore only weakly non-local. A weakly non-local theory obtained this way from an exact QET without neglecting any terms will be referred to as an *approximate QET*. If one then truncates the infinite series to some finite derivative order, the result is an even coarser but now local theory. A local theory obtained this way from an approximate QET will be referred to as a *truncated QET*. The states and operator algebras of all these theories are related in interesting and operationally distinguishable ways, as analysed in sections 3 and 5.

We conclude this section by setting some expectations as to which basic properties these theories may or may not possess, which will help in anticipating some findings later on. Suppose our starting point is a complete theory which is local, unitary, and causal, and assume its fields φ and Φ interact nontrivially. As emphasized in section 1.1, an intuitive way to think about the resulting exact QET is as a theory describing an open system where the environment has been integrated out. This suggests that the sector the exact QET describes will evolve non-unitarily as only part of the complete, interacting theory, an expectation which is borne out due to quantum wormholes. Strong non-localities would also seem to make acausal physics plausible, since in Lorentzian signature spacelike separated regions would be non-locally coupled. However, microcausality for operators within disjoint domains of dependence will still hold since any non-local interactions are mediated by the integrated-out fields, which behave causally in the complete theory. It is crucial to bear in mind that the way an exact QET may be non-local and non-unitary will always be in consistency with how the complete theory genuinely behaves as a local and unitary theory. Put differently, any apparently pathological behavior of an exact QET will be reproducing precisely what is required to restore locality, unitarity, and causality of the complete theory.

A logical converse naturally suggests that any approximation or truncation of an exact QET may fail in recovering this behavior of the complete theory. For instance, by destroying quantum wormholes, one may expect that an approximate QET no longer behaves like an open system and thus describes a decoupled system which should obey unitarily. Such an evolution in which the system is artificially decoupled from the environment cannot possibly be consistent with the actual coupled, unitary evolution of the complete theory. In a similar

[13] Doing a resummation before performing any integrals would give back a strongly non-local theory, but one where there no longer are quantum wormholes. If the local expansion is not resummed before path integrating, the results cannot possibly reproduce those of the exact QET, even if one keeps all infinitely many terms. This is because convergence of a Taylor series to a given function everywhere requires that function to be analytic everywhere; however, path integrals generally run over functions which are not even differentiable.

vein, the way a truncated QET describes φ in terms of local dynamics cannot possibly be reconciled with the actual local dynamics of φ coupled to Φ in the complete theory. As we know though, nothing prevents us from performing these approximations and working with local effective theories which, provided they satisfy certain effective theory bounds, are perfectly unitary and causal. The lesson starting from an exact QET is that restoring these properties via approximations is incompatible with describing a complete theory with those same properties. Indeed, from the perspective of the complete theory, the predictions of an exact QET for the φ are perfectly valid, whereas those of a truncated QET would appear non-local and non-unitary.

2.4 Factorization

This section addresses the behavior of the partition function of our different theories, and anticipates how subtle the appearance of quantum wormholes happens to actually be. Indeed, as it turns out, quantum wormholes will only make their first explicit appearance when preparing states in section 3.

In general, the manifold M on which we define a theory may be disjoint. Let $\{M_k\}$ denote the connected components of M, i.e., the maximal connected subsets which partition it. By definition, these are disjoint, nonempty, and their union is the whole space. If the theory is local, the action $I[\phi]$ involves a single integral which by additivity will decompose into a sum of integrals over the different connected components. Schematically, the action reads

$$\int_M (\,\cdot\,) = \sum_k \int_{M_k} (\,\cdot\,). \tag{2.6}$$

If ϕ_k is the restriction of ϕ to M_k, then clearly the integral over M_k only depends on ϕ_k. By extending ϕ_k trivially to all M such that it continues to only be supported on M_k, one can write $\phi = \sum_k \phi_k$. This way, the path integral in eq. (2.1) can be easily seen to factorize as[14]

$$Z = \prod_k Z_k, \qquad Z_k \equiv \int_{M_k} \mathcal{D}\phi_k \, e^{-I[\phi_k]}. \tag{2.7}$$

In fact, if the theory is only weakly non-local, then eq. (2.6) still holds and the path integral similarly factorizes. Indeed, higher derivatives lead to non-localities within path-connected parts of the manifold, but never to interactions across distinct connected components. Strongly non-local theories have the potential to behave rather differently. In particular, a path integral in such a theory would fail to factorize if there are quantum wormholes across connected components of the underlying space. While one could certainly make up a strongly non-local theory that behaves this way, an exact QET turns out to not do this at the level of partition functions.

Consider again calculating the partition function, now for an exact QET. This is a specific type of strongly non-local theory for which the starting point would be a complete theory defined on M. If this complete theory is local, the discussion that led to eq. (2.7) still

[14]For $I[\phi_k]$, it makes no difference whether the action integral is taken to run over M or just M_k, so continuing to use the same notation for the action is unambiguous.

JHEP05(2025)024

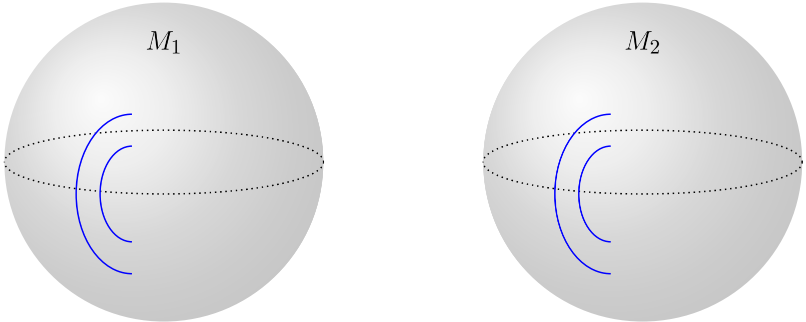

Figure 3. Representation of the partition function $\widetilde{Z}$ of an exact QET defined on a space M consisting of two distinct connected components, M_1 and M_2. The partition function factorizes as in eq. (2.8) since exact integration starting from a complete theory does not generate quantum wormholes. Nonetheless, strong non-localities do arise within each connected component of the space.

applies. Hence, one may start integrating out Φ fields from this factorized expression, with each path integral involving fields ϕ_k supported only on M_k. For each Z_k, there will clearly appear multi-integrals in the QET action $\widetilde{I}[\varphi_k]$, but these will only involve the M_k component of M. The non-localities are thus only generated within each connected component, but never across them. In other words, there are no quantum wormholes in the partition function of an exact QET, which still factorizes as

$$\widetilde{Z} = \prod_k \widetilde{Z}_k, \qquad \widetilde{Z}_k \equiv \int_{M_k} \mathcal{D}\varphi_k \, e^{-\widetilde{I}[\varphi_k]}, \tag{2.8}$$

where $\widetilde{I}$ is obtained as in eq. (2.2) by performing the path integral over Φ_k fields supported on just M_k. The structure of $\widetilde{Z}$ in eq. (2.8) is illustrated in figure 3. This situation shall be contrasted with the discussion that follows in sections 3 and 4, where it will be shown how other QET constructions do give rise to, and require, quantum wormholes.

Let us finish by pointing out that since fields are continuous maps on M, the constant function $\Phi(x) = \Phi_0$ is a perfectly valid configuration. However, if the path integral for Φ is restricted to just run over constant functions, what one gets is a dynamical cosmological constant taking the same value on every connected component of M. This is a very non-local object which would actually break factorization not only in eq. (2.8), but also in eq. (2.7). A global gauge symmetry would have similar consequences. This also illustrates an important point: in a theory where M itself is dynamically determined (cf. a gravitational path integral), non-factorization upon path integration would seem rather natural.

2.5 Examples

Here we cover some simple examples of the type of effective theories discussed in previous sections, which will serve as an illustration for the more general results of this paper.

Consider a Euclidean action of the form in eq. (2.5) involving two coupled scalars on a d-dimensional space M with free parts

$$I_\varphi[\varphi] = \frac{1}{2}\int_M d^dx \left[(\nabla\varphi(x))^2 + m_\varphi^2 \varphi(x)^2\right], \quad I_\Phi[\Phi] = \frac{1}{2}\int_M d^dx \left[(\nabla\Phi(x))^2 + m_\Phi^2 \Phi(x)^2\right], \tag{2.9}$$

JHEP05(2025)024

where the masses $m_\varphi, m_\Phi \geq 0$ are left arbitrary. For clarity, if M is a curved space we leave any metric determinants implicit in the integral measures and elsewhere.

The simplest possible interaction term one may consider is linear in Φ and of the form

$$I[\varphi, \Phi] = \lambda \int\limits_M d^d x \, f_\varphi(x) \, \Phi(x), \tag{2.10}$$

for some arbitrary function f_φ that depends only on φ.[15] For this interaction term the theory is well-defined so long as $m_\Phi > 0$. By virtue of being linear in Φ, eq. (2.10) admits the standard quantum field theory treatment of a source term. Integrating by parts with appropriate boundary conditions and completing the square to perform the functional Gaussian integral over Φ leads to

$$\int\limits_M \mathcal{D}\Phi \, e^{-I[\varphi,\Phi]} \propto \exp\left(-I_\varphi[\varphi] + \frac{\lambda^2}{2} \int\limits_{M^2} d^d x \, d^d y \, f_\varphi(x) \, G(x,y) \, f_\varphi(y) \right), \tag{2.11}$$

where G denotes the free propagator for Φ, i.e., the Green's function on M defined by

$$(-\nabla_x^2 + m_\Phi^2) \, G(x,y) = \delta(x-y). \tag{2.12}$$

From eq. (2.11) we can thus quote the final exact QET action as

$$\widetilde{I}[\varphi] = I_\varphi[\varphi] - \frac{\lambda^2}{2} \int\limits_{M^2} d^d x \, d^d y \, f_\varphi(x) \, G(x,y) \, f_\varphi(y), \tag{2.13}$$

which is our first example of a strongly non-local theory arising as an exact QET by integrating out Φ. In this case the strong non-locality takes the form of just a double integral or bilocal interaction involving two copies of the original space M.

The construction of an approximate QET from eq. (2.13) requires carefully checking that certain manipulations of the series expansion and integrals are allowed for a given M and corresponding G. Assuming this is the case for now and choosing x as the reference point, the desired local derivative expansion may be written

$$f_\varphi(y) = \sum_{k=0}^{\infty} \frac{f_\varphi^{(k)}(x)}{k!} (y-x)^k. \tag{2.14}$$

Using this in eq. (2.13) allows one to express the action as

$$\widetilde{I}[\varphi] = I_\varphi[\varphi] - \frac{\lambda^2}{2} \int\limits_M d^d x \sum_{k=0}^{\infty} \frac{g_k(x)}{m_\Phi^{2+k}} f_\varphi(x) f_\varphi^{(k)}(x), \tag{2.15}$$

where the g_k functions are the effective couplings induced by the multi-integrals,

$$g_k(x) \equiv \frac{m_\Phi^{2+k}}{k!} \int\limits_M d^d y \, G(x,y)(y-x)^k, \tag{2.16}$$

[15]This may be as simple as e.g. a quadratic term $f_\varphi = \frac{1}{2}\varphi^2$.

and the factors of m_Φ have been introduced to make g_k dimensionless. Now eq. (2.15) takes the form of the weakly non-local action of an approximate QET. In this action, higher orders in the derivative expansion are suppressed by powers of m_Φ. Hence this expansion is under control so long as the higher derivatives $f_\varphi^{(k)}$ are all small in units of m_Φ^{-1}. If this is the case, a truncation of the series to a finite order would be justified, thereby further reducing the theory to a truncated QET where locality is recovered.

For concreteness, suppose M is just a circle of radius R. Then the Green's function that obeys eq. (2.12) can be written

$$G(x,y) = \frac{1}{2m_\Phi}\frac{\cosh\left(m_\Phi(\pi R - |x-y|)\right)}{\sinh\left(m_\Phi \pi R\right)} = \frac{1}{2m_\Phi}\sum_{k\in\mathbb{Z}} e^{-m_\Phi|x-y+2\pi R\, k|}, \tag{2.17}$$

where on the circle $|x-y|<\pi R$. Notice that this propagator exhibits the familiar exponential decay with distance characteristic of massive fields.[16] Since G only depends on $|x-y|$, the couplings in eq. (2.16) are constants. These coupling constants vanish for odd k by symmetry. For even k they can be compactly written as finite series,

$$g_k = 1 - \frac{1}{\sinh\left(m_\Phi \pi R\right)}\sum_{l=1}^{k/2}\frac{(m_\Phi \pi R)^{2l-1}}{(2l-1)!}. \tag{2.18}$$

As $k\to\infty$ the finite series above becomes precisely the Taylor expansion of $\sinh\left(m_\Phi\pi R\right)$, implying that these dimensionless coupling constants go to zero at higher orders. In particular, at large k one finds the asymptotics

$$g_k \sim \frac{1}{k!}\frac{(m_\Phi\pi R)^k}{\sinh(m_\Phi\pi R)}. \tag{2.19}$$

Having introduced a new scale into the problem, we see by comparing eq. (2.19) to eq. (2.13) that the dimensionful parameter that is now important for the derivative expansion is R. Namely, for the expansion to be under control it must be the case that higher derivatives $f_\varphi^{(k)}$ are small in units of R. Since R is the length scale of the system, it becomes manifest that this local expansion only makes sense if the physics is indeed dominated by local interactions.

The example above only led to a bilocal coupling in the exact QET. Consider a slightly less trivial interaction term by making Φ quadratic [72],

$$I[\varphi,\Phi] = \frac{\lambda}{2}\int_M d^dx\, f_\varphi(x)\,\Phi(x)^2. \tag{2.20}$$

Integrating by parts, the resulting path integral over Φ gives rise to a functional determinant,

$$\int_M \mathcal{D}\Phi\, e^{-I[\varphi,\Phi]} \propto e^{-I_\varphi[\varphi]}\det\Big(-\nabla^2 + m_\Phi^2 + \lambda f_\varphi\Big)^{-1/2}, \tag{2.21}$$

which up to constants gives the exact QET action formally as[17]

$$\widetilde{I}[\varphi] = I_\varphi[\varphi] + \frac{1}{2}\operatorname{tr}\log\Big(-\nabla^2 + m_\Phi^2 + \lambda f_\varphi\Big). \tag{2.22}$$

[16]For instance, in flat space, at long distances satisfying $|x-y|\gg m_\Phi^{-1}$ the propagator would behave as $G(x,y)\sim|x-y|^{-d/2}e^{-m_\Phi|x-y|}$. For $m_\Phi=0$ it would be everywhere just the corresponding power law.

[17]We are implicitly assuming here that operators are trace class.

Under traces this gives the useful operator identity

$$\operatorname{tr}\log\Big(-\nabla^2+m_\Phi^2+\lambda f_\varphi\Big)=\operatorname{tr}\log\Big(-\nabla^2+m_\Phi^2\Big)+\operatorname{tr}\log(1+\lambda\, G\, f_\varphi). \tag{2.23}$$

The first term on the right-hand side is independent of φ and thus gives an uninteresting contribution to the quantum effective action. The φ-dependent term admits the expansion

$$\operatorname{tr}\log(1+\lambda\, G\, f_\varphi)=-\sum_{n=1}^{\infty}\frac{(-\lambda)^n}{n}\operatorname{tr}\left(G\, f_\varphi\right)^n, \tag{2.24}$$

where one can now explicitly write out the cyclic trace terms as

$$\operatorname{tr}\left(G\, f_\varphi\right)^n=\int\limits_{M^n} d^dx_1\cdots d^dx_n\, G(x_n,x_1)\, f_\varphi(x_1)\, G(x_1,x_2)\, f_\varphi(x_2)\cdots G(x_{n-1},x_n)\, f_\varphi(x_n). \tag{2.25}$$

Altogether, one finally arrives at the following action for the exact QET:

$$\widetilde{I}[\varphi]=I_\varphi[\varphi]-\frac{1}{2}\sum_{n=1}^{\infty}\frac{(-\lambda)^n}{n}\int\limits_{M^n} d^dx_1\cdots d^dx_n\prod_{k=1}^{n}G(x_{k-1},x_k)\, f_\varphi(x_k), \tag{2.26}$$

with $x_0\equiv x_n$. We thus see that already a simple Gaussian functional integral gives rise to a strongly non-local action involving infinitely many multi-integrals over arbitrarily many copies of the original space M. As for the previous example, one could similarly consider performing a derivative expansion of eq. (2.26) to simplify the theory in some regime. We leave this as an exercise for the reader.

Let us conclude by emphasizing that the construction of such strongly non-local actions characteristic of any exact QET in fact admits a standard perturbative treatment à la Feynman. This is because the local derivative expansion that destroys strong non-locality has nothing to do with the standard expansions at small couplings of perturbation theory. Importantly, this means that strongly non-local actions can actually be constructed using standard Feynman diagrams for arbitrarily complicated interacting theories at weak coupling. In other words, the framework that is relevant to this paper is as broad as perturbative quantum field theory techniques apply, and not at all limited to simple cases where exact path integration is directly doable as in the examples above. See figure 4 for a representation of the Feynman diagrams that arise in the construction of an exact QET.

3 State preparation

This section thoroughly studies the states that the path integrals of different theories naturally prepare, their general properties, and how they are related across theories.

3.1 Complete states

Let X be a manifold with boundary ∂X. Imposing boundary conditions ϕ_0 for the quantum fields on ∂X, the path integral on X computes the wavefunctional

$$\Psi[\phi_0]\equiv\langle\phi_0|\Psi\rangle=\frac{1}{\sqrt{Z}}\int\limits_{X}^{\phi|_{\partial X}=\phi_0}\mathcal{D}\phi\, e^{-I[\phi]}, \tag{3.1}$$

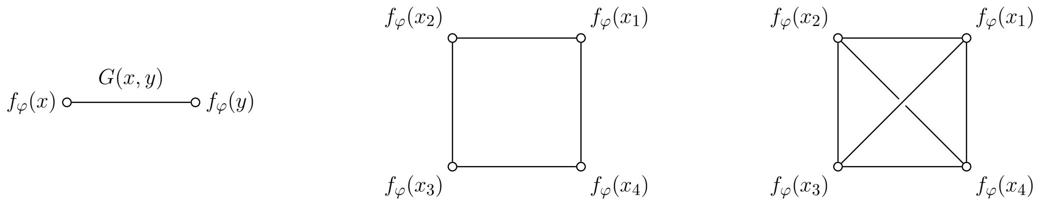

Figure 4. Feynman diagrams for the construction of the strongly non-local action of an exact QET. We consider a complete theory with interaction terms of the form $f_\varphi \Phi^k$ for $k \in \mathbb{Z}^+$ and where Φ is integrated out. The multi-integrals in the action come from integrating all points in these diagrams over M. On the left, the only diagram that contributes for $k = 1$, and which gives rise to the bilocal action in eq. (2.11). In the middle, the diagram which contributes the $n = 4$ term in eq. (2.26) for the theory with a $k = 2$ interaction term. Similar loops with all possible numbers n of vertices give all terms in the series in eq. (2.26). On the right, an example of a diagram which would also contribute a 4-local interaction among f_φ factors if the complete theory contained a $k = 3$ interaction.

where the normalization is given by eq. (2.1) with M constructed by gluing two copies of X along ∂X (cf. figure 2 with $\partial X \sim \Sigma$). Wavefunctionals corresponding to different states can be prepared by inserting distinct combinations of local operators in the path integral. The insertions that specify the state will be left implicit to avoid cluttering the notation. The wavefunctional in eq. (3.1) is a representation of the state $|\Psi\rangle$ in the space of field configurations ϕ_0 on ∂X. By leaving the boundary conditions on ∂X in eq. (3.1) unspecified, one obtains the state $|\Psi\rangle$ itself as a representation-independent functional operator,

$$|\Psi\rangle = \frac{1}{\sqrt{Z}} \int\limits_X \mathcal{D}\phi \, e^{-I[\phi]}. \tag{3.2}$$

The adjoint state $\langle\Psi|$ can be obtained by conjugating the fields in the path integral and evaluating it on a CRT-conjugate copy X^* of X. The Hilbert space $\mathcal{H}$ of the theory consists of the span of all quantum states $|\Psi\rangle$ prepared this way, endowed with the inner product that the path integral naturally defines,

$$\langle\Psi'|\Psi\rangle \equiv \int\limits_{\partial X \sim \partial X^*} \mathcal{D}\phi_0 \, \langle\Psi'|\phi_0\rangle\langle\phi_0|\Psi\rangle. \tag{3.3}$$

That states are appropriately unit-normalized can be easily seen by transitivity of the path integral. Namely, eq. (3.3) involves independent path integrals over X and X^* with identified boundary conditions along $\partial X \sim \partial X^*$ which are also integrated over. This is equivalent to just doing the path integral on M without cuts, which is by definition Z.[18] The following formal completeness relation can be read off from eq. (3.3):

$$\int\limits_{\partial X} \mathcal{D}\phi_0 \, |\phi_0\rangle \, \langle\phi_0| = \mathbb{I}_{\partial X}, \tag{3.4}$$

which provides a resolution of the identity on ∂X. This relation itself can be taken as the definition of the inner product on $\mathcal{H}$. Furthermore, it follows from eq. (3.3) that these

[18] With appropriate insertions corresponding to the state being prepared, if any.

JHEP05(2025)024

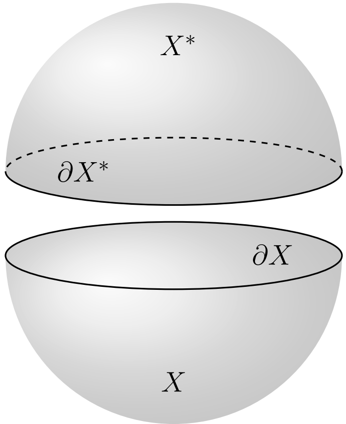

Figure 5. Representation of the path integral preparation of the density operator in eq. (3.6) by a complete theory. Because the theory is local, the path integral factorizes between the two dual spaces X and X^*. The path integral on X prepares $|\Psi\rangle$ on ∂X, and the CRT-conjugate path integral on X^* prepares $\langle\Psi|$ on ∂X^*. The identification $\partial X \sim \partial X^*$ defines the space $M \cong X \cup X^*$ on which the path integral computes the trace of the density operator.

configuration states obey the orthogonality condition[19]

$$\langle \phi_0' | \phi_0 \rangle = \delta(\phi_0' - \phi_0). \tag{3.5}$$

Together with the completeness relation in eq. (3.4), this implies that the $|\phi_0\rangle$ states form an orthonormal basis for the Hilbert space $\mathcal{H}$ of states of the theory on ∂X.

This formalism also allows one to directly prepare a density operator of the pure state by combining the preparation of a state and its dual into a single path integral,

$$|\Psi\rangle\langle\Psi| = \frac{1}{Z} \int\limits_{X \sqcup X^*} \mathcal{D}\phi \, e^{-I[\phi]}. \tag{3.6}$$

This is depicted in figure 5. Starting from the partition function definition of the path integral as in figure 2, the space $X \sqcup X^*$ can also be thought of as obtained by cutting open $M \cong X \cup X^*$ along $\Sigma \sim \partial X \sim \partial X^*$. Matrix elements of this operator are obtained by independently fixing boundary conditions for the quantum fields on ∂X and ∂X^*. Correspondingly, the trace is computed by identifying these configurations between both boundaries and path integrating over them as in eq. (3.3).

More generally, one may want to construct a state on some subspace $\Sigma \subseteq \partial X$. If $\Sigma = \partial X$ as above, it splits $M \cong X \cup X^*$ into two dual halves, so we refer to it as a *splitting* hypersurface (cf. Σ in figure 2).[20] If instead it is a proper subspace, $\Sigma \subset \partial X$, then cutting M open along

[19]To see this, note that by definition eq. (3.1) can also be written as

$$\langle \phi_0 | \Psi \rangle = \frac{1}{\sqrt{Z}} \int\limits_X \mathcal{D}\phi \, \delta(\phi|_{\partial X} - \phi_0) \, e^{-I[\phi]}.$$

Inserting eq. (3.4) on the left-hand side and comparing to the right-hand side leads to eq. (3.5). In canonical quantization, this orthogonality relation is simply declared. In the path integral formulation, it is a consequence of the functional formalism.

[20]If the manifold were Lorentzian, we would refer to such a hypersurface as a complete Cauchy surface.

JHEP05(2025)024

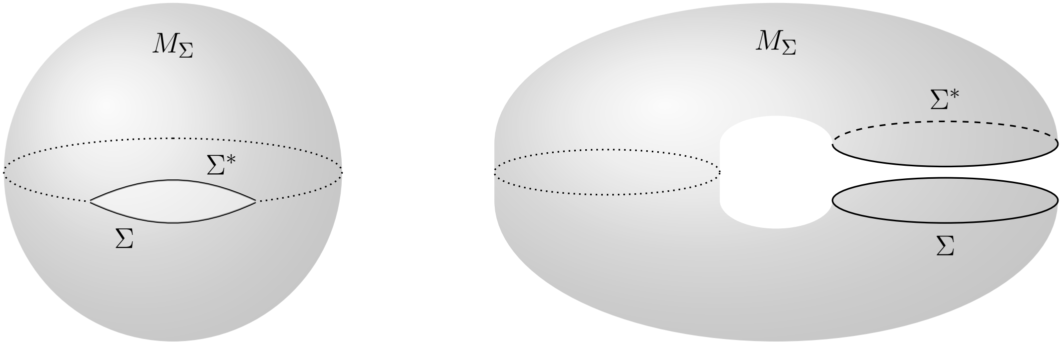

Figure 6. Representation of the path integral on spaces M_Σ obtained by cutting open M along a non-splitting hypersurface Σ. The non-splitting Σ on the left has $\partial\Sigma \neq \varnothing$, whereas on the right it is a cycle with $\partial\Sigma = \varnothing$. By leaving unspecified the boundary conditions for field configurations on the two sides of the cut Σ and Σ^*, the path integral prepares a general density operator on Σ. In field theory, such a state is generally mixed.

Σ does not split it into two dual halves; rather, X and X^* remain glued along the portion of their boundaries outside Σ. This is illustrated in figure 6. In general, the space obtained by cutting open M along Σ will be denoted by M_Σ. Performing a path integral on M_Σ can be easily seen to be equivalent to doing it on $X \sqcup X^*$, identifying the fields along $\Sigma \smallsetminus \partial X$ and $\Sigma \smallsetminus \partial X^*$, and integrating over configurations therein. This is nothing but the path integral implementation of a partial trace over configurations outside Σ, which is precisely what is needed to obtain the state of quantum fields on only Σ. Hence the path integral on M_Σ prepares the state on Σ as a reduced functional density operator which we write as[21]

$$\rho = \frac{1}{Z} \int\limits_{M_\Sigma} \mathcal{D}\phi\, e^{-I[\phi]}. \tag{3.7}$$

In field theory this is generally a mixed state unless Σ splits M. Since we used the path integral of the complete theory defined in eq. (2.1) to prepare the state in eq. (3.7), we regard such a ρ as a *complete state*. Computing the trace of ρ involves a path integral that now just runs over field configurations on $\Sigma \subset \partial X$, and a subspace completeness relation follows:

$$\int\limits_{\Sigma} \mathcal{D}\phi_0 \left|\phi_0\right\rangle \left\langle\phi_0\right| = \mathbb{I}_\Sigma. \tag{3.8}$$

The orthogonality condition from eq. (3.5) similarly holds on Σ. For future reference, we also quote the relations this would induce under the splitting of ϕ into φ and Φ sectors,

$$\mathbb{I}_\Sigma = \mathbb{I}_\Sigma^\varphi \otimes \mathbb{I}_\Sigma^\Phi, \qquad \int\limits_{\Sigma} \mathcal{D}\varphi_0 \left|\varphi_0\right\rangle \left\langle\varphi_0\right| = \mathbb{I}_\Sigma^\varphi, \quad \int\limits_{\Sigma} \mathcal{D}\Phi_0 \left|\Phi_0\right\rangle \left\langle\Phi_0\right| = \mathbb{I}_\Sigma^\Phi, \tag{3.9}$$

[21]The operator algebra of a quantum field theory on splitting hypersurfaces Σ is generally type-I von Neumann, which guarantees that pure states and traces like those given in eqs. (3.4) and (3.6) are indeed well-defined. For a non-splitting Σ states will be mixed, but so long as Σ itself has no boundaries, the algebra will be type II and admit a trace. Otherwise, if Σ has boundaries the algebra is generally type III, which does not admit pure states, density operators, or even a trace. The reader is welcome to assume Σ has no boundaries for all purposes of this paper, whether it is splitting or not.

JHEP05(2025)024

where the Hilbert space of φ fields $\mathcal{H}_\varphi$ is spanned by $|\varphi_0\rangle$ states, and the Hilbert space of Φ fields $\mathcal{H}_\Phi$ is spanned by $|\Phi_0\rangle$ states. The same argument that led to eq. (3.5) for the complete states of the theory applies to the φ and Φ states using the inner products from eq. (3.9),

$$\langle\varphi_0'|\varphi_0\rangle = \delta(\varphi_0' - \varphi_0), \qquad \langle\Phi_0'|\Phi_0\rangle = \delta(\Phi_0' - \Phi_0). \tag{3.10}$$

3.2 Exact effective states

If interested only in the physics of φ fields, however, one may want to consider the preparation of states using just the QET whose action is given by eq. (2.2). The state analogous to eq. (3.7) that the QET prepares on Σ can be obtained similarly by cutting open the path integral in eq. (2.3) along the same hypersurface Σ,[22]

$$\widetilde{\rho} = \frac{1}{\widetilde{\widetilde{Z}}} \int\limits_{M_\Sigma} \mathcal{D}\varphi\, e^{-\widetilde{I}[\varphi]}. \tag{3.11}$$

Since this state preparation involves the path integral of the exact QET defined in eq. (2.3), we regard such a $\widetilde{\rho}$ as an *exact QET state*. Importantly, the matrix elements of $\widetilde{\rho}$ involve boundary conditions for the φ fields only; the Φ fields were already integrated out in computing the quantum effective action $\widetilde{I}$. States of this kind define the Hilbert space of the exact QET, and their inner products are taken by the path integral for $\mathbb{I}^\varphi_\Sigma$ in eq. (3.9).

There is a very important lesson to be drawn from eq. (3.11) about state preparation in strongly non-local theories, particularly when they descend from some complete theory as an exact QET. For simplicity, here let M be connected and suppose Σ is not splitting, such that after cutting M open, the space M_Σ is still connected. All integrals inside the QET action $\widetilde{I}$ in eq. (3.11) then clearly run over all of M_Σ. By continuously extending Σ into a splitting hypersurface, one eventually reaches the disjoint limit in which $M_\Sigma \cong X \sqcup X^*$, for X and X^* the two dual half-spaces that make up M. Expecting continuity of the path integral in this limiting procedure, all integrals inside the QET action $\widetilde{I}$ will thus eventually have to run over both X and X^*. Schematically, for a double integral, this limit corresponds to

$$\int\limits_{M_\Sigma}\int\limits_{M_\Sigma}(\,\cdot\,) = \int\limits_{X}\int\limits_{X}(\,\cdot\,) + \int\limits_{X^*}\int\limits_{X^*}(\,\cdot\,) + 2\int\limits_{X}\int\limits_{X^*}(\,\cdot\,). \tag{3.12}$$

The appearance of the last term is remarkable: it couples the two disconnected spaces X and X^*. This is our first encounter of quantum wormholes, which we illustrate in figure 7. This cross-term in eq. (3.12) may seem surprising, but as will be seen turns out to be required by consistency of the framework. The basic reason it must be there is that the functional integral over Φ fields that defined the exact QET action $\widetilde{I}$ involved no cuts on the space M. As a result, the strong non-localities already present in the theory at the level of $\widetilde{I}$ are insensitive to the Σ cut. This in particular means that even if Σ is splitting, the two dual halves X and X^* of the resulting disjoint space M_Σ will remain coupled.

Correlations between the two dual spaces X and X^* in the preparation of a state are general indicators of mixing. One should thus suspect that quantum wormholes may cause

[22] As emphasized in footnote 9, if the QET action $\widetilde{I}$ depends on Φ field insertions, so will this QET state.

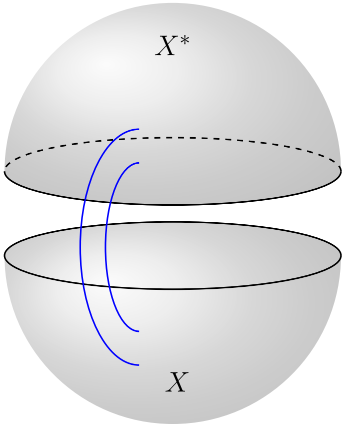

Figure 7. Representation of the path integral preparation of the density operator in eq. (3.11) by an exact QET. Here Σ has been taken to be a splitting hypersurface giving $M_\Sigma \cong X \sqcup X^*$. Because the theory is strongly non-local, the path integral does not factorize, but involves quantum wormholes coupling the two dual spaces X and X^*. As a result, what would be the 'bra' and 'ket' of the state are in fact coupled, meaning that the resulting state is generally mixed despite Σ being splitting.

exact QET density operators to be mixed. The linearity of the path integrals involved in eq. (3.11) allow for a suggestive rewriting that confirms this intuition:

$$\widetilde{\rho} = \frac{1}{\widetilde{Z}} \int_M \mathcal{D}\Phi \int_{M_\Sigma} \mathcal{D}\varphi \, e^{-I[\varphi,\Phi]}, \tag{3.13}$$

where we have simply used the definition of $\widetilde{I}$ from eq. (2.2). In particular, one can easily see that, altogether, eq. (3.13) is the result of splitting the path integral in eq. (3.7) into φ and Φ fields, and computing the one over the latter with Φ identified across the cut. From this point of view, $\widetilde{\rho}$ is the reduced density operator obtained from a state ρ by performing a partial trace over Φ fields. Hence eq. (3.11) can be interpreted as[23]

$$\widetilde{\rho} = \mathrm{Tr}_\Phi \, \rho, \tag{3.14}$$

where the trace over the Φ sector is defined by the path integral for $\mathbb{I}^\Phi_\Sigma$ in eq. (3.9). The structure and consequences of this trace will be the subject of section 3.4. For now, eq. (3.14) already demystifies why we should expect quantum wormholes to generally cause exact QET states to be mixed even when Σ is splitting. The identification of $\widetilde{\rho}$ as a reduced density operator for the φ fields will offer useful insights into the behavior of state replicas in section 4 and its operational distinguishability from states prepared by non-exact QET path integrals in section 5.

[23]In the context of Wilsonian renormalization in momentum space, this observation was also made in [73], where an explicit perturbative calculation of the entanglement entropy of the low-energy reduced density operator was performed for various scalar theories.

JHEP05(2025)024

3.3 Approximate effective states

A reasonable question at this point is whether the inclusion of quantum wormholes in the preparation of a state is optional.[24] Here we address what the consequences are of choosing not to include quantum wormholes.

Suppose one took an exact QET and decided to treat it as if it were the full theory of a system, i.e., as if it were a (strongly non-local) complete theory. Then one could consider constructing two completely separate pure states as in eq. (3.2). By dualizing one of them and then putting them together into a functional density operator as in eq. (3.6), one would obtain a pure state which we will call $\rho_\times$. In the path integral that prepares $\rho_\times$, the strongly non-local action would not involve any multi-integrals coupling X and X^*. For the double integral example in eq. (3.12), the last cross-term would be absent. Following this recipe, one is essentially preparing a pure state by truncating away quantum wormholes.

One may thus wonder how the pure state $\rho_\times$ is related to the mixed state $\widetilde{\rho}$ of the actual exact QET, both prepared by seemingly the same theory. In particular, how did $\rho_\times$ come to be a pure state of the φ fields despite their generic correlations with Φ fields in the complete theory? The answer is that in changing the rules of the path integral, one is in fact defining a strictly different theory. The resulting theory is precisely the outcome of erasing quantum wormholes from the exact QET. As explained in section 2.3 (cf. footnote 13), this is basically what happens if one performs a local derivative expansion and then decides to resum it back. Put differently, this new theory that prepares pure states like $\rho_\times$ is secretly no more than an approximate QET behaving like a weakly non-local theory with no quantum wormholes.[25]

Let us dissect further the preparation of these states to ascertain how much information is being lost by the approximate QET. A crucial realization is that strong non-localities are very important even when Σ is not splitting. To see this, suppose one decided to start by reducing an exact QET to an approximate QET in which the non-localities are only caused by derivatives of the fields of arbitrarily high order. As emphasized above, for a splitting Σ there would be no way of making sense of cross terms like those in eq. (3.12). However, more generally, even if a state is being prepared on a non-splitting Σ, such higher derivatives would be unable to couple fields right across the cut. In the exact QET these are genuinely coupled: indeed, from the perspective of Φ fields, points right across the cut are in a local neighborhood of each other. Once again, this is because the exact QET is obtained by integrating out the Φ fields without cuts, so the non-localities these induce persist across Σ as if the manifold had not been cut open (because for the Φ fields it was not). All this information is completely lost once one performs a derivative expansion of the strongly non-local theory, and then cuts open the path integral.

In intuitive terms, the exact QET still remembers the Φ sector, whereas the approximate QET has forgotten much of it. This is because in the preparation of $\widetilde{\rho}$ in eq. (3.11), the path integral is still treating the complete theory with action $\widetilde{I}$ as the full theory, and the

[24] The quantum gravity reader may find this reminiscent of the analogous question of whether or not to include gravitational bra-ket wormholes in the preparation of the quantum state of the universe [74–78].

[25] As a weakly non-local theory in disguise, the path integral inner product of this new theory would behave rather unconventionally. Namely, upon sewing $X \sqcup X^*$ by integrating over boundary conditions of fields identified across Σ, the multi-integrals would continue to run only over either X or X^* independently. This would be inequivalent to performing a full strongly non-local path integral over $X \cup X^*$ without cuts.

states being prepared know they are missing information about the Φ fields. This is how the exact QET path integral genuinely works, and mixed states like $\widetilde{\rho}$ are the states the exact QET path integral naturally prepares.[26] In contrast, in preparing $\rho_\times$, the path integral is instead treating the exact QET with action $\widetilde{I}$ as the full theory. In so doing, the only degrees of freedom this path integral accounts for are those associated to φ fields, and the Φ fields from the old complete theory no longer have any imprint on the states the path integral prepares. As a result the state $\rho_\times$ of φ fields on a splitting Σ hypersurface is a perfectly valid pure state of this approximate QET.

3.4 Ensembles of states

In trying to learn about complete theories, our primary interest is in understanding exact QET states $\widetilde{\rho}$ and how the exact QET path integrals that prepare them behave. The goal of this section is to first obtain a representation of $\widetilde{\rho}$ as a classical mixture of pure states of the exact QET, and then use these to identify ensembles of pure states of the complete theory, all of which yield the same $\widetilde{\rho}$ upon partial traces. This exercise will crystallize the sense in which exact QET states correspond to no single complete state of the theory, but to ensembles thereof.

Recall that we are denoting the Hilbert space of φ fields of the exact QET by $\mathcal{H}_\varphi$, and that of Φ fields of the complete theory by $\mathcal{H}_\Phi$. As usual, a mixed state on a given Hilbert space admits expansions into pure states on the same Hilbert space. These states may or may not be orthogonal, and in general this expansion is highly non-unique. Suppose $\widetilde{\rho}_*[\alpha]$ is an orthogonal basis of pure QET states for $\mathcal{H}_\varphi$ on Σ labeled by some function α. By definition these must obey the orthogonality and completeness relations

$$\mathrm{Tr}\left(\widetilde{\rho}_*[\alpha]\widetilde{\rho}_*[\alpha']\right) = \delta(\alpha-\alpha'), \qquad \int \mathcal{D}\alpha\, \widetilde{\rho}_*[\alpha] = \mathbb{I}^\varphi_\Sigma, \tag{3.15}$$

for some appropriate formal measure over α. On the left, the trace is over $\mathcal{H}_\varphi$, and is defined by the path integral over φ fields in eq. (3.9). The desired expansion of a mixed, exact QET state $\widetilde{\rho}$ in terms of the $\widetilde{\rho}_*$ states then takes the form

$$\widetilde{\rho} = \int \mathcal{D}\alpha\, P_*[\alpha]\, \widetilde{\rho}_*[\alpha], \tag{3.16}$$

where the probability density P_* can be obtained by the orthogonality property in eq. (3.15),

$$P_*[\alpha] = \mathrm{Tr}\left(\widetilde{\rho}_*[\alpha]\, \widetilde{\rho}\right). \tag{3.17}$$

Put differently, $\widetilde{\rho}$ can be interpreted as an average over an ensemble of $\widetilde{\rho}_*[\alpha]$ states with probability density function $P_*[\alpha]$. Of course, this density depends on the choice of spanning $\widetilde{\rho}_*$ states, which here has been arbitrary. However, as it turns out, a canonical choice of pure states spanning $\widetilde{\rho}$ descends rather naturally from the state of the complete theory.

[26] We remain agnostic as to whether there are pure states in an exact QET. For splitting hypersurfaces, the algebra of operators of an exact QET may be of type II, in contrast with the type I algebra of a complete theory (cf. a complete theory on a non-splitting hypersurface with no boundary, as on the right of figure 6).

Such a canonical choice can be constructed by exploiting the reduced density operator form of $\widetilde{\rho}$ in eq. (3.14). Making eq. (3.13) even more explicit, the functional form of the partial trace over Φ fields reads

$$\widetilde{\rho} = \int\limits_{\Sigma} \mathcal{D}\Phi_0 \, \langle\Phi_0|\, \rho \,|\Phi_0\rangle \,, \qquad \langle\Phi_0|\, \rho \,|\Phi_0\rangle = \frac{1}{\widetilde{Z}} \int\limits_{M_\Sigma} \mathcal{D}\varphi \int\limits_{M}^{\Phi|_\Sigma=\Phi_0} \mathcal{D}\Phi \, e^{-I[\varphi,\Phi]}. \tag{3.18}$$

In obtaining the second expression, we have simply noticed that the projection $\langle\Phi_0|\, \rho \,|\Phi_0\rangle$ corresponds to a constrained preparation of the state ρ as in eq. (3.7) with the Φ fields clamped onto the configuration Φ_0 on Σ. Appropriately normalized, each such projection of ρ naturally defines a state on the φ sector:

$$\widetilde{\rho}[\Phi_0] \equiv \frac{\widetilde{Z}}{Z^*_{\Phi_0}} \langle\Phi_0|\, \rho \,|\Phi_0\rangle = \frac{1}{Z^*_{\Phi_0}} \int\limits_{M_\Sigma} \mathcal{D}\varphi \int\limits_{M}^{\Phi|_\Sigma=\Phi_0} \mathcal{D}\Phi \, e^{-I[\varphi,\Phi]}, \tag{3.19}$$

where the normalization factor can be easily seen to be

$$Z^*_{\Phi_0} \equiv \int\limits_{M} \mathcal{D}\varphi \int\limits_{M}^{\Phi|_\Sigma=\Phi_0} \mathcal{D}\Phi \, e^{-I[\varphi,\Phi]}. \tag{3.20}$$

The states defined by eq. (3.19) are states on $\mathcal{H}_\varphi$ for the φ fields, and will henceforth be referred to as *projected states*.

We would like to identify these projected states as the canonical choice for an expansion like eq. (3.16) of an exact QET state. In particular, we are interested in the case in which the state is mixed solely because the theory that prepares it is an exact QET, and not the complete theory of the system. In other words, we want to understand the expansion of $\widetilde{\rho}$ on hypersurfaces Σ which are splitting, and on which the complete state ρ would be pure. For the projected states to be such a canonical choice, one just needs to check that they indeed are a set of pure states spanning $\widetilde{\rho}$. In fact, it will also be possible to think of them as an orthogonal basis in a precise sense from the perspective of the Φ sector.

That they span $\widetilde{\rho}$ trivially follows from their definition in eq. (3.19) applied to eq. (3.18). Regarding purity, note that by construction these states are obtained by projections of the Φ sector. In general, if ρ is a pure state of the complete theory, so will be any projection, which includes these projected states. Although this argument suffices, understanding how this happens from the path integral preparation in eq. (3.19) is informative. If the projection were not there, this would precisely correspond to the preparation of an exact QET state as in eq. (3.11), which generally is mixed due to correlations between φ and Φ fields. What the projection does is destroy correlations between φ and Φ fields from the get go. More precisely, for a splitting Σ, the projection on Σ causes the path integral over Φ fields to factorize into separate path integrals over each half-space of M. Hence, if this path integral is performed first, no quantum wormholes are generated between the two dual spaces into which M is being cut open. The result is thus the preparation of two state vectors dual to each other.[27] Explicitly, following the notation in eq. (3.2) and below, if $\rho = |\Psi\rangle\langle\Psi|$, projected states as functional vectors in $\mathcal{H}_\varphi$ read

$$|\psi\rangle_{\Phi_0} \equiv \frac{\langle\Phi_0|\Psi\rangle}{\sqrt{Z^*_{\Phi_0}}} = \frac{1}{\sqrt{Z^*_{\Phi_0}}}\int_X \mathcal{D}\varphi \int_X^{\Phi|_{\partial X}=\Phi_0} \mathcal{D}\Phi\, e^{-I[\varphi,\Phi]}, \tag{3.21}$$

which manifestly demonstrates their purity. It thus follows that the states $\widetilde{\rho}[\Phi_0]$ defined in eq. (3.19) form the desired canonical spanning set of pure states for the exact QET state in eq. (3.18). Explicitly, the decomposition of the mixed state $\widetilde{\rho}$ in eq. (3.16) takes the form

$$\widetilde{\rho} = \int_\Sigma \mathcal{D}\Phi_0\, P[\Phi_0]\, \widetilde{\rho}[\Phi_0], \qquad P[\Phi_0] \equiv \frac{Z^*_{\Phi_0}}{\widetilde{Z}}, \tag{3.22}$$

where $Z^*_{\Phi_0}$ was defined in eq. (3.20). This realizes the exact QET state $\widetilde{\rho}$ as an ensemble average over $\widetilde{\rho}[\Phi_0]$ states with probability density $P[\Phi_0]$, and motivates introducing the following notation for expectation values in this ensemble:

$$\langle \cdot \rangle_P \equiv \int_\Sigma \mathcal{D}\Phi_0\, P[\Phi_0]\, (\,\cdot\,), \tag{3.23}$$

which note is indeed correctly normalized to give $\langle 1\rangle_P = 1$. Using this, eq. (3.22) reads

$$\widetilde{\rho} = \langle \widetilde{\rho}[\Phi_0]\rangle_P\,. \tag{3.24}$$

As quoted, eq. (3.24) holds in fact for states on any Σ hypersurface. If Σ is splitting, then eq. (3.24) indeed is a decomposition of the mixed exact QET state into pure projected states. If instead Σ is not splitting, then this is a decomposition into mixed states, but whose mixing is solely caused by spatial correlations of quantum fields (both φ and Φ) between Σ and its complementary subsystem. This way, for general Σ, eq. (3.22) succeeds in decomposing $\widetilde{\rho}$ into a family of states with no correlations between φ and Φ fields.

While eq. (3.22) takes the same form as eq. (3.16), the projected states do not enjoy an orthogonality property like eq. (3.15) on $\mathcal{H}_\varphi$, and thus their P density measure cannot be extracted as in eq. (3.17). In general the failure of orthogonality only allows for the bound $\mathrm{Tr}(\widetilde{\rho}[\Phi_0]\widetilde{\rho}) \geq P[\Phi_0]$. The measure $P[\Phi_0]$ is more naturally associated to Φ field configurations Φ_0, which do obey the orthogonality condition in eq. (3.10). Our projected states on $\mathcal{H}_\varphi$ can be lifted back to states acting on the full Hilbert space of the complete theory $\mathcal{H}$ in arbitrary ways, such as

$$\rho_\xi[\Phi_0] \equiv \widetilde{\rho}[\Phi_0] \otimes |\xi_{\Phi_0}\rangle\langle\xi_{\Phi_0}|\,, \qquad \widetilde{\rho}[\Phi_0] = \mathrm{Tr}_\Phi\, \rho_\xi[\Phi_0], \tag{3.25}$$

for any pure state $|\xi_{\Phi_0}\rangle$ in $\mathcal{H}_\Phi$. Since eq. (3.26) holds for any family of states $|\xi_{\Phi_0}\rangle$ labeled by Φ_0, we indeed see that this provides an infinite family of ensembles of pure states of the complete theory, all of which can be combined to reproduce the exact same quantum

[27] The argument for purity from the perspective of the complete theory is trivial and uninformative: eq. (3.19) is obviously pure as it corresponds to preparing a complete state as in eq. (3.6), but with partially specified boundary conditions on the sector of Φ fields.

JHEP05(2025)024

state $\widetilde{\rho}$ on $\mathcal{H}_\varphi$. There is though an obvious canonical choice corresponding to $\xi_{\Phi_0} = \Phi_0$ and which we denote by

$$\Omega[\Phi_0] \equiv \widetilde{\rho}[\Phi_0] \otimes |\Phi_0\rangle \langle\Phi_0| \,. \tag{3.26}$$

This lifting does not invert the projection of ρ in eq. (3.18) (i.e., $\Omega \neq \rho$ in general); rather, it identifies a canonical ensemble of states of the complete theory with orthogonal Φ field configurations which precisely reproduces the exact QET state $\widetilde{\rho}$ in its φ sector. Using eq. (3.23), this ensemble of states we are referring to can be written

$$\Omega \equiv \langle\Omega[\Phi_0]\rangle_P \,. \tag{3.27}$$

Equipped with this notation, we reproduce the statistics of the φ sector that the exact QET state captures as coming from a classical ensemble via (cf. eq. (3.18))

$$\widetilde{\rho} = \mathrm{Tr}_\Phi\, \Omega. \tag{3.28}$$

In addition, because the $\Omega[\Phi_0]$ states are orthogonal on $\mathcal{H}_\Phi$, the P density measure can now be extracted from the ensemble by projection onto the Φ sector,

$$P[\Phi_0] = \mathrm{Tr}\left(|\Phi_0\rangle \langle\Phi_0|\ \Omega\right), \tag{3.29}$$

where we are using that projected states are properly normalized to $\mathrm{Tr}\,\widetilde{\rho}[\Phi_0] = 1$.

Instead of recovering orthogonality by lifting projected states to act on $\mathcal{H}$, one may consider constructing an orthogonal set out of them. For a given $\widetilde{\rho}$, the projected states $\widetilde{\rho}[\Phi_0]$ may only act on a subspace of $\mathcal{H}_\varphi$. However, more generally, the set of all $\widetilde{\rho}[\Phi_0]$ for all possible exact QET states $\widetilde{\rho}$ is guaranteed to act on all of $\mathcal{H}_\varphi$. Here we will focus on constructing an orthogonal set only using projected states $\widetilde{\rho}[\Phi_0]$ for a given $\widetilde{\rho}$. We would like to diagonalize the subspace $\widetilde{\mathcal{H}} \subseteq \mathcal{H}_\varphi$ on which these act. By construction, in general $\dim\widetilde{\mathcal{H}}$ is upper bounded by the smallest of $\dim\mathcal{H}_\varphi$ and $\dim\mathcal{H}_\Phi$. If $\dim\widetilde{\mathcal{H}} < \dim\mathcal{H}_\Phi$, the projected states will be guaranteed to form an over-complete basis for $\widetilde{\mathcal{H}}$. In other words, there will be combinations of them which give null states,

$$\int_\Sigma \mathcal{D}\Phi_0\, \nu[\Phi_0]\, \widetilde{\rho}[\Phi_0] = 0, \tag{3.30}$$

for certain choices of functionals ν. The space of null states corresponds to the kernel of the map $\Omega[\Phi_0] \to \widetilde{\rho}[\Phi_0]$, whose dimension will be $\dim\mathcal{H}_\Phi - \dim\widetilde{\mathcal{H}}$ or zero, whichever is greater. Combinations of projected states can also be used to form an orthonormal basis for $\widetilde{\mathcal{H}}$. To make contact with the notation in eqs. (3.15)–(3.17), we write these as

$$\widetilde{\rho}_*[\alpha] \equiv \int_\Sigma \mathcal{D}\Phi_0\, \alpha[\Phi_0]\, \widetilde{\rho}[\Phi_0]. \tag{3.31}$$

In other words, we are identifying α as the functionals giving combinations of the projected states obeying the orthogonality relation in eq. (3.15). The states in eq. (3.31) for a fixed $\widetilde{\rho}$ may only obey a completeness relation on $\widetilde{\mathcal{H}}$, but that is all we need. As combinations of

JHEP05(2025)024

projected states, the $\widetilde{\rho}_*[\alpha]$ states can be thought of as coming from a superposition of states of Φ fields with different configurations on Σ. Because they form an orthonormal basis of operators on $\widetilde{\mathcal{H}}$, they can always be identified as pure states on that subspace. This notion of purity corresponds to purity on the exact QET Hilbert space $\mathcal{H}_\varphi$ when $\dim \mathcal{H}_\Phi \geq \dim \mathcal{H}_\varphi$ and $\widetilde{\mathcal{H}} \cong \mathcal{H}_\varphi$, but is also guaranteed by purity of the projected states if Σ is splitting even if $\widetilde{\mathcal{H}} \subset \mathcal{H}_\varphi$. The expansion for $\widetilde{\rho}$ with density measure $P_*[\alpha]$ over $\widetilde{\rho}_*[\alpha]$ states can be related to the expansion with density measure $P[\Phi_0]$ over $\widetilde{\rho}[\Phi_0]$ states using eq. (3.17), obtaining

$$P_*[\alpha] = \int\limits_{\Sigma} \mathcal{D}\Phi_0 \int\limits_{\Sigma} \mathcal{D}\Phi_0' \, \alpha[\Phi_0] \, P[\Phi_0'] \, \mathrm{Tr}\left(\widetilde{\rho}[\Phi_0] \, \widetilde{\rho}[\Phi_0']\right). \tag{3.32}$$

Altogether, the discussion in this section allows for a precise identification of the exact QET state $\widetilde{\rho}$ as descending from an ensemble average over orthogonal states $\Omega[\Phi_0]$ of the complete theory with probability density functional $P[\Phi_0]$, or as an ensemble average over orthogonal states $\widetilde{\rho}_*[\alpha]$ with probability density functional $P_*[\alpha]$. In the former interpretation, we are taking the viewpoint that there is a unique complete theory, and that the ensemble captures the failure of the QET to distinguish Φ field configurations. In the latter, we are simply using the complete theory to select a natural over-complete basis for the exact QET states, and identifying sets of null states and spanning states. These constructs are the field-theoretic analogue to α-states in the wormhole literature, as discussed in section 6.1.

3.5 Ensembles of theories

An interesting alternative perspective arises if one decides to begin by declaring a family of theories functionally parameterized by Φ_0, with actions defined by

$$\frac{1}{\widetilde{Z}} e^{-I_{\Phi_0}[\varphi]} \equiv \frac{1}{Z^*_{\Phi_0}} \int\limits_{M}^{\Phi|_\Sigma = \Phi_0} \mathcal{D}\Phi \, e^{-I[\varphi,\Phi]}, \tag{3.33}$$

where $\widetilde{Z}$ and $Z^*_{\Phi_0}$ appear as normalization constants for fixed Φ_0. Because of their obvious relation to the projected states defined in eq. (3.19), these theories will be referred to as *projected theories*. In terms of this family of theories, the exact QET state is prepared by

$$\widetilde{\rho} = \frac{1}{\widetilde{Z}} \int\limits_{M_\Sigma} \mathcal{D}\varphi \, \left\langle e^{-I_{\Phi_0}[\varphi]} \right\rangle_P. \tag{3.34}$$

This expression allows one to interpret $\widetilde{\rho}$ as being prepared by an ensemble of projected theories, rather than by any single theory. Correspondingly, the QET partition function takes the form of an ensemble average over partition functions of these theories,

$$\widetilde{Z} = \langle Z_{\Phi_0} \rangle_P \,, \qquad Z_{\Phi_0} \equiv \int\limits_{M} \mathcal{D}\varphi \, e^{-I_{\Phi_0}[\varphi]}. \tag{3.35}$$

The objects Z_{Φ_0} in eq. (3.35) and $Z^*_{\Phi_0}$ in eq. (3.33) should not be confused. In particular, $Z^*_{\Phi_0}$ is just a normalization constant that is evaluated and set once and for all when the projected theory action $I_{\Phi_0}[\varphi]$ is defined. In contrast, the partition function Z_{Φ_0} defines a functional integral over φ fields on M. In fact, in the absence of insertions, note that numerically Z_{Φ_0} evaluates to $\widetilde{Z}$, not $Z^*_{\Phi_0}$. With insertions, however, the path integral that Z_{Φ_0} defines gives results which depend on Φ_0 nontrivially, whereas that of $\widetilde{Z}$ of course does not.

A projected theory with partition function Z_{Φ_0} may be regarded as a complete theory in its own right. It is not an exact QET: the operation being performed on the Φ fields in eq. (3.33) is a projection onto a specific field configuration state $|\Phi_0\rangle$ on Σ, not a partial trace over states on Σ. This theory is strongly non-local because of the preparation of the state of the Φ fields, but it has no quantum wormholes across Σ due to the projection onto Φ_0. Hence these projected theories naturally prepare pure states on splitting hypersurfaces, which are no other than the projected states in eq. (3.19).

In terms of state preparation, these projected theories are complete theories for the φ fields in a superselection sector of the original theory where the configuration state of the Φ fields on Σ is given by Φ_0. This state is superselected in the sense that no matter what operators one decides to insert into the path integral, the theory stays within the sector in which $\Phi|_\Sigma = \Phi_0$. This notion of superselection is entirely analogous to that of α-states and baby universes in quantum gravity, as we elaborate on in section 6.1.5. From the perspective of the original complete Euclidean theory on M, these projected theories are the result of postselecting on the state of Φ fields on the Σ hypersurface. Indeed, the path integral is not preparing a state that happens to freely evolve deterministically to the configuration Φ_0 for the Φ fields on Σ; rather, this configuration is enforced by postselection.

Optionally, one could consider performing a derivative expansion of the strong non-localities of a projected theory and a truncation down to a local theory. The ensemble average analogous to eq. (3.35) over such truncated projected theories would give rise to an averaged truncated QET. At the level of the truncated QET action, the coupling constants would thus acquire a statistical meaning as coming from an ensemble of theories with different coupling constants. Every theory in the ensemble comes from a different state of the Φ fields of the original complete theory, and thus with an associated probability of being measured by an observer. This literally implies an indeterminacy in the couplings an observer would measure in trying to describe the complete theory in terms of a truncated QET. More strikingly, in general the states of the Φ fields will have no symmetries and be position dependent. Correspondingly, one should expect a poor description of the physics of the complete theory if the couplings are assumed to be constant across space. Observations of this flavor have also been made in the context of effective theories in quantum gravity [79–84]. In string theory phenomenology it is well understood that low-energy couplings are really expectation values of dynamical background fields. Here we see these as general QET features, with no need to appeal to gravity or string theory.

In conclusion, we find that the theory ensemble interpretation of the exact QET that eq. (3.35) provides and the previous state ensemble interpretation of the exact QET states from eq. (3.28) are two sides of the same coin.[28] Summarizing: the exact QET is an ensemble of projected theories, the exact QET states are ensembles of projected states, and the connection between these two facts is that projected states are precisely the states that the path integral of the projected theory prepares.

[28] See e.g. [85, 86] for a quantum gravity realization of this idea.

4 Replica path integrals

A sharp distinguishing feature among the theories in section 3 is whether and why the states they prepare are pure or mixed. This section is devoted to quantifying this in terms of entropic measures which are computable via path integrals, and thus further elucidate how exact QET path integrals work.

4.1 Entropic measures

From eq. (3.14), we see that the QET state $\widetilde{\rho}$ is the reduced density operator obtained by tracing out the Φ fields from ρ. From a quantum mechanical perspective, one would like to know how much information about the φ fields in the complete state ρ is still retained in $\widetilde{\rho}$ intrinsically within φ fields. This depends on the amount of correlations between the φ and Φ fields, particularly their shared entanglement. Information theoretic measures of such correlations typically involve the calculation of entropies.

Given a system with density operator σ, a quantity of interest is its von Neumann entropy,

$$S(\sigma) \equiv - \operatorname{Tr} \sigma \log \sigma. \tag{4.1}$$

Because S is a nonlinear functional of σ, standard path integral techniques cannot be straightforwardly applied to compute the trace above. Instead, a fruitful strategy is to employ the following mathematical identity:

$$\log x = \lim_{n \to 0} \frac{x^n - 1}{n} = \lim_{n \to 0} \partial_n x^n. \tag{4.2}$$

Applying this to eq. (4.1) and using $\operatorname{Tr} \sigma = 1$ leads to

$$S(\sigma) = - \lim_{n \to 1} \frac{\operatorname{Tr} \sigma^n - 1}{n - 1} = - \lim_{n \to 1} \partial_n \operatorname{Tr} \sigma^n. \tag{4.3}$$

This is the starting point for a path integral procedure known as the replica trick for the von Neumann entropy. This mathematical trick consists of calculating traces of powers of the density operator, and analytically continuing them near $n = 1$. At values $n \neq 1$, these objects are of interest themselves and compute Rényi entropies,

$$S_n(\sigma) \equiv - \frac{1}{n - 1} \log \operatorname{Tr} \sigma^n, \tag{4.4}$$

which can be easily checked to yield the limit $\lim_{n \to 1} S_n(\sigma) = S(\sigma)$.[29] Crucially, for positive integers $n \in \mathbb{Z}^+$, an appropriate path integral construct does allow one to compute these Rényi entropies. This amounts to calculating traces of powers of density operators, $\operatorname{Tr} \sigma^n$, to which we refer as *state replicas*.

4.2 Replica manifolds

Consider a state ρ prepared on $\Sigma \subset M$ by the local path integral of a complete theory on M_Σ. It will be convenient here to refer to the two sides of the cut of M_Σ by $\Sigma^\pm$. Recall that the path integral calculates $\operatorname{Tr} \rho$ by identifying fields across the Σ cut and integrating over

[29] Note, in particular, that one can write $\lim_{n \to 1} \partial_n \operatorname{Tr} \sigma^n = \lim_{n \to 1} \partial_n \log \operatorname{Tr} \sigma^n$.

them. As explained in section 3, this is equivalent to performing the path integral on M_Σ after identifying $\Sigma^+ \sim \Sigma^-$, which gives back the manifold M.

Consider now $\operatorname{Tr}\rho^2$, which requires two copies of the manifold, $M_{\Sigma 1}$ and $M_{\Sigma 2}$, to build each replica of the state inside the trace. At the level of matrix elements it is clear what the path integral should do. For the product ρ^2, we would like to identify the field configurations at Σ_1^+ and Σ_2^-, and integrate over them. To then take the trace $\operatorname{Tr}\rho^2$, we additionally identify the fields at Σ_2^+ and Σ_1^-, and integrate over them. Again, this field identifications can also be implemented by performing analogous identifications at the level of the spaces themselves, specifically $\Sigma_1^+ \sim \Sigma_2^-$ and $\Sigma_2^+ \sim \Sigma_1^-$. The resulting space consists of two copies of M_Σ which are appropriately identified along their respective cuts. Because the theory is local, doing the path integral on each copy separately with the desired identification of the fields along cuts is equivalent to just doing the path integral on the space consisting of the two copies of M_Σ that are themselves identified along their cuts.

The more general case of calculating $\operatorname{Tr}\rho^n$ for $n \in \mathbb{Z}^+$ replicas of the state ρ easily follows. The desired periodic identification between field configurations from cut to cut across replicas under the trace can be equivalently implemented by the construction of an appropriate *replica manifold* $M_\Sigma^{(n)}$. This space is obtained by taking n copies of M_Σ, and cyclically gluing them along their Σ cuts, such that the Σ_k^+ side of the k^{th} replica is identified with the Σ_{k+1}^- side of the $(k+1)^{\text{th}}$ replica, with $k+n \sim k$. The resulting replica manifold $M_\Sigma^{(n)}$ can be easily described: it is an n-sheeted covering of M_Σ branched over Σ.

It will clearly be important in this section to keep track of which manifold a path integral is to be evaluated on. We will make so explicit as follows: if Z is the partition function of the theory that prepares the state ρ, then we will use $Z[X]$ to denote the path integral of this theory performed on X. The upshot of the discussion above is thus the following path integral prescription for computing state replicas in local theories:[30]

$$\operatorname{Tr}\rho^n = \frac{Z[M_\Sigma^{(n)}]}{Z[M]^n}. \tag{4.5}$$

The denominator is a normalizing factor that makes $\operatorname{Tr}\rho^n = 1$ for ρ a pure state. To see this, recall that for a field theory a path integral may prepare a pure state on Σ only if it is a splitting hypersurface. When this is the case, the components in which M is split by Σ can be cyclically re-glued into copies of M itself. In particular, if $M_{\Sigma k} \cong X_k^- \sqcup X_k^+$, the half-space X_k^+ gets glued to the half-space X_{k+1}^- of the next replica, thus giving a space isomorphic to M. Altogether, the replica manifold reassembles itself into $M_\Sigma^{(n)} \cong M^n$, i.e., $M_\Sigma^{(n)}$ is isomorphic to n disjoint copies of M with no cuts. If the theory is local, then from the factorization property in eq. (2.7) it would follow that $Z[M^n] = Z[M]^n$, thus yielding $\operatorname{Tr}\rho^n = 1$, as desired.

For later comparison, it will be useful to write out eq. (4.5) explicitly. Using the definition of the path integral for the complete theory in eq. (2.1), the replica calculation for such a local theory reads

$$\operatorname{Tr}\rho^n = \frac{1}{Z^n}\int\limits_{M_\Sigma^{(n)}} \mathcal{D}\phi\, e^{-I[\phi]} = \frac{1}{Z^n}\int\limits_{M_\Sigma^{(n)}} \mathcal{D}\varphi \int\limits_{M_\Sigma^{(n)}} \mathcal{D}\Phi\, e^{-I[\varphi,\Phi]}, \tag{4.6}$$

[30] A naïve application of this formula to a strongly non-local theory would be wrong.

where in the second equality we are simply showcasing the splitting into φ and Φ fields. In either expression, locality of the theory means the action takes the form of a single integral over $M_\Sigma^{(n)}$. Hence if $M_\Sigma^{(n)} \cong M^n$, the action just decomposes as the right-hand side of eq. (2.6) with M_k labeling the k^{th} replica of M. Correspondingly, the path integral in eq. (4.6) factorizes into a product of n identical integrals over M.

A potential pitfall at this point is to conclude that, just because $M_\Sigma^{(n)}$ looks like n disjoint copies of M for Σ splitting, for a general path integral the replica calculation on $M_\Sigma^{(n)}$ will always factorize. In fact, as explained in the next section, the replica path integral for an exact QET on $M_\Sigma^{(n)}$ does not factorize. This is to be contrasted with the behavior of an exact QET on M^n, which as in eq. (2.8) would indeed give a factorizing answer. The crucial point is that, although isomorphic, the spaces $M_\Sigma^{(n)}$ and M^n are obtained very differently.

4.3 Replica wormholes

Before jumping into the technical details, let us first emphasize why the factorization of a general replica path integral for $M_\Sigma^{(n)} \cong M^n$ would be an inconsistency of the framework itself. Consider the path integral that prepares the exact QET state $\widetilde{\rho}$ on Σ. In eq. (3.14), $\widetilde{\rho}$ was identified as a reduced density operator obtained from ρ by tracing out Φ fields. The state $\widetilde{\rho}$ will be mixed so long as ρ has entanglement between φ and Φ fields, whether or not Σ is splitting. In other words, how mixed $\widetilde{\rho}$ is depends on two phenomena: entanglement between φ fields inside and outside Σ (if Σ is not splitting), and entanglement between φ fields in Σ and Φ fields everywhere. If only the latter happens, the replica manifold will be isomorphic to M^n, as argued above, but the entanglement between φ and Φ fields will still cause $\widetilde{\rho}$ to be mixed. A general fact about a mixed density operator is that $\operatorname{Tr}\widetilde{\rho}^n$ is a strictly decreasing function of n so, in particular, $\operatorname{Tr}\widetilde{\rho}^n < 1$ for all $n > 1$.[31] Hence, in a replica calculation like eq. (4.5), the QET path integral with partition function $\widetilde{Z}$ that prepares $\widetilde{\rho}$ cannot possibly factorize. How this non-factorization occurs at the level of path integrals is explained next.

As emphasized throughout section 4.2, locality played an important role in deriving eq. (4.5) as a path integral prescription for calculating state replicas. However, suppose the path integral under consideration is that of an exact QET with a strongly non-local action. Naïvely, given eq. (4.5), to calculate state replicas for the exact QET state $\widetilde{\rho}$, one could proceed to evaluate $\widetilde{Z}[M_\Sigma^{(n)}]$. If Σ is splitting, by the isomorphism $M_\Sigma^{(n)} \cong M^n$ one would be tempted to declare a factorizing answer as in eq. (2.8), which once again would be wrong. But even worse, if Σ is not splitting, this would correspond to evaluating the full strongly non-local path integral on all of $M_\Sigma^{(n)}$, meaning that all multi-integrals in the action would run over the full replica manifold $M_\Sigma^{(n)}$. This would be very different from what the replica calculation is supposed to do.

The reason is simple: evaluating the strongly non-local theory on the full replica manifold would lead to an action with multi-integrals coupling all replicas $M_{\Sigma k}$ that make up $M_\Sigma^{(n)}$. In particular, there would be multi-integrals mixing $M_{\Sigma k}$ and $M_{\Sigma k'}$ for all $k' \neq k$. This should not happen, since the preparation of identical copies of the state should involve completely independent path integrals. Namely, the full path integral on $M_{\Sigma k}$ that prepares the k^{th} copy of ρ involves strong non-localities with multi-integrals in the action over $M_{\Sigma k}$ only. Field

[31]This follows from monotonicity of the p-norm, which also leads to monotonicity of the Rényi entropies.

configurations are then to be identified along Σ cuts, but never does this step introduce strong non-localities taking the form of multi-integrals over distinct replicas.

From these observations, we learn that in the replica calculation for the $\widetilde{Z}$ path integral, the multi-integrals in the action should not run over the full replica manifold, but only over each different replica independently. We capture this behavior of the action for a replica calculations on $M_\Sigma^{(n)}$ with the following notation:

$$\widetilde{I}_\circ \equiv \sum_{k=1}^{n} \widetilde{I}_k, \tag{4.7}$$

where $\widetilde{I}_k$ is the full exact QET action evaluated on the k^{th} replica $M_{\Sigma k}$. Correspondingly, the desired path integral to be performed on $M_\Sigma^{(n)}$ is

$$\widetilde{Z}_\circ[M_\Sigma^{(n)}] \equiv \int\limits_{M_\Sigma^{(n)}} \mathcal{D}\varphi\, e^{-\widetilde{I}_\circ[\varphi]}, \tag{4.8}$$

where even though the action decomposes as in eq. (4.7), the path integral still is over continuous field configurations φ on $M_\Sigma^{(n)}$. The appropriate generalization of eq. (4.5) for the calculation of state replicas is thus

$$\operatorname{Tr}\widetilde{\rho}^{\,n} = \frac{\widetilde{Z}_\circ[M_\Sigma^{(n)}]}{\widetilde{Z}[M]^n}. \tag{4.9}$$

This prescription now holds for the path integral of any theory, be it a complete theory, an exact QET, an approximate QET, or a truncated QET.

In light of the decomposition of the action in eq. (4.7), one may worry that the path integral $\widetilde{Z}_\circ$ would end up factorizing for $M_\Sigma^{(n)} \cong M^n$, contradicting the consistency arguments at the beginning of this section. In fact, factorization does not occur, and the mechanism for this is as follows. Assume Σ is splitting, since this is the sharp case where factorization would imply a qualitative failure of the replica calculation. The k^{th} replica of the manifold M_Σ on which $\widetilde{\rho}$ is prepared then splits as

$$M_{\partial X_k} \cong X_k^- \sqcup X_k^+, \tag{4.10}$$

where $X_k^\pm$ are the two dual halves that make up M when glued along their respective boundaries $\partial X_k^\pm$. As explained in section 4.2 and particularly below eq. (4.5), these are then glued back into the replica manifold $M_{\partial X}^{(n)}$ in a cyclic fashion. Namely, the cut identifications are of the form $\partial X_k^+ \sim \partial X_{k+1}^-$, thereby joining the X_k^+ half of the k^{th} replica to the X_{k+1}^- half of the next one. As a result, the reassembling of half-spaces into n disjoint copies of M involves dual halves from consecutive replicas giving spaces $X_k^+ \cup X_{k+1}^- \cong M$, with $k+n \sim k$. Though isomorphic, the replica labels matter, since these keep track of which specific spaces the action integrals run over. The replica manifold for a splitting hypersurface one obtains is precisely given by:

$$M_{\partial X}^{(n)} = \bigsqcup_{k=1}^{n} M_{k|k+1}, \qquad M_{k|k+1} \equiv X_k^+ \cup X_{k+1}^-. \tag{4.11}$$

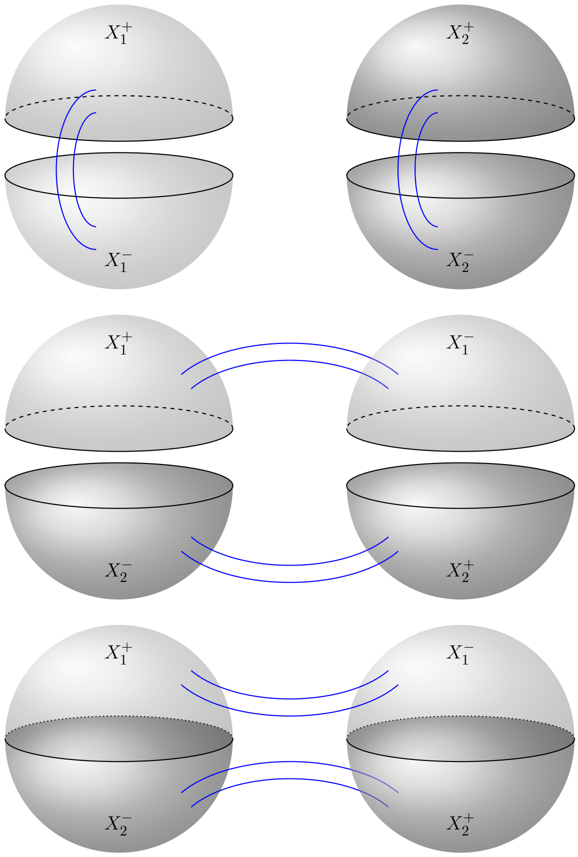

Figure 8. Step-wise representation of the state replica calculation in eq. (4.9) for an exact QET. For simplicity, the figures show the $n = 2$ case that computes the purity, and consider a state prepared on a splitting hypersurface. In the first figure, the two replicas of the state are separately prepared by independent path integrals on distinct copies of $M_{\partial X} \cong X^- \sqcup X^+$ (cf. replicating figure 7). The second figure is just a convenient redrawing of the first one anticipating the cyclic identifications that the trace implements, keeping track of which spaces are coupled by quantum wormholes. In the last figure, the trace is performed by identifying $\partial X_1^+ \sim \partial X_2^-$ and $\partial X_2^+ \sim \partial X_1^-$, i.e., by acting with the 'bra' of the first replica onto the 'ket' of the second one and by cyclicity also acting with the 'bra' of the latter onto the 'ket' of the former. The resulting replica manifold, $M_{\partial X}^{(2)}$, though isomorphic to M^2, swaps halves and is given by eq. (4.11), not eq. (4.12). This is how replica wormholes arise.

An illustration of this structure can be seen in figure 8. In contrast, the usual notation M^n refers to a standard disjoint union of the n original replicas,

$$M^n = \bigsqcup_{k=1}^{n} M_k, \qquad M_k \equiv X_k^+ \cup X_k^- . \tag{4.12}$$

While obviously isomorphic $M_{\partial X}^{(n)} \cong M^n$ as claimed throughout, these two spaces are different and explain why the path integral in eq. (4.8) indeed does not factorize. Written out using the notation in eq. (4.7), we have

$$\widetilde{Z}_\circ[M_{\partial X}^{(n)}] = \int\limits_{\bigsqcup_{k=1}^{n} M_{k|k+1}} \mathcal{D}\varphi \prod_{k=1}^{n} e^{-\widetilde{I}_k[\varphi]}, \tag{4.13}$$

where recall $\widetilde{I}_k$ is the action with integrals only on $M_k = X_k^- \cup X_k^+$. Hence we see that although the path integral is over spaces $M_{k|k+1}$ which are disjoint for distinct k, the multi-integrals in the action $\widetilde{I}_k$ which run over M_k will couple the X_k^- half of $M_{k-1|k}$ to the X_k^+ half of $M_{k|k+1}$. Because the disjoint spaces the path integral runs over are genuinely coupled by quantum wormholes, as illustrated in figure 8, it is clear that the replica path integral does not factorize. In addition, because mixing causes $\operatorname{Tr}\widetilde{\rho}^n \leq 1$, for general $n \in \mathbb{Z}^+$ one expects

$$\widetilde{Z}_\circ[M_{\partial X}^{(n)}] \leq \widetilde{Z}[M]^n, \tag{4.14}$$

with equality if and only if $n = 1$ or the state $\widetilde{\rho}$ prepared by the path integral on $M_{\partial X}$ is pure. Moreover, the monotonicity property of $\operatorname{Tr}\widetilde{\rho}^n$ discussed at the end of section 4.2 implies that for all $n \in \mathbb{Z}^+$,

$$\widetilde{Z}_\circ[M_{\partial X}^{(n+1)}] \leq \widetilde{Z}_\circ[M_{\partial X}^{(n)}]\,\widetilde{Z}[M], \tag{4.15}$$

which leads to eq. (4.14) as a special case. Of course, this phenomenon relies on the fact that $\widetilde{I}_k$ is a strongly non-local action that can sustain quantum wormholes. In contrast, even a non-local theory that is only weakly so would break into separate integrals over X_k^- and X_k^+, thereby restoring factorization. Hence, once again we see that quantum wormholes are crucial for the exact QET replica calculation to give a sensible answer.

4.4 Ensembles of replicas

The quantum wormholes giving non-factorizing replicas came from the path integral over Φ fields. It will be illustrative to restore these fields by going back to the complete theory in order to understand precisely how the non-factorization in the exact QET occurs. We can do this by expressing the QET action in eq. (4.13) back in terms of the path integral that defines it using eq. (2.2). This leads to

$$\widetilde{Z}_\circ[M_{\partial X}^{(n)}] = \int\limits_{\bigsqcup_{k=1}^n M_{k|k+1}} \mathcal{D}\varphi \int\limits_{\bigsqcup_{k=1}^n M_k} \mathcal{D}\Phi\, e^{-I[\varphi,\Phi]}, \tag{4.16}$$

where we have used locality of the complete theory to combine the path integrals that define the QET on each M_k into a single path integral over M^n. From eq. (4.16) we see that the path integrals over Φ and φ are perfectly out of phase. This is what causes the strong non-localities arising from the path integral over the former to lead to quantum wormholes for the latter.

Going back to the general case in which we have a hypersurface Σ which is not necessarily splitting, eqs. (4.13) and (4.16) lead to the following explicit form for eq. (4.9):

$$\operatorname{Tr}\widetilde{\rho}^n = \frac{1}{\widetilde{Z}^n} \int\limits_{M_\Sigma^{(n)}} \mathcal{D}\varphi \prod_{k=1}^n e^{-\widetilde{I}_k[\varphi]} = \frac{1}{\widetilde{Z}^n} \int\limits_{M_\Sigma^{(n)}} \mathcal{D}\varphi \int\limits_{M^n} \mathcal{D}\Phi\, e^{-I[\varphi,\Phi]}. \tag{4.17}$$

The difference between eqs. (4.6) and (4.17) is here manifest: in the latter, the Φ fields are only allowed to propagate on disjoint copies of M instead of on the replica manifold $M_\Sigma^{(n)}$. In other words, when calculating $\operatorname{Tr}\widetilde{\rho}^n$, the path integral first traces out the Φ fields to define a QET on each replica of M, and then the calculation on the replica manifold

$M_\Sigma^{(n)}$ is done for the only dynamical fields remaining, which are the φ fields. In contrast, when calculating $\operatorname{Tr}\rho^n$, the full theory was path-integrated on $M_\Sigma^{(n)}$, allowing both φ and Φ fields to propagate across the replica manifold.

The lack of factorization of $\operatorname{Tr}\widetilde{\rho}^n$ also admits an interpretation in terms of the averages over ensembles discussed in sections 3.4 and 3.5. In particular, consider expanding $\widetilde{\rho}$ in terms of the $\widetilde{\rho}_*[\alpha]$ states from eq. (3.31). Using orthogonality, the state replicas for $\widetilde{\rho}$ can be easily seen to reduce to

$$\operatorname{Tr}\widetilde{\rho}^n = \int \mathcal{D}\alpha\, P_*[\alpha]^n \, \operatorname{Tr}\widetilde{\rho}_*[\alpha]^n, \tag{4.18}$$

where the trace term is unity if $\widetilde{\rho}_*[\alpha]$ is pure. It is straightforward to apply the replica trick in eq. (4.3) to eq. (4.18), obtaining

$$S(\widetilde{\rho}) = H[P_*] + \langle S(\widetilde{\rho}_*[\alpha])\rangle_{P_*}\,, \tag{4.19}$$

where we have defined the Shannon entropy H of a classical probability distribution,

$$H[P_*] \equiv -\int \mathcal{D}\alpha\, P_*[\alpha]\, \log P_*[\alpha]. \tag{4.20}$$

This is of course what one would obtain by simply applying eq. (4.1) to eq. (3.24). This result says how for the state $\widetilde{\rho}$, as a classical mixture of quantum states, the von Neumann entropy decomposes into the Shannon entropy $H[P_*]$ of the classical ensemble, plus the average von Neumann entropy of the $\widetilde{\rho}_*[\alpha]$ states across the ensemble. The former accounts for purely classical uncertainty in the state $\widetilde{\rho}$ due to ignorance of the state of Φ fields, whereas the latter captures genuinely quantum correlations intrinsic to the φ fields within each member of the ensemble. Recall in particular that for a sufficiently large Hilbert space $\mathcal{H}_\Phi$, the $\widetilde{\rho}_*[\alpha]$ may be pure even for a non-splitting Σ. Hence we expect the classical term in eq. (4.19) to be extensive in $\mathcal{H}_\Phi$, and the quantum term to only be nontrivial if the $\widetilde{\rho}_*[\alpha]$ states fail to act with unit rank on the Hilbert space $\mathcal{H}_\varphi$ of the exact QET.

We can also express the state replicas for $\widetilde{\rho}$ in terms of traces of the spanning family of projected states defined in eq. (3.19). However, this family is not orthogonal on $\mathcal{H}_\varphi$ and consists of pure states only if Σ is splitting. The path integrals can still be pulled out by linearity, but the resulting expression does not simplify as eq. (4.18). We can write it as

$$\operatorname{Tr}\widetilde{\rho}^n = \left\langle \operatorname{Tr}\left(\prod_{k=1}^{n} \widetilde{\rho}[\Phi_{0k}]\right)\right\rangle_{P^n}. \tag{4.21}$$

The notation for the ensemble average $\langle\,\cdot\,\rangle_{P^n}$ in eq. (4.21) instructs us to introduce a copy of the measure $P[\Phi_{0k}]$ for every replica, i.e.,

$$\langle\,\cdot\,\rangle_{P^n} \equiv \int_{\Sigma^n} \prod_{k=1}^{n} \left(\mathcal{D}\Phi_0\, P[\Phi_{0k}]\right)(\,\cdot\,), \tag{4.22}$$

which is correctly normalized to $\langle 1\rangle_{P^n} = 1$ for any n. The von Neumann entropy of $\widetilde{\rho}$ one gets now reads

$$S(\widetilde{\rho}) = H[P] + \langle S(\widetilde{\rho}[\Phi_0])\rangle_P\,, \tag{4.23}$$

which has a structure similar to eq. (4.19). Of course both equations give the same total entropy for $\widetilde{\rho}$. Here, however, what each term corresponds to is qualitatively different. The classical term in eq. (4.23) accounts for correlations of the Φ fields on Σ with every other field, whereas the quantum term quantifies spatial correlations between φ fields on Σ and any other field outside of Σ. Hence, if Σ is splitting, the former captures correlations between φ and Φ fields, and the latter vanishes. In such a case it is clear that $H[P]$ will scale with the interaction term in eq. (2.5). In particular, were one to introduce a dimensionless parameter κ in front of I_{int}, one would find that $H[P] \propto \kappa$, with a vanishing entropy as $\kappa \to 0$. In other words, the ensemble trivializes as the interactions between φ and Φ fields are turned off.

One may expect that, even if the projected states $\widetilde{\rho}[\Phi_0]$ are not trace-orthogonal, the trace on the right-hand side of eq. (4.21) may still be dominated by the diagonal term where all states are equal across replicas. This is plausible if there are not many null states among projected states like in eq. (3.30), and also motivated by the fact that such configurations respect the replica symmetry that the path integral manifestly satisfies. With this expectation, one may write

$$\operatorname{Tr} \widetilde{\rho}^{\,n} = \int\limits_{\Sigma} \mathcal{D}\Phi_0 \, P[\Phi_0]^n \, \operatorname{Tr} \widetilde{\rho}[\Phi_0]^n + O(\mathrm{RSB}_{\,\Phi|_\Sigma}) \geq \langle \operatorname{Tr} \widetilde{\rho}[\Phi_0]^n \rangle_{P^n} \,, \tag{4.24}$$

where the corrections correspond to replica-symmetry breaking (RSB) configurations of Φ across the different Σ cuts, and are guaranteed to be nonnegative. On the right-most expression, note that we are using the notation introduced in eq. (4.24), but with the understanding that, because every replica has the same configuration $\Phi_{0k} = \Phi_0$, there is just a single integral over Φ_0 and every copy of P is evaluated on this Φ_0. For $n > 1$, the trace over projected states in eq. (4.24) gives unity if and only if Σ is splitting, and for $n = 1$ eq. (4.24) becomes a trivial equality. Using eqs. (3.22) and (4.9), the analogous expression to eq. (4.24) that holds at the level of the exact QET partition function for a splitting Σ is

$$\widetilde{Z}_{\circ}[M_\Sigma^{(n)}] = \langle Z_{\Phi_0}[M]^n \rangle_{P^n} + O(\mathrm{RSB}_{\,\Phi|_\Sigma}) \geq \langle Z_{\Phi_0}[M]^n \rangle_{P^n} \,, \tag{4.25}$$

where we are using the partition functions of projected theories from eq. (3.35), and the fact that $\widetilde{\rho}[\Phi_0]$ is pure for Σ splitting. This says that the replica calculation for the exact QET partition function on disconnected replicas does not factorize, but is approximated and lower bounded by an ensemble average over partition functions of theories which do factorize across replicas. In section 6.1.3 we relate eq. (4.25) to the factorization problem in holography.

In fact, the right-hand side of eq. (4.24) does provide an exact expression for the replicas of a different object: the classical ensemble Ω of pure states of the complete theory defined in eq. (3.27). This is because the lifted states $\Omega[\Phi_0]$ from eq. (3.26), which act on the full Hilbert space $\mathcal{H}$ of the complete theory, are actually orthogonal on the Φ sector. As a result, the replica calculation for this ensemble yields

$$\operatorname{Tr} \Omega^n = \langle \operatorname{Tr} \widetilde{\rho}[\Phi_0]^n \rangle_{P^n} \,, \tag{4.26}$$

an equation that holds exactly due to the orthogonality property in eq. (3.10). The von Neumann entropy of the Ω still nicely reduces to that of $\widetilde{\rho}$, as expected by construction,

$$S(\Omega) = S(\widetilde{\rho}). \tag{4.27}$$

JHEP05(2025)024

These expressions provide precise relations between the exact QET state $\widetilde{\rho}$, and the statistics of the corresponding canonical ensemble Ω of states of the complete theory.

4.5 Multi-swap operators

In this section, we obtain alternative representations of the path integrals for state replicas by performing field space projections rather than identifications on the underlying manifolds. The projection operators we obtain can be understood as multi-swap operators. These operators realize an alternative way of thinking of state replicas as measurements performed on states prepared by different copies of a theory, rather than as identical copies of a state prepared by a single theory.[32] This exercise will also make the differences and similarities between the replica calculations in a complete theory and in an exact QET more transparent.

The path integrals that compute state replicas started life as preparations of states on independent copies of the cut-open space M_Σ. In previous sections we implemented field identifications across the Σ_k cuts of the different replicas by doing the path integral on a space constructed by gluing copies of M_Σ appropriately. This gave rise to the replica manifold $M_\Sigma^{(n)}$, and also to the disjoint union M^n. One can reproduce these identifications in terms of projection operators in field space instead. In particular, here we will keep the path integrals just running over the disjoint union M_Σ^n of n independent copies of M_Σ. Because these path integrals are completely independent of each other, one may think of them as n copies of the same theory, each independently preparing a single copy of the pertinent state on a different M_{Σ_k}. In the cyclic trace, these copies then get related by a projection operator which can be interpreted as performing a measurement across all replicas (cf. a multi-swap test).

Comparing eqs. (4.6) and (4.17), we see that at the level of the path integral for the φ fields these two calculations are identical. What actually distinguishes the replica path integrals for complete and exact QET states is fully captured by what happens to the Φ fields. The focus here will thus be on understanding the differences between $\operatorname{Tr}\rho^n$ and $\operatorname{Tr}\widetilde{\rho}^n$ in terms of projection operators on the Φ fields.

For $\operatorname{Tr}\widetilde{\rho}^n$, the space M^n corresponds to simply gluing back each M_{Σ_k} into M_k along its own Σ_k cut. The identifications this implements on Φ can be captured by

$$\mathbf{P}_{\Pi_n}[\Phi] \equiv \prod_{k=1}^{n} \delta\left(\Phi|_{\Sigma_k^-} - \Phi|_{\Sigma_k^+}\right). \tag{4.28}$$

Explicitly, what this means is that the Φ path integral in eq. (4.17) can equivalently be written as the following path integral on M_Σ^n:

$$\int_{M^n} \mathcal{D}\Phi\,(\,\cdot\,) = \int_{M_\Sigma^n} \mathcal{D}\Phi\, \mathbf{P}_{\Pi_n}[\Phi]\,(\,\cdot\,). \tag{4.29}$$

On the other hand, for $\operatorname{Tr}\rho^n$, the field identifications that the replica manifold $M_\Sigma^{(n)}$ enforces can be implemented by

$$\mathbf{P}_{\Xi_n}[\Phi] \equiv \prod_{k=1}^{n} \delta\left(\Phi|_{\Sigma_k^+} - \Phi|_{\Sigma_{k+1}^-}\right), \tag{4.30}$$

[32]In quantum theory this distinction is operationally meaningful while mathematically innocuous. In quantum gravity, these two interpretations of state replicas can lead to different results [87, 88].

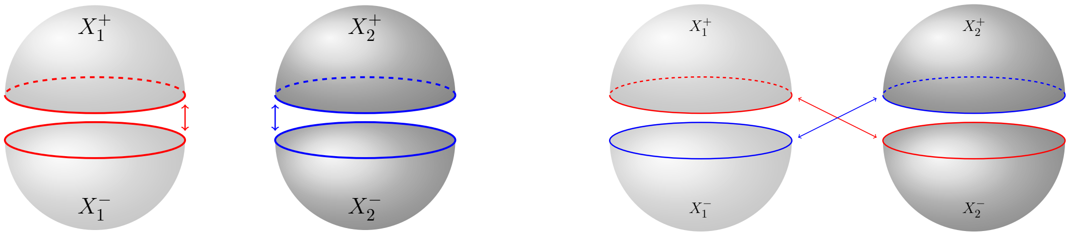

Figure 9. Illustration of the projection operators implementing field identifications in replica path integrals for $n = 2$. On the left, the operator $\mathbf{P}_{\Pi_n}$ defined in eq. (4.28) performs identifications of field configurations between the boundaries of dual half-spaces within each replica. On the right, the operator $\mathbf{P}_{\Xi_n}$ defined in eq. (4.30) performs these identifications between dual half-spaces of consecutive replicas. For $n = 2$, this corresponds to a standard swap operator. The combined projection operator $\mathbf{P}_{\mathbb{Z}_n}$ defined in eq. (4.33) would implement all identifications depicted above at once.

where $k + n \sim k$. The Φ path integral in eq. (4.6) can thus be reproduced on M_Σ^n by

$$\int_{M_\Sigma^{(n)}} \mathcal{D}\Phi\,(\,\cdot\,) = \int_{M_\Sigma^n} \mathcal{D}\Phi\,\mathbf{P}_{\Xi_n}[\Phi]\,(\,\cdot\,). \tag{4.31}$$

The operators defined in eqs. (4.28) and (4.30) are depicted in figure 9. Notice that for both eqs. (4.6) and (4.17) the projector in action for the φ fields is $\mathbf{P}_{\Xi_n}[\varphi]$, since these fields are dynamical in both the complete and the exact QET.

By inserting the projector $\mathbf{P}_{\Pi_n}$ into eq. (4.31) and the projector $\mathbf{P}_{\Xi_n}$ into eq. (4.29) one gets matching right-hand sides, which leads to the following identity for the left-hand sides:

$$\int_{M_\Sigma^{(n)}} \mathcal{D}\Phi\,\mathbf{P}_{\Pi_n}[\Phi]\,(\,\cdot\,) = \int_{M^n} \mathcal{D}\Phi\,\mathbf{P}_{\Xi_n}[\Phi]\,(\,\cdot\,). \tag{4.32}$$

This path integral provides a precise relation between eqs. (4.6) and (4.17). We can use this to better understand what makes the complete theory and the exact QET differ in calculations of state replicas. From eq. (4.32), we learn that in order to relate the path integrals on $M_\Sigma^{(n)}$ and M^n one has to project out configurations from both. Namely, there are field configurations on the replica manifold $M_\Sigma^{(n)}$ which need to be projected out by $\mathbf{P}_{\Pi_n}$, and also field configurations on the disjoint union M^n which need to be projected out by $\mathbf{P}_{\Xi_n}$. Put differently, the path integrals on $M_\Sigma^{(n)}$ and M^n respectively receive contributions which are generally not accounted for by the other.

However, suppose the path integral over Φ fields admits a saddle point approximation. In both cases, the functional extremization problem over Φ takes place on M_Σ^n. The difference between extremizing fields on $M_\Sigma^{(n)}$ and M^n is which saddles are allowed to contribute, which depends on boundary conditions. If the extrema on M_Σ^n happen to automatically be consistent with the identifications that both $\mathbf{P}_{\Xi_n}$ and $\mathbf{P}_{\Pi_n}$ impose, then the saddle-point approximations to the path integrals on $M_\Sigma^{(n)}$ and M^n will naturally agree. Hence at the level of saddle points, the state replicas for $\widetilde{\rho}$ and ρ agree if the dominant saddles are not projected out by the joint action of $\mathbf{P}_{\Xi_n}$ and $\mathbf{P}_{\Pi_n}$. Using Σ_k to indistinctively refer to both

$\Sigma_k^\pm$ cut sides, this combined projection can be expressed as

$$\mathbf{P}_{\mathbb{Z}_n}[\Phi] \equiv \mathbf{P}_{\Xi_n}[\Phi]\,\mathbf{P}_{\Pi_n}[\Phi] = \int\limits_{\Sigma} \mathcal{D}\Phi_0 \prod_{k=1}^{n} \delta\left(\Phi|_{\Sigma_k} - \Phi_0\right), \tag{4.33}$$

i.e., their joint action collapses the configurations on all Σ_k cuts to be identical. Because $\mathbf{P}_{\mathbb{Z}_n}$ enforces the $\mathbb{Z}_n$ replica symmetry at the level of configurations on the Σ_k cuts, we refer to it as the *replica symmetry operator.*[33] In the path integrals that compute state replicas, even if off-shell, there are always field configurations which break replica symmetry, and are thus in the kernel of $\mathbf{P}_{\mathbb{Z}_n}$. All such configurations of the Φ fields will make replica calculations for the complete theory and the exact QET automatically differ. The path integral over Φ fields with the operator in eq. (4.33) inserted yields

$$\int\limits_{M_\Sigma^n} \mathcal{D}\Phi\,\mathbf{P}_{\mathbb{Z}_n}[\Phi]\,e^{-I[\varphi,\Phi]} = \int\limits_{\Sigma} \mathcal{D}\Phi_0 \prod_{k=1}^{n} P[\Phi_{0k}]\,e^{-I_{\Phi_{0k}}[\varphi]}\Big|_{\Phi_{0k}=\Phi_0}, \tag{4.34}$$

where the P measures were introduced in eq. (3.22), and $I_{\Phi_{0k}}$ is the action on the k^{th} replica of a projected theory, as defined in eq. (3.33). By additionally acting with the path integral over φ on $M_\Sigma^{(n)}$ to prepare the corresponding state replica we obtain

$$\int\limits_{M_\Sigma^{(n)}} \mathcal{D}\varphi \int\limits_{M_\Sigma^n} \mathcal{D}\Phi\,\mathbf{P}_{\mathbb{Z}_n}[\Phi]\,e^{-I[\varphi,\Phi]} = \frac{1}{\widetilde{Z}^n} \langle \operatorname{Tr} \widetilde{\rho}[\Phi_0]^n \rangle_{P^n} = \frac{1}{\widetilde{Z}^n} \operatorname{Tr} \Omega^n, \tag{4.35}$$

where we have used the notation from eq. (4.22), and in the last equality we have recognized the replica result for the Ω ensemble in eq. (4.26). One may interpret eq. (4.35) as giving the expectation value of the replica symmetry operator $\mathbf{P}_{\mathbb{Z}_n}[\Phi]$ in the complete theory. Hence contributions to $\operatorname{Tr}\rho^n$ and $\operatorname{Tr}\widetilde{\rho}^n$ which are common to both path integrals are given precisely by $\operatorname{Tr}\Omega^n$. Conversely, it is contributions to the Φ path integral in the kernel of $\mathbf{P}_{\mathbb{Z}_n}[\Phi]$ which make the state replicas for complete and exact QET states differ, as claimed above.

The focus in this section up to this point has been on understanding the difference between the complete theory and the exact QET when it comes to replica calculations, which as discussed at the beginning is fully captured by the difference in how the Φ fields are treated. There is however also an important observation to be made about replica symmetry and its breaking at the level of the φ path integrals. Using the operator $\mathbf{P}_{\Xi_n}$ from eq. (4.30), we can write the replica path integral in eq. (4.8) as

$$\widetilde{Z}_\circ[M_\Sigma^{(n)}] = \int\limits_{M_\Sigma^n} \mathcal{D}\varphi\,\mathbf{P}_{\Xi_n}[\varphi]\,e^{-\widetilde{I}_\circ[\varphi]}. \tag{4.36}$$

Consider now constructing a resolution of the identity by using the operator in eq. (4.28) and its negation,

$$\mathbb{I} \equiv \mathbf{P}_{\Pi_n} + \not{\mathbf{P}}_{\Pi_n}, \tag{4.37}$$

[33]This operator was already secretly in action when going from eq. (4.21) to eq. (4.24).

which can in fact be taken as a definition of $\not{\mathbf{P}}_{\Pi_n}$. Combined with $\mathbf{P}_{\Xi_n}$, using eq. (4.33) gives

$$\mathbf{P}_{\Xi_n} \equiv \mathbf{P}_{\mathbb{Z}_n} + \mathbf{P}_{\Xi_n} \, \not{\mathbf{P}}_{\Pi_n}. \tag{4.38}$$

This is simply saying that the multi-swap operator $\mathbf{P}_{\Xi_n}$ which implements the replica manifold identifications in eq. (4.36) can be decomposed into a projection onto replica-symmetric configurations captured by $\mathbf{P}_{\mathbb{Z}_n}$, and RSB ones captured by the other term. Note that here we are talking about replica symmetry at the level of field configurations on Σ only, so the path integral is free to run over all possible configurations away from the cut in both terms. Since $\mathbb{Z}_n$ is a symmetry of the path integral, it is often reasonable to assume that RSB contributions are subleading, so we may write eq. (4.36) as

$$\widetilde{Z}_\circ[M_\Sigma^{(n)}] = \int\limits_{M_\Sigma^n} \mathcal{D}\varphi \, \mathbf{P}_{\mathbb{Z}_n}[\varphi] \, e^{-\widetilde{I}_\circ[\varphi]} + O(\mathrm{RSB}_{\varphi|_\Sigma}). \tag{4.39}$$

Using eq. (4.33), for the replica-symmetric path integral we find the expression

$$\int\limits_{M_\Sigma^n} \mathcal{D}\varphi \, \mathbf{P}_{\mathbb{Z}_n}[\varphi] \, e^{-\widetilde{I}_\circ[\varphi]} = \int\limits_{\Sigma} \mathcal{D}\varphi_0 \, \widetilde{P}[\varphi_0]^n \, \widetilde{Z}_{\varphi_0}^n, \tag{4.40}$$

where we have defined a partition function $\widetilde{Z}_{\varphi_0}$ and a normalized distribution $\widetilde{P}$ by

$$\widetilde{Z}_{\varphi_0} \equiv \frac{\widetilde{Z}}{\widetilde{Z}_{\varphi_0}^*} \int\limits_{M}^{\varphi|_\Sigma = \varphi_0} \mathcal{D}\varphi \, e^{-\widetilde{I}[\varphi]}, \qquad \widetilde{P}[\varphi_0] \equiv \frac{\widetilde{Z}_{\varphi_0}^*}{\widetilde{Z}}, \tag{4.41}$$

and the constant $\widetilde{Z}_{\varphi_0}^*$ is given by

$$\widetilde{Z}_{\varphi_0}^* \equiv \int\limits_{M}^{\varphi|_\Sigma = \varphi_0} \mathcal{D}\varphi \, e^{-\widetilde{I}[\varphi]}. \tag{4.42}$$

The normalizations here have been picked so as to mimic the definition of the projected theories Z_{Φ_0} in eq. (3.35).[34] Once again, as explained below eq. (3.35) for those, here $\widetilde{Z}_{\varphi_0}^*$ is also just a constant to be evaluated as in eq. (4.42), whereas $\widetilde{Z}_{\varphi_0}$ is to be treated as a partition function which defines a path integral for the φ fields on M. Hence we obtain the following form for eq. (4.39) (cf. eq. (4.25) for the Φ fields):

$$\widetilde{Z}_\circ[M_\Sigma^{(n)}] = \left\langle \widetilde{Z}_{\varphi_0}[M]^n \right\rangle_{\widetilde{P}^n} + O(\mathrm{RSB}_{\varphi|_\Sigma}), \tag{4.43}$$

where we are using the notation from eq. (4.22), but for the new measure $\widetilde{P}$ and for configurations of φ_0 of φ. Applied to eq. (4.9), this says that up to RSB effects the state

[34] The reader might wonder what motivates the convoluted definitions in eqs. (4.41) and (4.42), since plugging these into eq. (4.40) just gives $(\widetilde{Z}_{\varphi_0}^*)^n$ for the integrand. The point is that partition functions are only meaningful up to normalization, and we are using this freedom to extract a normalized measure $\widetilde{P}$. This measure may seem arbitrary at this level, but what fixes it uniquely is that it is the measure that appears when relating observables computed by $\widetilde{Z}$ and $\widetilde{Z}_{\varphi_0}$, for which normalizations drop out. In other words, it is the measure $\widetilde{P}$ that picks out a preferred normalization for the partition functions with fixed φ_0 configurations.

replicas for $\widetilde{\rho}$ are given by an ensemble average over $\widetilde{Z}_{\varphi_0}$ partition functions. This is used in section 6.1.3 to offer a different perspective on the factorization problem in holography.

To conclude, let us quote the identities which allow one to interpret the operators introduced in this section as performing measurements on states prepared by different copies of a theory. The Hilbert space of n copies of the exact QET on each M_k replica is a tensor product $\mathcal{H}_\varphi^{\otimes^n}$, i.e., a multi-theory Hilbert space where each factor corresponds to one exact QET. The states prepared by these theories similarly separate as $\widetilde{\rho}^{\otimes^n}$. Then if Tr_n denotes the trace on the multi-theory Hilbert space, one has the multi-swap test identity

$$\mathrm{Tr}_n\left(\mathbf{P}^{\varphi}_{\Xi_n}\widetilde{\rho}^{\otimes^n}\right) = \mathrm{Tr}\,\widetilde{\rho}^{\,n}, \tag{4.44}$$

where the operator $\mathbf{P}^{\varphi}_{\Xi_n}$ is being used here to multi-swap the φ field states. From the point of view of eq. (4.44), the state replicas for $\widetilde{\rho}$ can be understood as expectation values of the multi-swap operator $\mathbf{P}^{\varphi}_{\Xi_n}$.

5 Operator algebras

The algebraic picture that follows helps clarify the qualitative difference among the various types of theories considered throughout. Namely, what distinguishes the algebra of operators of a complete theory (local), an exact QET (strongly non-local), an approximate QET (weakly non-local), and a truncated QET (local). For this algebraic detour we follow [89–92].

5.1 Algebraic duality

Let us begin by introducing some basic notions that we will need to analyse the algebraic structure of these theories. Since we are interested in understanding the properties of very different theories, no assumptions will be made about relations that do not simply follow axiomatically. We denote algebras by $\mathcal{A}$. The set of operators that commute with a subalgebra $\mathcal{A}$ is denoted by $\mathcal{A}'$ and called the commutant of $\mathcal{A}$. For a quantum theory on some space M, let S be a connected splitting hypersurface $S \subset M$. To any subspace $\Sigma \subseteq S$, consider associating an algebra of local operators supported on it, and write it as $\mathcal{A}(\Sigma)$. Doing this for the complementary subspace $\Sigma' \equiv S \setminus \Sigma$ one obtains the algebra $\mathcal{A}(\Sigma')$. In general, for such locally generated algebras the following causality relation holds:

$$\mathcal{A}(\Sigma) \subseteq (\mathcal{A}(\Sigma'))'. \tag{5.1}$$

Indeed, for a Lorentzian theory where S is spacelike, this is the statement of microcausality. The algebra associated to Σ is said to satisfy duality (or Haag duality for ball regions in vacuum [93, 94]) if

$$\mathcal{A}(\Sigma) = (\mathcal{A}(\Sigma'))'. \tag{5.2}$$

This relation does not always holds, as will be important for us. In particular, duality can be violated by making Σ disjoint or non-simply connected [90]. It will be simpler to investigate the former case in which Σ consists of more than one connected component.

With this purpose, consider two subspaces $\Sigma_1, \Sigma_2 \subseteq S$. If one is contained in the other, their algebras obey isotony,

$$\mathcal{A}(\Sigma_1) \subseteq \mathcal{A}(\Sigma_2), \qquad \Sigma_1 \subseteq \Sigma_2. \tag{5.3}$$

If instead they are disjoint, let $\Sigma \equiv \Sigma_1 \cup \Sigma_2 \subseteq S$. The smallest algebra one can possibly associate to Σ is the one generated additively by the local operators in the algebras $\mathcal{A}(\Sigma_1)$ and $\mathcal{A}(\Sigma_2)$. This defines the minimal additively generated algebra

$$\mathcal{A}_{\text{min}}(\Sigma) \equiv \mathcal{A}(\Sigma_1) \vee \mathcal{A}(\Sigma_2), \tag{5.4}$$

where $\mathcal{A}_1 \vee \mathcal{A}_2 \equiv (\mathcal{A}_1 \cup \mathcal{A}_2)''$. If $\mathcal{A}$ is a von Neumann algebra, then $\mathcal{A}'' = \mathcal{A}$ by the bicommutant theorem [92]. The notation $\mathcal{A}(\Sigma)$ used up to this point implicitly refers to $\mathcal{A}_{\text{min}}(\Sigma)$; hereon, only the latter, explicit notation will be used when referring to locally generated algebras of Σ subspaces. This is important because there generally is no canonical algebra one may associate to a subspace. In particular, there is also a largest possible algebra one may associated to Σ in a way that is still consistent with the causality relation in eq. (5.1). This corresponds to taking the set of all operators which commute with those additively generated in the complementary subspace. More explicitly, this maximal algebra is given by the commutant of the additively generated algebra of the complement,

$$\mathcal{A}_{\text{max}}(\Sigma) \equiv (\mathcal{A}_{\text{min}}(\Sigma'))\,'. \tag{5.5}$$

The causality relation in eq. (5.1) becomes the general statement that

$$\mathcal{A}_{\text{min}}(\Sigma) \subseteq \mathcal{A}_{\text{max}}(\Sigma). \tag{5.6}$$

These definitions allow one to make concrete statements about the algebraic structure of a theory without having to commit to any choice of association of algebras to subspaces. This in particular means we can extend the notion of duality in eq. (5.2) to a general statement about the theory. The algebra of the theory is said to satisfy duality if for any subspace Σ, minimal and maximal algebras agree, i.e.,

$$\mathcal{A}_{\text{min}}(\Sigma) = \mathcal{A}_{\text{max}}(\Sigma), \qquad \forall \Sigma \subseteq S. \tag{5.7}$$

In [89–91], theories are characterized as complete if this notion of duality holds, and non-complete otherwise. The complete, local theories studied throughout this paper are assumed to be complete in this sense. Because violations of duality for effective theories will be important for us, let us briefly comment on both their cause and consequences.

A violation of duality takes the form of the proper inclusion $\mathcal{A}_{\text{min}}(\Sigma) \subset \mathcal{A}_{\text{max}}(\Sigma)$ for some subspace Σ. In words, this says that there are operators in the maximal algebra one could associate to Σ which cannot be locally constructed within Σ. Put differently, the mismatch between $\mathcal{A}_{\text{min}}$ and $\mathcal{A}_{\text{max}}$ is solely caused by the existence of non-local operators in the latter which are absent in the former. In the context of this paper, this is obviously why duality violations are relevant. The algebraic consequences of the failure of duality are noteworthy too. Assuming von Neumann's bicommutant theorem applies, one can easily see that if Σ violates duality, so will its complement:

$$\mathcal{A}_{\text{max}}(\Sigma') = (\mathcal{A}_{\text{min}}(\Sigma))' \supset (\mathcal{A}_{\text{max}}(\Sigma))' = \mathcal{A}_{\text{min}}(\Sigma'). \tag{5.8}$$

This in particular says that $\mathcal{A}_{\max}(\Sigma')$ contains operators outside the commutant of $(\mathcal{A}_{\max}(\Sigma))'$. In other words, there are operators in $\mathcal{A}_{\max}(\Sigma')$ which do not commute with some operators in $\mathcal{A}_{\max}(\Sigma)$, and vice versa. As a result, it follows that the expectation values of such operators cannot be locally determined, but depend non-locally on each other. As understood in [91], these are related by generalized symmetries. In our context, it is the explicit existence of non-local couplings which will relate those expectation values.

5.2 Effective theory algebras

Equipped with this machinery, it is now possible to make precise distinctions between the structural properties of the algebras of operators in our theories of interest. In each theory, we would like to characterize the kind of subspaces on which duality is violated, if any. On one end we have some complete, local theory, which as mentioned above is taken to respect duality. On the opposite end we have a truncated QET, which by virtue of being again local may also obey duality, although this is not necessarily always true. The intermediate cases of an exact QET and an approximate QET do violate duality in general and are thus more interesting. Given that non-localities are strictly stronger in an exact than in an approximate QET, so do we expect violations of duality to be strictly stronger in the former than in the latter. Let us introduce the minimal structure necessary to qualitatively distinguish these two on the same setting.

As we saw in section 2.4, partition functions factorize across connected components of a space even for an exact QET. Hence let our theory of interest be defined on a connected space $M \cong X \cup X^*$, where X and X^* are the usual CRT-dual half-spaces. The path integrals in section 3 allowed us to construct a Hilbert space for a given theory by preparing states and defining an inner product on ∂X. We would like to understand the algebras of operators acting on this Hilbert space and, in particular, their structure at the level of subspaces $\Sigma \subseteq \partial X$. As it turns out, it suffices to consider a connected boundary ∂X, and to make Σ consist of just two connected components, Σ_1 and Σ_2. This setting is illustrated on the left in figure 10.

Consider first an approximate QET, which recall is a weakly non-local theory. These theories are characterized by actions with arbitrarily high derivative orders, which give rise to non-local interactions within connected components. By virtue of being connected in X, the subspaces Σ_1 and Σ_2 will correspondingly be non-locally coupled. Hence we expect there to be non-local operators across these two subspaces with nontrivial expectation values. This would imply a violation of duality on Σ. This type of violation is characteristic of an approximate QET, but would of course also happen for an exact QET. A failure of duality that occurs for an exact QET but not for an approximate one must take place at a more global level.

An exact QET is a strongly non-local theory, the characteristic non-localities of which are associated to quantum wormholes. Recall that quantum wormholes correspond to non-local couplings across connected components, which are in particular lost in an approximate QET. Remarkably, this suggests that we should expect the violation of duality characteristic of an exact QET to occur even on ∂X, which is a full connected component. This is surprising because ∂X is the full space where the path integral defines the Hilbert space of the theory. In particular, ∂X has an empty complement, $\partial X' = \varnothing$, and the empty set has no local operators, so those in its commutant which define $\mathcal{A}_{\max}(\partial X)$ must genuinely give the full algebra of the

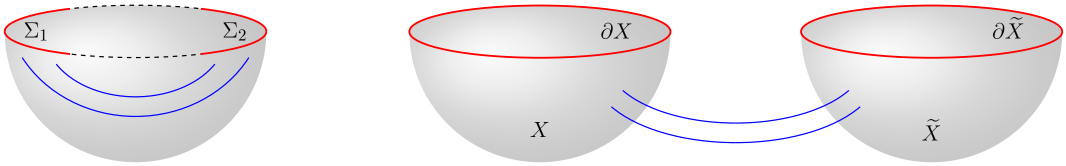

Figure 10. Illustration of the settings consider for understanding the violations of duality in QET operator algebras. On the left, the maximal algebra for $\Sigma = \Sigma_1 \cup \Sigma_2$ includes non-local operators straddling between Σ_1 and Σ_2 which are not in the minimal algebra. This gives the characteristic violation of duality in an approximate QET due to weak non-locality. On the right, the maximal algebra for the full space ∂X includes non-local operators which cannot be locally generated within the theory on ∂X. These are associated to quantum wormholes which are more naturally understood as connecting the two half-spaces which prepare a canonical purification of the state on ∂X. These globally non-local operators give the characteristic violation of duality in an exact QET due to strong non-locality and, in particular, the existence of quantum wormholes.

theory. Hence we see that a violation of duality on ∂X would imply that the full algebra of the theory includes operators which cannot be locally generated on the full ∂X space.

A more concrete way of understanding how a global violation of duality on ∂X can possibly happen is to remember that X is just one half of $M \cong X \cup X^*$, the space where the exact QET is originally defined. Once the path integral is cut open, quantum wormholes straddle between X and X^*, and the density operators prepared on ∂X are thus mixed. But if one considers CRT-conjugating X^* back to another copy $\widetilde{X}$ of X, then what the corresponding path integral would be preparing are canonical purifications of those density operators. This is illustrated in figure 10. From this viewpoint, $\partial\widetilde{X}$ can be thought of as the complementary subspace to ∂X, a more familiar setting when discussing algebras associated to subspaces. Then quantum wormholes are non-local interactions between ∂X and $\partial\widetilde{X}$, which would give rise to nontrivial expectation values between non-local operators associated to $\partial X \sqcup \partial\widetilde{X}$. This identification allows for a more explicit realization of the maximal algebra of the exact QET on ∂X as

$$\mathcal{A}_{\max}(\partial X) = (\mathcal{A}_{\min}(\partial\widetilde{X}))' \tag{5.9}$$

The point of view of canonical purifications used above also serves to demystify the global violation of duality of an exact QET. Namely, similarly to how for an approximate QET the maximal algebra of a subspace Σ is given by the commutant of the local algebra of its purifying subspace Σ', for an exact QET so is the maximal algebra of the theory given by the commutant of the local algebra of a purifying theory on $\partial\widetilde{X}$.

Intuitively, an approximate QET reveals its non-completeness when spatial regions are traced out, but is otherwise globally complete in the sense that duality holds on splitting hypersurfaces. In contrast, an exact QET reveals its non-completeness already at the global level of splitting hypersurfaces. This is because an exact QET still remembers that it came from a complete theory by tracing out some sector. Because this global type of non-completeness corresponds to a violation of duality at the level of the full theory, one may say that the non-local operators in $\mathcal{A}_{\max}$ which are not in $\mathcal{A}_{\min}$ are actually not part of the theory at all. However, we already saw quantum wormholes in action in section 4 giving nontrivial

JHEP05(2025)024

correlations across replicas which would result in nontrivial expectation values for non-local operators across replicas. The crux to reconcile these observations is to embrace the idea that the global violation of duality of an exact QET is in a specific sense not associated to non-local operators on a single copy of the theory, but to non-local operators on multiple copies of it. This perspective was made precise in section 4.5 by interpreting state replicas as multi-swap measurements. As usual, this emphasizes how global the properties that replica calculations probe are.

Let us conclude with an interesting observation about the global algebras of our theories. For the approximate QET, the absence of quantum wormholes guarantees duality on splitting hypersurfaces. For the exact QET, it is solely quantum wormholes which cause a violation of duality. But these strong non-localities associated to quantum wormholes are precisely what distinguish an exact QET from an approximate QET. This is particularly manifest if one decides to treat an exact QET as if it were a complete theory, which as explained in sections 2.3 and 3.3 is secretly equivalent to operating with an approximate QET (cf. a resummed derivative expansion). This leads to the expectation that the locally generated algebras of both theories should agree on splitting hypersurfaces. Using duality of the approximate QET, this means

$$\mathcal{A}_{\text{max}}(\partial X)^{\text{approx}} \cong \mathcal{A}_{\text{min}}(\partial X)^{\text{exact}}, \tag{5.10}$$

i.e., the full operator algebra of the approximate QET should be isomorphic to the full locally generated operator algebra of the exact QET. One may also generalize this to a relation involving the maximal algebra of the approximate QET for an arbitrary subspace Σ. In the exact QET, this would be reproduced by an intermediate algebra consisting of all operators in $\mathcal{A}_{\text{max}}(\Sigma)$ which are not in the commutant of $\mathcal{A}_{\text{min}}(\partial X)$.

5.3 Breakdown of effective theory

In a complete theory, minimal and maximal algebras for any subspace agree. This way, duality grants a canonical association of an operator algebra to any subspace, and also guarantees that this canonical algebra is always locally generated. In a non-complete theory, the failure of duality makes such an association of algebras to subspaces ambiguous. If one associates $\mathcal{A}_{\text{min}}(\Sigma)$ to Σ, non-local operators will be missed. If instead one associates $\mathcal{A}_{\text{max}}(\Sigma)$, non-local operators will be accounted for, but the algebras of complementary regions will contain operators which do not commute. An operational way of distinguishing the physics these two algebras capture when acting on states is thus desirable. When addressing our different theories, this will provide a diagnostic of which physics they do or do not capture.

Given some theory, consider the preparation of a state on a subspace Σ. A failure of duality on Σ signals an incompleteness of the description of that state that is attainable solely by local observables on Σ. In particular, this means that there will be distinct states the theory can prepare on Σ which will be indistinguishable from the perspective of the algebra of local operators on Σ. Indistinguishable states on Σ will be those differing only by the expectation values of non-local operators in $\mathcal{A}_{\text{max}}(\Sigma)$ which are not in $\mathcal{A}_{\text{min}}(\Sigma)$.

This conditional expectation can be manufactured by partially tracing the state over non-local sectors (thereby removing non-local information) and then lifting it back to an

JHEP05(2025)024

operator in $\mathcal{A}_{\text{max}}(\Sigma)$ with a maximally mixed identity operator in the non-local sector. The resulting state thus retains all information about expectation values of local observables, but is completely ignorant of non-local ones and yields zero for these. Put differently, the two states give identical expectation values for operators in $\mathcal{A}_{\text{min}}(\Sigma)$, and may only differ for those only in $\mathcal{A}_{\text{max}}(\Sigma)$. For instance, given a state on $\Sigma \equiv \Sigma_1 \sqcup \Sigma_2$ for the setup on the left of figure 10, one could consider constructing the reduced density operators $\widetilde{\rho}_1$ and $\widetilde{\rho}_2$ on each subspace, and then lifting these back to the product state on Σ given by $\widetilde{\rho}_1 \otimes \widetilde{\rho}_2$.

More generally, let the state on Σ be $\widetilde{\rho}$, and denote the state conditioned on having trivial expectation values with respect to non-local operators on Σ by $\widetilde{\rho}_{\text{loc}}$. With an information theoretic mindset, we would like to consider a situation in which $\widetilde{\rho}$ is the true state on Σ, and $\widetilde{\rho}_{\text{loc}}$ is a local observer's hypothesis for what the state on Σ is. We want to know the number N of copies of $\widetilde{\rho}$ we would need to perform measurements on in order to conclude that the hypothesis $\widetilde{\rho}_{\text{loc}}$ is wrong. This is quantified by requiring $NS(\widetilde{\rho}\,||\,\widetilde{\rho}_{\text{loc}}) \gg 1$, where $S(\widetilde{\rho}\,||\,\widetilde{\rho}_{\text{loc}})$ is the relative entropy of $\widetilde{\rho}$ with respect to $\widetilde{\rho}_{\text{loc}}$, defined by [95]

$$S(\widetilde{\rho}\,||\,\widetilde{\rho}_{\text{loc}}) = \operatorname{Tr}\widetilde{\rho}\,(\log\widetilde{\rho} - \log\widetilde{\rho}_{\text{loc}})\,. \tag{5.11}$$

The relative entropy is a nonnegative quantity by Klein's inequality, and vanishes if and only if the two states are identical. We emphasize again that this trace shall be taken in $\mathcal{A}_{\text{max}}(\Sigma)$; if the trace were taken in $\mathcal{A}_{\text{min}}(\Sigma)$, the relative entropy of $\widetilde{\rho}$ with respect to $\widetilde{\rho}_{\text{loc}}$ would be identically zero by construction. Equivalently, this means that different states $\widetilde{\rho}_1 \neq \widetilde{\rho}_2$ with the same conditional expectation $\widetilde{\rho}_{\text{loc}}$ are perfectly indistinguishable with only access to local operators in $\mathcal{A}_{\text{min}}(\Sigma)$. Indeed, the value of the relative entropy in eq. (5.11) quantifies the uncertainty of $\widetilde{\rho}$ in $\mathcal{A}_{\text{max}}(\Sigma)$ given the knowledge in $\mathcal{A}_{\text{min}}(\Sigma)$. Importantly, relative entropies are always well-defined across algebra types in quantum field theory.[35] If the conditional expectation that defines $\widetilde{\rho}_{\text{loc}}$ is trace-preserving, eq. (5.11) reduces to a simple difference of von Neumann entropies [89],

$$S(\widetilde{\rho}\,||\,\widetilde{\rho}_{\text{loc}}) = S(\widetilde{\rho}_{\text{loc}}) - S(\widetilde{\rho}), \tag{5.12}$$

an expression which now looks just like a conditional entropy. Here nonnegativity is easily seen to follow from the maximally mixed conditional expectation in $\widetilde{\rho}_{\text{loc}}$. Furthermore, if $\widetilde{\rho}$ is a state on $\Sigma \equiv \Sigma_1 \sqcup \Sigma_2$ that reduces to $\widetilde{\rho}_{\text{loc}} = \widetilde{\rho}_1 \otimes \widetilde{\rho}_2$, then eq. (5.12) becomes the mutual information between Σ_1 and Σ_2 on the state $\widetilde{\rho}$, i.e.,

$$S(\widetilde{\rho}\,||\,\widetilde{\rho}_{\text{loc}}) = I(\Sigma_1 : \Sigma_2)_{\widetilde{\rho}}. \tag{5.13}$$

The splitting case in which $\Sigma = \partial X$ is noteworthy. In particular, because in an exact QET duality holds globally, there would seem to be distinct states of the complete theory on Σ which are nonetheless perfectly indistinguishable by exact QET operators in $\mathcal{A}_{\text{min}}(\partial X)$. Remarkably, this would mean that there are states which operators on a single copy of the theory cannot possibly distinguish. To distinguish such states, one is necessarily forced to consider operators in multiple copies of the theory (as in eq. (5.9) for two copies), or quantities like the state replicas in section 4.

[35] This includes type-III von Neumann algebras, even though for these neither the traces nor the density operators in eq. (5.11) actually exist when Σ is not a splitting hypersurface.

This situation is demystified by thinking of the exact QET as preparing a canonical purification of $\widetilde{\rho}$ as on the right of figure 10. Then it is clear that $\mathcal{A}_{\text{min}}(\partial X)$ just has access to $\widetilde{\rho}$, the reduced density operator on ∂X, but not to any information about the purifying subsystem $\partial\widetilde{X}$. Any state prepared on $\partial X \sqcup \partial\widetilde{X}$ with operator insertions in $\widetilde{X}$ which do not change the state on ∂X will be perfectly indistinguishable by observables in $\mathcal{A}_{\text{min}}(\partial X)$. In this canonical purification picture, $\widetilde{\rho}_{\text{loc}}$ is the tensor product of the reduced density operators on ∂X and $\partial\widetilde{X}$. This state differs from the canonical purification of $\widetilde{\rho}$ in that quantum wormholes have been traced out, thereby decorrelating the two sides of the system. Correspondingly, eq. (5.13) quantifies quantum wormhole correlations, which are given by the mutual information between the state on ∂X and any purification. Because quantum wormholes are the remnant of correlations between φ and Φ fields in the complete theory, eq. (5.13) can also be thought of as measuring the amount of entanglement between these two sectors in pure states of the complete theory. Hence we expect the mutual information in such cases to scale with the interaction term between φ and Φ fields in the complete theory.

The discussion above has concerned distinguishability of states within a single theory. In fact, given how our QET constructions are related to each other, the relative entropy can also be used to distinguish between states prepared by different theories. This will in particular allow us to determine to what extend the states that an approximate or truncated QET prepare fail to capture the expectation values of complex operators of the true theory.

To be more precise, let the state of the complete theory on Σ be ρ. We are just interested in the state of the φ fields, which can be obtained by tracing out the Φ sector. By eq. (3.14) the resulting state is precisely $\widetilde{\rho}$, the exact QET state. Hence the exact QET perfectly describes the φ fields, and should be used as a benchmark. Now, suppose our best description of the φ fields is an approximate or truncated QET, which prepares the state $\rho_\times$ on Σ. Theoretical predictions made by computing observables on $\rho_\times$ are just approximations to the true results one would obtain by acting with those observables on the actual state $\widetilde{\rho}$ of the system. More precisely, we expect there to be a class of observables $\mathcal{O}_\times$ for which

$$\mathrm{Tr}(\rho_\times \mathcal{O}_\times) \approx \mathrm{Tr}(\widetilde{\rho}\,\mathcal{O}_\times) \tag{5.14}$$

with the approximation holding in some perturbative sense, and some different class of observables $\widetilde{\mathcal{O}}$ for which

$$\mathrm{Tr}(\rho_\times \widetilde{\mathcal{O}}) \neq \mathrm{Tr}(\widetilde{\rho}\,\widetilde{\mathcal{O}}), \tag{5.15}$$

with the two sides differing arbitrarily. For instance, suppose our observables of interest are correlation functions of a small number of insertions of the φ fields. If the lightest Φ field that was integrated out from the complete theory had a mass $m_\Phi > 0$, one may expect eq. (5.14) to hold so long as the operator insertions have energies $\Lambda \ll m_\Phi$, whether $\rho_\times$ comes from an approximate or a truncated QET. If instead one considers operators with energies comparable to or larger than m_Φ, the neglect of higher derivative terms will clearly make eq. (5.14) fail for a truncated QET. For an approximate QET, however, it is not obvious whether or not eq. (5.14) will hold. This is because all that an approximate QET is missing is quantum wormholes, but otherwise preserves all non-localities within connected components (see section 2.3). There is nonetheless a simple argument for why there must exist observables $\widetilde{\mathcal{O}}$ for which even an approximate QET must break down in the sense of eq. (5.15).

Consider the same logic that led us to introduce the relative entropy in eq. (5.11). Namely, if all we have access to is an approximate QET, our best hypothesis for the state on Σ is $\rho_\times$ as prepared by the approximate QET. To quantify how hard it is to realize that our hypothesis $\rho_\times$ for the state of the system is wrong, we consider the relative entropy

$$S(\widetilde{\rho}\,||\,\rho_\times) = \mathrm{Tr}\,\widetilde{\rho}\,(\log\widetilde{\rho} - \log\rho_\times)\,. \tag{5.16}$$

In principle, one may want to take the trace above in the exact QET algebra $\mathcal{A}_{\tilde{}}(\Sigma)$. However, since we are considering states prepared by a single copy of the theory, we can get the same answer by just taking the trace in $\mathcal{A}_{\tilde{}}(\Sigma)$. This is also desirable since, as argued at the end of section 5.2, we expect the $\mathcal{A}_{\tilde{}}(\Sigma)$ algebras of the exact and approximate QET to agree. The key observation to be made about eq. (5.16) is sharpest when Σ is a splitting hypersurface. Then, as explained in section 3.3, the state $\rho_\times$ is pure, while the state $\widetilde{\rho}$ is mixed due to quantum wormholes. As a result, the relative entropy in eq. (5.16) is in fact formally infinite. This is more easily seen in terms of the underlying probability density functions of each density operator, for which the relative entropy becomes a classical Kullback-Leibler, and the infinity corresponds to the support of $\widetilde{\rho}$ overlapping with the kernel of $\rho_\times$.

The divergence of eq. (5.16) has operational consequences. From the interpretation of the relative entropy above eq. (5.11), it means that a single measurement on a single copy of the system should suffice to distinguish $\widetilde{\rho}$ from $\rho_\times$. In other words, there must exist $\widetilde{\mathcal{O}}$ operators which cause the approximate QET to break down in the sense of eq. (5.15). As pointed out above, it is not clear that this breakdown is necessarily associated to energy scales. Rather, it must correspond to observables which have support on $\widetilde{\rho}$ but not on $\rho_\times$, such that their expectation values are zero for the latter but nontrivial for the former. Needless to say, these operators will also cause any truncated QET to break down.

The kind of observables which an approximate QET will miss correspond to operators which can tell apart a mixed from a pure state. These must in some sense be complex operators capable of probing the spectrum of the density operator. Intuitively, the more mixed the true state $\widetilde{\rho}$ is, the easier it should be to realize that the hypothesis state $\rho_\times$ is wrong. The mixing of $\widetilde{\rho}$ is caused by quantum wormholes, which arise from the interactions between φ and Φ fields in the complete theory. These quantum wormholes may ultimately be suppressed by the mass m_Φ of the lightest Φ field that was integrated out. As a result, the different QET states will be harder to distinguish the heavier m_Φ is. Hence one may expect any reasonable notion of operator complexity for the $\widetilde{O}$ observables that break the approximate QET to grow with m_Φ. À priori this does not imply that $\widetilde{O}$ itself must be heavy: it may be possible to scale the complexity of an operator that distinguishes states without scaling its total energy.[36]

Let us conclude by pointing out that the failure of approximate QET states to capture complex operators is akin to how no single projected state $\widetilde{\rho}[\Phi_0]$ in the ensemble that makes up $\widetilde{\rho}$ can possibly make predictions all consistent with $\widetilde{\rho}$ itself (see section 3.4). The true

[36]I thank Chris Akers and Netta Engelhardt for very insightful conversations on the breakdown of effective theory with complexity, and also Aidan Herderschee for highlighting the possibility of a separation of scales.

state of the system is $\widetilde{\rho}$, and its predictions carry classical uncertainty due to its ensemble nature. This classical uncertainty was in section 3.5 formulated in terms of $\widetilde{\rho}$ being prepared by an ensemble of theories, rather than any single one. Correspondingly, we expect a single approximate QET to fail to capture the true physics of the φ fields, but an ensemble thereof to potentially make consistent predictions. Such an ensemble would naturally be over couplings of the theory, since these ultimately come from integrating out genuinely quantum Φ fields.

6 Quantum gravity

There are two complementary but qualitatively distinct ways in which the lessons from the previous sections can also be applied to effective theories in the context of quantum gravity. Holographic duality offers an arena where the gravitational bulk can be understood as an emergent theory that ought to realize every expectation from a boundary field theory, including an exact QET. Alternatively, starting from a fully non-perturbative theory of quantum gravity, one might ask which features of the exact QET framework described here remain valid in a gravitational effective theory (GET). In section 6.1, we first explore the holographic perspective, followed by a discussion in section 6.2 on how a GET fundamentally differs from a QET.

6.1 Effective holography

Great progress has been made in understanding quantum gravity within the holographic paradigm by harvesting the connection between entanglement on the boundary and spacetime connectivity in the bulk [96, 97]. The notion that spacetime emerges from quantum correlations has led to significant advances in the holographic dictionary, particularly in understanding holographic entanglement [98–106] and bulk reconstruction [107–115].

6.1.1 Replica wormholes

An important development in recent years has in particular brought to the forefront of holography the importance of wormhole contributions in holographic calculations of entropic quantities [32–34, 116, 117]. In the context of black hole evaporation, we have come to understand how crucial it is to include connected wormhole topologies in the gravitational path integral in order to obtain a Page curve consistent with unitarity for the entropy of Hawking radiation [29–31, 118]. A specific setting where these wormholes make an appearance is in replica calculations in which the path integral is used to prepare the state of an evaporating black hole [32–34]. Explicitly, as quantum entanglement builds between the black hole and radiation, wormhole geometries begin to dominate replica calculations, connecting replicas even when they lie on disjoint spaces. Put differently, quantum correlations between replicas are encoded geometrically in the bulk as replica wormholes. In this section, we would like to understand these replica wormholes from the perspective of the boundary theory.[37]

Although the bulk naturally allows for a spatial splitting along Cauchy surfaces to distinguish black hole and radiation degrees of freedom, this distinction becomes less clear in

[37]A coarse-graining mechanism by which quantum wormholes reproduce the permutation structure of geometric replica wormholes was studied in detail by [119]. The relation between the quantum wormholes that arise in open quantum systems and those in holographic replica calculations was also explored by [52].

the dual boundary theory. To simplify matters, an auxiliary non-gravitating bath system is often introduced, coupling to both theories and making the degrees of freedom easier to separate. This way, in the boundary picture, the bath captures most radiation (early radiation only), whereas the genuinely holographic theory captures mostly the black hole (plus late radiation). The need for introducing such an artificial artifact, and nonetheless still getting a rather ambiguous splitting between the two sectors, emphasizes an important point: these two sectors are not naturally localized on any distinct spatial regions on the boundary, but rather simply correspond to distinct sectors of quantum fields of the theory.

The observations above translate naturally into the state preparation formalism used in this paper. In particular, we have argued that the state of the radiation on the boundary cannot simply be obtained by partially tracing some state of the complete theory over some spatial region. Rather, the state of the radiation corresponds to partially tracing over an internal sector of black hole states of the complete theory. As shown in section 3, what constructs such a state is precisely an exact QET.

The boundary theory we are interested in here is obtained by integrating out the black hole sector from the complete theory. The resulting exact QET holographically describes the radiation as a strongly non-local theory. The fact that the black hole evaporates is a manifestation of the non-unitarity that is characteristic of any exact QET due to correlations between system and environment. As explained in section 2.3, the non-locality and non-unitarity of an exact QET are not only consistent with unitarity and locality of the complete theory, but required. Indeed, in the exact QET the correlations between black hole and radiation are fully captured by quantum wormholes, and would be completely missed by any approximate QET. In particular, by the arguments in section 5.3, all possible radiation states the exact QET could prepare differing only by such correlations would be perfectly indistinguishable by an approximate QET. In the exact QET these states are distinguishable by sufficiently complex operators, but will require multiple copies of the state. In other words, the QET framework predicts the failure of any local effective theory to distinguish the state of Hawking radiation from a thermal state, and also the need for replicas where the inclusion of wormholes allows one to recover results consistent with unitarity of the complete theory.

Let us restate that in the above paragraph we are arguing for all of these phenomena purely from a boundary perspective. By holographic duality, these predictions must be reproduced by the bulk theory. Since more than one copy of the theory is needed to make meaningful statements about the state of the radiation, consider the replica path integrals of section 4. On the boundary, the quantum wormholes that couple the replicas come to life as soon as the black hole and radiation get coupled, i.e., as soon as the black hole begins to evaporate. Correspondingly, in the bulk one should expect wormholes connecting all replicas to exist as off-shell configurations of the gravitational path integral at all stages of black hole evaporation. Whether replica wormholes dominate or not is a different question that arises in a semiclassical limit. Namely, it refers to the competition between the limit that allows for saddle-point approximations, and the build-up of strongly non-local correlations to overcome local ones, such that the connected saddle dominates.

In line with the old and fruitful intuition that spacetime geometrizes entanglement, we once again are compelled to claim that geometric replica wormholes precisely capture the ways in which replicas are strongly non-locally coupled in any exact QET by quantum wormholes. The contrapositive of this statement is even more illuminating. As explained in section 3.3, one rarely works with an exact QET, and it is customary to reduce such a strongly non-local theory to a local one by performing a derivative expansion and truncating. This procedure may lead to a seemingly unitary theory (if unitarity bounds are respected), but which will make predictions strictly inconsistent with unitarity of the complete theory (cf. section 2.3). Intuitively, this is because a truncated QET is quite literally obtained by throwing away information about the complete theory. In addition, the destruction of quantum wormholes also means that replica path integrals in any such truncated QET cannot possibly couple disjoint spaces. In other words, the violation of unitarity in the complete theory comes hand in hand with the absence of quantum wormholes across replicas in an effective theory. The situation in the bulk side is entirely analogous: the moment one decides to not include contributions from replica wormholes, unitarity manifestly breaks down in the form of information loss during black hole evaporation.[38]

More generally, the relation above between wormholes and unitarity is a manifestation of an unavoidable connection between ensembles and unitarity. Namely, only an exact QET is consistent with unitarity of an underlying complete theory, and the way it is so is by producing such ensembles (see [52] for related results). Indeed, as explained in section 3.4, an exact QET generally has quantum wormholes which cause mixing in the preparation of states, and the resulting mixed states can be identified with ensembles of pure states of the complete theory. Equivalently, section 3.5 showed how these ensembles of states can also be interpreted as states prepared by an ensemble of complete theories. These two interchangeable pictures provide a precise mechanism for and are consistent with the interpretations of [8, 85, 86, 120–122].

6.1.2 Non-replica wormholes

What follows is a potentially polarizing observation, since it may seem as trivial to some as provocative to others. For quantum wormholes to arise, it was crucial that the way the replica manifold reassembles into disjoint copies of the original manifold is by combining half-spaces from consecutive replicas. This is fundamentally why the eventually disjoint spaces remain coupled: quantum wormholes within each original replica end up straddling across the reassembled copies in the form of quantum wormholes. This construction, which is specific to state replicas, differs fundamentally from simply placing a complete theory on multiple copies of a space and integrating out a sector to construct an exact QET. Indeed, were one to do this, the partition function of the resulting exact QET would just factorize across connected components, as shown in eq. (2.8). Hence the words 'performing a path integral on a space with multiple connected components' do not immediately imply factorization or lack thereof even in field theory. In particular, if the theory at hand is strongly non-local, it all depends on how the multi-integrals in the action behave, which is non-optionally determined by which calculation one is doing.[39]

[38] I thank Marija Tomašević for emphasizing these similarities.

[39] Alternatively, one may think there is a choice to include or not include quantum wormholes. But as we saw, their inclusion corresponds to a calculation in the exact QET, whereas ignoring them corresponds to using an approximate QET. It may be that the dichotomy as for whether one should include geometric wormholes in the gravitational path integral is analogous.

If geometric wormhole contributions to the gravitational path integral are understood as geometrizing quantum wormholes of the boundary theory, then the lesson would be: wormholes must be included in holographic replica calculations for an exact QET, but they must not be included when simply evaluating a partition function of a theory on a set of disjoint spaces. Correspondingly, replica calculations should not factorize and holographically they would not, whereas partition functions on disjoint spaces should factorize and holographically they would. Assuming this interpretation makes sense, using the gravitational path integral correctly would not lead to a factorization problem.

A potential counterargument arises from calculations in JT gravity and dual matrix ensembles, which appear to compute the partition function of a theory on disjoint spaces. In these cases, including wormhole contributions successfully reproduces the expected spectral statistics of black hole microstates. There are actually two reasons why this may be reconciled with the statements above. We offer a possible explanation for non-replica wormholes here, and a reason why these wormholes may secretly actually be replica wormholes in section 6.1.3.

The explanation is as follows. In holography, we often write $Z(M)$ to refer to the path integral of a field theory on a boundary manifold M, and then $\mathcal{P}(M^n)$ to refer to whatever the gravitational path integral computes with n copies of M as its boundary conditions. Upon the inclusion of wormholes we then obtain the statistical interpretation of non-factorization in $\mathcal{P}(M^n) \equiv \langle Z(M)^n \rangle$ as caused by averaging over complete theories. However, because we are treating gravity as an effective theory, we already see that a more natural interpretation would read $\mathcal{P}(M^n) \equiv \langle \widetilde{Z}(M)^n \rangle$, where the average is over exact QET partition functions. But then a simple explanation for what happens when we allow for wormholes to connect the n boundaries is that we are simply evaluating $\mathcal{P}(M^n) = \langle \widetilde{Z}(M^n) \rangle$, and allowing for quantum wormholes in the exact QET to straddle across distinct copies of M. In other words, we would be taking the exact QET partition function $\widetilde{Z}(M)$ and deciding to define it as a theory on all n boundaries by simply replacing $M \to M^n$ everywhere in the action. This would be very unnatural from the perspective of the exact QET, but would nonetheless give statistical correlations across copies of M which ultimately are nontrivial precisely because the exact QET does genuinely have quantum wormholes.

This way, $\widetilde{Z}(M)^n$ and $\widetilde{Z}(M^n)$ would already be different from the boundary perspective, and the distinction between them, though unphysical, would be clear. In the factorizing case of $\widetilde{Z}(M)^n$, the exact QET has strong non-localities only within connected components, whereas in the non-factorizing case of $\widetilde{Z}(M^n)$ these would extend to multiple components in the form of quantum wormholes. One could then speculate that this is also the mechanism at play for gravitational wormholes in the bulk. If so, to compute $\widetilde{Z}(M)^n$ one would be instructed to include all possible topologies anchoring to a single boundary, but to not allow for wormholes connecting multiple boundaries. Notice how this naturally gives single-boundary wormholes and multi-boundary wormholes a qualitatively different status, and a field theory argument for why one should include the former but not the latter in certain calculations. As for the non-factorizing case of $\widetilde{Z}(M^n)$, the artificial quantum wormholes would get geometrized by similarly artificial multi-boundary gravitational wormholes. Any statistical correlations across components of M^n would then seem to be capturing the physics of objects which do exist in the theory (quantum wormholes and gravitational wormholes), but in a setting where they would actually not arise physically.

6.1.3 Non-replica wormholes as replica wormholes

A typical calculation where geometric multi-boundary wormholes are believed to contribute involves providing the gravitational path integral with n copies of a thermal partition function Z_β as its boundary conditions [6, 26, 41–47]. The non-factorizing answer that wormholes lead to is statistically interpreted as corresponding to the calculation of an n^{th} moment of Z_β in some ensemble of theories, which we may denote by $\langle Z_\beta^n \rangle$. Explicitly, let M represent the manifold boundary condition corresponding to each Z_β. Then we identify

$$\mathcal{P}(M^n) \equiv \langle Z_\beta^n \rangle. \tag{6.1}$$

Since the theories on each copy of M are assumed to be independent, the average above provides a natural interpretation of the observed non-factorization. In fact, there are other quantities $\mathcal{P}(M^n)$ could be computing, with identical boundary conditions for the gravitational path integral, and consistent with having a single complete theory on the boundary.

The non-factorizing result $\langle Z_\beta^n \rangle$ when a factorizing Z_β^n answer was expected is just a paradigmatic example of a more general concern. Namely, if Z is the partition function of a complete theory, we expect a factorizing result, and not something like $\langle Z^n \rangle$. Actually, as explained in eq. (2.7), also the partition function $\widetilde{Z}$ of an exact QET must factorize. In what follows, we identify replica calculations which are perfectly consistent with both sides of eq. (6.1), but where the wormholes on the left are replica wormholes, and the average on the right is consistent with having a single, complete theory.

Suppose one has an exact QET on the boundary and considers preparing a state $\widetilde{\rho}$ on a splitting Σ. This is done by performing a path integral on the cut-open space M_Σ. To prepare n independent copies of the state, one can do n independent path integrals on the disjoint space M_Σ^n. Upon identifications under the trace, as explained in section 4, the replica path integral for $\operatorname{Tr}\widetilde{\rho}^n$ can be performed on a replica manifold $M_\Sigma^{(n)}$ which turns out to be isomorphic to M^n for Σ splitting. Despite being isomorphic to M^n, the resulting space connects different replicas as in eq. (4.11). Hence as we have already discussed multiple times the resulting calculation does not factorize due to quantum wormholes, and in addition admits an interpretation as an ensemble average over factorized objects like eq. (4.21). To make the connection to the factorization problem explicit, we would like to relate such an ensemble average to one over partition functions of different theories. We now give two general situations in which $\operatorname{Tr}\widetilde{\rho}^n$ can indeed be understood this way.

The first situation follows from eq. (4.25) and yields

$$\operatorname{Tr}\widetilde{\rho}^n = \frac{1}{Z^n}\,\langle Z_{\Phi_0}^n \rangle_{P^n} + O(\mathrm{RSB}_{\,\Phi|_\Sigma}), \tag{6.2}$$

where all partition functions are evaluated on M. Here, recall that Z_{Φ_0} are the partition functions of projected theories defined in eq. (3.35), and which do factorize in the usual sense, i.e., $Z_{\Phi_0}[M^n] = Z_{\Phi_0}[M]^n$. That the RSB corrections in eq. (6.2) are small relies on the assumption that the projected states from eq. (3.19) are approximately trace-orthogonal. This will generically be the case if the projection from the complete theory does not generate too many null states. More precisely, this assumes that $\dim \mathcal{H}_\Phi$ is not much larger than $\dim \mathcal{H}_\varphi$. Altogether, this means that the replica-symmetric approximation that the first term in eq. (6.2) gives should hold so long as the Hilbert space of states that was integrated out from the complete theory was not too large. Then, the statement of eq. (6.2) is that the replica calculation for an exact QET state takes precisely the form of an ensemble average over partition functions on M^n which do factorize.

The second situation in which this happens does not require assuming anything about the Hilbert space of the Φ fields in the complete theory. From eq. (4.43), we obtain

$$\mathrm{Tr}\,\widetilde{\rho}^n = \frac{1}{Z^n}\left\langle \widetilde{Z}^n_{\varphi_0} \right\rangle_{\widetilde{P}^n} + O(\mathrm{RSB}_{\varphi|_\Sigma}) \tag{6.3}$$

where all partition functions are again evaluated on M. The partition functions Z_{φ_0} were defined in eq. (4.41), and note that these also factorize due to the projection of φ to φ_0 on Σ. Now the RSB term in eq. (6.3) corresponds to φ field configurations across replicas that break replica symmetry on the Σ cuts. The assumption that his RSB term is small is the usual assumption of replica symmetry motivated by the cyclic $\mathbb{Z}_n$ symmetry of the overall path integral. Namely, one typically expects replica-symmetric configurations to dominate on the basis of symmetry arguments. This is generically a safe assumption for positive $n \in \mathbb{Z}^+$, particularly when the φ path integral admits a saddle-point approximation which respects the $\mathbb{Z}_n$ symmetry.[40] As a result, eq. (6.3) provides another situation in which a replica calculation takes the form of an ensemble average over theories.

Crucially, note that in both eqs. (6.2) and (6.3) we are considering replica calculations for a state $\widetilde{\rho}$ on a splitting Σ, such that $M^{(n)}_\Sigma \cong M^n$. In a bulk calculation, it would thus be natural to provide the gravitational path integral with M^n as its boundary conditions. The inclusion of geometric wormholes in this case makes sense, since from the boundary perspective there genuinely are quantum wormholes coupling the replicas. This provides an interpretation of $\mathcal{P}(M^n)$ with wormholes included as a bulk calculation of $\mathrm{Tr}\,\widetilde{\rho}^n$ (where factorization should not happen) rather than a calculation of $\widetilde{Z}^n$ (where factorization should happen).

Alternatively, the constructions above may be thought of as a version of the factorization puzzle which places the focus on the room for ambiguity that exists in how boundary conditions for the gravitational path integral are prescribed. Namely, one lesson from this exercise is that generally $\mathcal{P}(M^n)$ is not a well-defined object: simply feeding the manifold boundary M^n to the gravitational path integral is not enough to single out a specific quantity.[41]

Let us address the concrete case of thermal partition functions in eq. (6.1) more directly. Consider a complete theory defined on a space M with a U(1) symmetry around a non-contractible cycle of length β. Let Σ be a hypersurface splitting M into two $\mathbb{Z}_2$-symmetric dual halves X and X^*. Each such half-space contains a half-cycle, or segment, of length $\beta/2$. Hence ∂X consists of two disjoint boundary components, one at each end-point of the segment, and similarly for ∂X^*. We refer to these as left and right boundaries, respectively Σ_L and Σ_R, and write $\Sigma \equiv \Sigma_L \sqcup \Sigma_R$. The state that the exact path integral prepares on the cut-open space M_Σ is the density operator of the thermofield-double (TFD) state. Suppose, however, that we are not using the complete theory on the boundary, but an exact QET. This theory cannot possibly prepare a pure TFD. Instead, it would compute some generally

[40] Analytic continuations for replica tricks, specifically to $n < 1$, are a typical example of when RSB contributions become non-negligible, and in fact oftentimes dominant.

[41] A related construct which would also gives these ambiguities is the thermo-mixed double of [123, 124].

mixed state $\widetilde{\rho}_{\text{TFD}}$ obtained by partially tracing out whichever sector was integrated out from the complete theory. If we further trace out one side of the TFD, what we obtain is a similarly approximate thermal state $\widetilde{\rho}_\beta$ on the other side.

There are two inequivalent calculations one may perform with the states $\widetilde{\rho}_{\text{TFD}}$ and $\widetilde{\rho}_\beta$,

$$(\operatorname{Tr} \widetilde{\rho}_\beta)^n \neq \operatorname{Tr}(\widetilde{\rho}_{\text{TFD}}{}^n). \tag{6.4}$$

Both of these take the form of a boundary path integral on a space isomorphic to M^n, but only one of them factorizes. On the left, we have a multi-trace quantity giving $\widetilde{Z}(\beta)^n$ if $\widetilde{\rho}_\beta$ is left un-normalized, i.e., the exact QET computation of n copies of the thermal partition function. This calculation takes place on M^n and obviously factorizes since it would in particular be the same calculation in the complete theory. On the right, we have a single-trace replica calculation for the state $\widetilde{\rho}_{\text{TFD}}$ which obviously cannot factorize because the state is mixed. This calculation takes place on $M_\Sigma^{(n)} \cong M^n$. In other words, both calculations take place on the same manifold. It would thus seem that a gravitational path integral of the form $\mathcal{P}(M^n)$ cannot possibly tell the difference between the distinct quantities on the left and right of eq. (6.4). This is particularly worrisome because the discrepancy between these quantities is not merely quantitative, but qualitative: the one on the left factorizes, whereas the one on the right does not.

This exercise suggests that the gravitational wormholes for $\left\langle Z_\beta^n \right\rangle$ in eq. (6.1) may in fact be replica wormholes and actually computing $\operatorname{Tr}(\widetilde{\rho}_{\text{TFD}}{}^n)$ in eq. (6.4). In the concrete setting of JT gravity, one may then wonder what the dual matrix ensemble of Hamiltonians corresponds to. In other words, in what sense does the exact QET object $\operatorname{Tr}(\widetilde{\rho}_{\text{TFD}}{}^n)$ capture the spectral statistics of an ensemble of theories? The answer to this question follows from sections 3.4 and 3.5, and can be intuitively addressed by thinking of the exact QET as an open quantum system. Specifically, just as an open system evolves under an influence functional without a single-Hamiltonian description due to dissipation, the exact QET does not exhibit a deterministic spectrum. Instead, its spectral statistics are naturally reproduced by an ensemble of complete Hamiltonians consistent with the dynamics of the relevant subsector.[42]

6.1.4 Quantum error correction

The encoding and reconstruction of bulk operators in terms of boundary ones is now well understood in terms of operator algebra quantum error correction [111–115]. The QET framework naturally mirrors this structure using the operator algebras discussed in section 5. This connection can be most clearly illustrated by comparing the operator algebras of the approximate and exact QET following section 5.2, without needing to reference the complete theory.[43]

The relevant identifications to be made with the ingredients of quantum error correction are as follows. The physical theory is the exact QET, which has a Hilbert space $\mathcal{H}_{\text{TFD}}$ of states with quantum wormholes, and an algebra $\mathcal{A}_{\text{TFD}}$ which includes both locally generated operators and also non-local ones whose expectation values are nontrivial due to quantum

[42]This is currently work in progress [125].

[43]The exact QET already has all the structure needed to reconstruct operators of the approximate QET, and the additional information the complete theory contains is unnecessary. The extreme case of a complete theory where the sector that is integrated out does not interact with the sector of interest illustrates this.

wormholes. This algebra can thus be written

$$\mathcal{A}_{\text{exact}} = \mathcal{A}_{\text{loc}} \vee \mathcal{A}_{\text{worms}}. \tag{6.5}$$

The logical theory is the approximate QET, whose Hilbert space $\mathcal{H}_{\text{approx}}$ is isomorphic to a code subspace $\mathcal{H}_{\text{code}} \subset \mathcal{H}_{\text{exact}}$ of states where quantum wormholes are projected out. The algebra $\mathcal{A}_{\text{approx}}$ of the approximate QET is locally generated and, by eq. (5.10), is isomorphic to the local sector of $\mathcal{A}_{\text{exact}}$, i.e.,

$$\mathcal{A}_{\text{approx}} \cong \mathcal{A}_{\text{loc}}. \tag{6.6}$$

That $\mathcal{A}_{\text{approx}}$ is reconstructible from $\mathcal{A}_{\text{exact}}$ is thus obvious at a global level. The more interesting setting corresponds to reconstruction within an arbitrary subspace Σ. In this case the minimal algebras locally generated within Σ of both effective theories are still isomorphic, and so are the maximal algebras if the $\mathcal{A}_{\text{worms}}$ subalgebra of the exact QET is truncated away.[44] Reconstruction of operators in the minimal algebra for Σ in the approximate QET will be possible with access to operators in the minimal algebra for Σ in the exact QET, and similarly for maximal ones. However, unsurprisingly, it will not be possible to reconstruct non-local operators in the maximal subalgebra for Σ in the approximate QET with just access to the locally generated minimal algebra for Σ in the exact QET.

To see how this works, it is instructive to consider the standard theorem addressing the necessary and sufficient conditions for recoverability in operator algebra quantum error correction [111, 126, 127]. The relevant elements for error correction on Σ are an arbitrary logical operator $O \in \mathcal{A}_{\text{loc}}$ encoded to act only on $\mathcal{H}_{\text{code}}$, and physical operators in $\mathcal{A}_{\text{exact}}(\Sigma)$ and its commutant. We remain for now agnostic as to whether the algebras assigned to Σ are minimal or maximal in either theory. The basic statement of the theorem is that O is recoverable from $\mathcal{A}_{\text{exact}}(\Sigma)$ if and only if[45]

$$\langle\Psi|\,[O,X]\,|\Psi\rangle = 0, \qquad \forall X \in (\mathcal{A}_{\text{exact}}(\Sigma))', \quad \forall\,|\Psi\rangle \in \mathcal{H}_{\text{code}}. \tag{6.7}$$

It is not hard to figure out for which O this condition holds. Obviously any operator in the bicommutant, $O \in (\mathcal{A}_{\text{exact}}(\Sigma))'' = \mathcal{A}_{\text{exact}}(\Sigma)$, obeys this condition, where we have assumed the algebra is von Neumann for simplicity. However, we expect more operators to be recoverable given the projection onto $\mathcal{H}_{\text{code}}$ that eq. (6.7) implements. The algebra $(\mathcal{A}_{\text{exact}}(\Sigma))'$ takes the form of eq. (6.5), but the projection in eq. (6.7) removes any potential contribution from the wormhole sector. Hence if an operator X obeys the commutation relation in eq. (6.7), any other operator in $(\mathcal{A}_{\text{exact}}(\Sigma))'$ that only differs from X due to quantum wormholes will do so too. This suffices to conclude that the algebra of recoverable operators is given by

$$\{O \in \mathcal{A}_{\text{min}} \mid O \in \mathcal{A}_{\text{exact}}(\Sigma) \vee \mathcal{A}_{\text{worms}}\}, \tag{6.8}$$

where recall $\mathcal{A}_{\text{worms}}$ captures globally non-local operators associated to the full theory, and not to any specific subspace Σ. To diagnose recoverability, it only remains to specify which

[44] Maximal algebras on Σ now both involve non-local operators which are not generated locally. However, these are associate to weak non-localities, and not to quantum wormholes. This non-wormhole sector is what $\mathcal{A}_{\text{loc}}$ labels in this case.

[45] This theorem is usually stated for X in the algebra of the complementary subspace Σ'. Such a statement would agree with ours for algebras obeying duality. In our case duality is generally violated, so referring to the algebra of the complementary subspace would be ambiguous.

operators we actually have access to on Σ, i.e., which algebra we are associating to Σ in the exact QET. If it is the minimal one, then the algebra of recoverable operators is simply $\mathcal{A}_{\text{min}}(\Sigma)$, since no operator in $\mathcal{A}_{\text{wormh}}$ is in $\mathcal{A}_{\text{min}}$. In this case accessible and recoverable algebras are isomorphic, so the decoding map would also be unique up to isomorphism. If instead we associate the maximal one, then the algebra of recoverable operators consists of those operators in $\mathcal{A}_{\text{max}}(\Sigma) \vee \mathcal{A}_{\text{wormh}} = \mathcal{A}_{\text{max}}(\Sigma)$ which are also in the full $\mathcal{A}_{\text{loc}}$. In other words, these are the type of operators that a maximal algebra on Σ would include in the approximate QET, but which in the exact QET have to be restricted to being additively generated within the full theory so that they have no support on $\mathcal{A}_{\text{wormh}}$. This case is more interesting, since now the accessible algebra includes $\mathcal{A}_{\text{wormh}}$, whereas the recoverable algebra does not.

The last situation above is more reminiscent of holography, where there is a large amount of non-local boundary freedom in reconstructing local bulk operators. In other words, there exist decoding maps which allow to reconstruct bulk operators from many different boundary regions. This freedom here comes from the quantum wormholes inherited from the complete theory, and which respect eq. (6.8) by virtue of only acting nontrivially outside the code subspace. States in the code subspace of the exact QET can be perfectly represented by the approximate QET, much like semiclassical bulk gravity is dual to holographic boundary states in the code subspace. On the opposite extreme one has wormhole states of the exact QET with no support in the code subspace, which cannot be represented at all by the approximate QET. In holography, these are presumably highly quantum gravitational states with no geometric interpretation.

Continuing with this line of reasoning, one may expect that states of the exact QET which are mostly but not completely in the code subspace will be very accurately described by the approximate QET. The description of the latter will however only be approximate, with errors corresponding to its inability to describe quantum wormholes. These errors can be quantified by the relative entropy measures introduced in section 5.3, which diagnose how distinguishable states of the two theories are. As explained therein, operationally these are distinguishable by sufficiently complex operators. In holography, this nicely connects with the idea that highly complex and non-local observables are needed in order to capture any departure of bulk physics from semiclassical gravity [128–136].

In black hole physics, this may explain how the local approximation leads to information loss in Hawking radiation, violating the Bekenstein bound after the Page time. In particular, building on the discussion in section 5.3, our framework provides an explicit mechanism for the breakdown of local effective theory with complexity, even at low energies. This occurs because an enhancement of non-local interactions in an exact QET can cause a local theory to fail in capturing non-negligible non-perturbative correlations. More generally, it may clarify how local approximations in quantum gravity result in an arbitrarily large Hilbert space populated with unphysical null states (cf. non-isometry in holographic maps [135]). Moreover, the non-perturbative non-localities characterized in this work are ubiquitous, particularly during black hole evaporation, and evade the small corrections theorem. It is intriguing to explore the extent to which these effects contribute to the purification of Hawking radiation and a dynamical resolution of the information problem.[46]

[46]These developments will be addressed in future work [137].

JHEP05(2025)024

6.1.5 Baby universes

Another set of ideas that is precisely captured by our constructs in sections 3.4 and 3.5 are the baby universes and α-states of [80–82, 138–146]. In our language, consider a complete theory defined on a manifold M, and a splitting hypersurface Σ consisting of arbitrarily many connected components $\{\Sigma_k\}$. Evaluating the path integral on the cut-open space M_Σ yields the pure state ρ on Σ. However, using the path integral of an exact QET on M_Σ instead produces the mixed state $\widetilde{\rho}$, which can be interpreted as an ensemble of pure states due to its mixed nature.

Our boundaries for baby universes are the distinct connected components $\{\Sigma_k\}$, and the choice of sources on each corresponds to the specification of field configurations Φ_{0k} for the Φ fields. In the path integral for the complete theory on M_Σ, we may thus consider inserting projections of the form $\delta(\Phi|_{\Sigma_k} - \Phi_{0k})$. In the quantum mechanical language these would be the operators

$$\widehat{B[\Phi_{0k}]} \equiv |\Phi_{0k}\rangle \langle\Phi_{0k}| \,. \tag{6.9}$$

These operators commute, as is most obviously seen from the fact that they are associated to delta functions in the path integral. The state one obtains by inserting no operators in the path integral on M_Σ is the pure state ρ above. We identify this as our Hartle-Haking state, and write $\rho \equiv |\mathrm{HH}\rangle \langle\mathrm{HH}|$. Generally, $|\mathrm{HH}\rangle$ is acted on nontrivially by all possible operators in eq. (6.9). Using this state, one can thus think of the path integral on M_Σ with source insertions as actions of the operators in eq. (6.9) on $|\mathrm{HH}\rangle$. Similarly, the path integral with insertions on M without cuts corresponds to computing expectation values of these operators in the Hartle-Hawking state.

The states the operators in eq. (6.9) give rise to are those in eq. (3.26), which correspond to lifts of the projected states from eq. (3.19) to the Hilbert space $\mathcal{H}$ of the complete theory. The Hilbert space of baby universes is thus more naturally associated to the projected states, which here we denote by $\widetilde{\rho}[\{\Phi_{0k}\}]$. Recall that these states are pure only on splitting hypersurfaces, and are otherwise mixed. Therefore here purity corresponds to having choices of sources $\{\Phi_{0k}\}$ specified on every Σ_k boundary. Otherwise, for any finite number of sources, these $\widetilde{\rho}[\{\Phi_{0k}\}]$ states are mixed. The span of all possible such states defines a (not-necessarily proper) subspace of $\mathcal{H}_\varphi$. Because the Φ fields define the space of sources for baby universes, it follows that there will be null state combinations of the $\widetilde{\rho}[\{\Phi_{0k}\}]$ states if $\dim \mathcal{H}_\Phi$ exceeds the dimensionality of the Hilbert space of baby universes, as observed in eq. (3.30).

Because the operators in eq. (6.9) all commute, they can be simultaneously diagonalized. The orthonormal eigenstates to all of them, known as α-states, are here given by states of the form $|\{\Phi_{0k}\}\rangle$ with delta sources specified on every single Σ_k boundary by a projection operator $\alpha[\{\Phi_{0k}\}]$. These were constructed as combinations of projected states in eq. (3.31), but here can be more easily seen to come from specifications of infinite combinations of sources. That these are orthonormal is just the statement of eq. (3.10). The overlaps of α-states with the Hartle-Hawking state give precisely the probability density functions introduced in eqs. (3.22) and (3.32),

$$P_*[\alpha] = \frac{|\langle\mathrm{HH}|\{\Phi_{0k}\}\rangle|^2}{\langle\mathrm{HH}|\mathrm{HH}\rangle}. \tag{6.10}$$

This is the probability of projecting the state of the complete theory onto an α-sector. In terms of the exact QET, the state one obtains without projection is $\widetilde{\rho}$, which by eq. (3.24) is an ensemble of baby universes. By projecting onto an α-sector, one obtains the state $\widetilde{\rho}[\{\Phi_{0k}\}]$, which is a pure state corresponding to a definite baby universe state.

There is, however, a fundamental difference between the usual narrative for α-states in the literature and our construction. The picture the QET framework provides for α-states more closely parallels the perspective of [82]. In particular, different α-states correspond to different states of the Φ fields of the complete theory, which are not merely global parameters but genuine quantum fields with nontrivial position-dependent configurations. Correspondingly, an average over α-sectors does not take the form of a integral over ad-hoc global α-parameters that generate strong non-localities that do not decay with distance. Instead, our average over α-sectors is a functional integral over physical degrees of freedom that were already there in the complete theory. In section 6.2.3 we provide an example of a (wrong) way of truncating a complete theory that leads to an exact QET with the features of a traditional α-parameter integral where strong non-localities do not decay with distance.

6.2 Effective gravity

Although holographic duality provides a non-perturbative definition of quantum gravity in terms of a non-gravitational quantum field theory, one would ultimately hope there to also exist an equally complete but inherently gravitational description of quantum gravity. Whether this is string theoretic or not, dual to a field theory or not, for such a non-perturbative formulation to have a gravitational flavor, one would at least expect it to reduce in some appropriate limit to a geometric theory of metric manifolds. In other words, a theory of quantum gravity should reproduce semiclassical gravity as we know it: a GET where spacetime is dynamical, with all its dimensions manifest and solely governed by a (potentially effective rather than fundamental) metric field.

Given the explorations in this paper on QET, one would hope to be able to draw some lessons about GET as well. However, to say that a GET should somehow behave like a QET just because both are effective theories begs the questions: which aspects of the derivation of a QET from a complete field theory do still makes sense in quantum gravity? While answering this question reliably would require actually having a complete theory of quantum gravity, a few distinguishing features the gravity framework exhibits which are generically absent in field theory and their expected consequences can already be pointed out. In what follows, we will first explain how IR/UV mixing in gravity may naturally cause a GET to exhibit more dramatic non-localities than a QET obtained by renormalization does, and also touch on the ways in which Kaluza-Klein renormalization intrinsically truncates gravity in ways that Wilsonian renormalization in field theory does not.

6.2.1 IR/UV mixing

As emphasized at the beginning of section 2.2, the construction of a QET does not rely in any way on the idea of renormalization à la Wilson, and simply requires a splitting of degrees of freedom into two sectors. For instance, in section 6.1 these two sectors were the black hole and the radiation. The notion of a GET does however encompass the idea of working with gravity at low energies or, more precisely, low curvatures. But the distinction between high and low energies in gravity is a very subtle one, as it turns out, due to the phenomenon of IR/UV mixing.[47] The paradigmatic thought experiment illustrating IR/UV mixing in quantum gravity proceeds as follows.

Consider a particle physics collision at some center of mass energy E, and let E_P denote the Planck energy. In the IR, at energies E many orders of magnitude below E_P, gravity plays no role and the Standard Model applies. At higher energies but still with $E \ll E_P$ gravity begins becoming important, but it can still be treated semiclassically as quantum fields propagating on a fixed curved background. A perturbative treatment of quantum gravity is also possible, and allows one to account for graviton contributions to particle scattering. However, as soon as scales $E \sim E_P$ are reached, one loses perturbative control and this framework breaks down: gravity becomes genuinely quantum. At such energies we expect the collision to involve the formation of a Planck-sized black hole, the physics of which is unknown. In the context of string theory, this is the regime in which Horowitz-Polchinski describes a continuous transition as the energy is raised between a state of highly excited strings and D-branes, and a string-size black hole [147]. Remarkably though, as one continues to raise the energies towards $E \gg E_P$, semiclassical gravity takes over again. The outcome of such an ultra-high energy collision is again well understood and takes the form of a black hole whose Hawking radiation can be well described semiclassically. And if E is many orders of magnitude larger than E_P, the black hole will be sufficiently large that quantum effects in fact become negligible and the end state of the collision is just a classical black hole.

The fact that the deep UV is controlled by IR physics is a characteristic feature of gravity that is rarely seen in field theory. Conversely, in gravity one also observes that the very IR energy scale of the Hawking radiation a large black hole emits is governed inversely by the very UV scale of the mass of the black hole. These are clear manifestations of the phenomenon of IR/UV mixing, which makes assuming a decoupling between IR and UV sectors in a GET naïve. Another reason for expecting significant IR/UV mixing effects in gravity which explicitly also leads to non-local physics is diffeomorphism invariance. The very fact that there exist no local observables in quantum gravity and that any operator has to be gravitationally dressed gives rise in and on itself to a mixing between length scales with an intrinsically non-local flavor [148, 149].

IR/UV mixing does not invalidate the concept of a GET. As we have emphasized, effective theories can be defined without reference to specific energy scales. However, the mechanisms that control the suppression or enhancement of certain effects in a GET may differ significantly from those predicted by standard field theory renormalization. In particular, the non-localities the effective theory exhibits may not merely be exponentially decaying in distance scales determined by some mass of the lightest field that was integrated out (if massive at all). Instead, the scale of non-localities will be governed by IR/UV mixing effects which may actually be dynamically determined and not as simple as some parameter of the theory.

This intuition comes from somewhat exotic field theory constructions which accomplish IR/UV mixing effects by virtue of being defined on non-commutative spaces [150–153]. These non-commutative field theories exhibit momentum-dependent non-localities of arbitrarily

[47] I thank Jake McNamara for pointing out the importance of IR/UV mixing in effective gravity.

long range which actually resemble the physics of strings in many ways [154, 155]. Such non-commutative spaces also arise when studying string theory itself with a background magnetic field [156, 157]. In string theory, a simple argument for IR/UV mixing arises simply from worldsheet modular invariance. In particular, this symmetry relates the low and high energy spectrum of the string, suggesting that the string tension may set a scale for IR/UV mixing.

All in all, one may expect that the natural consequence of IR/UV mixing for a GET would be to further exacerbate the non-localities relative to those expected for a QET. This would lend further support to the idea put forward in section 6.1 that gravitational wormholes may be a manifestation in a GET of what we have been calling quantum wormholes in a QET. This is compelling from a holographic perspective since, as we have seen, gravitational wormholes behave in very much the same way as and capture similar physics to quantum wormholes. But from an intrinsically quantum gravity perspective it also suggests an intriguing possibility. Generally, gravitational wormholes can act as mediators of non-local interactions between quantum fields propagating on them, and thus integrating over gravitational wormholes induces non-localities on otherwise local quantum fields. If gravitational wormholes are analogous to quantum wormholes, it would be worth further exploring whether integrating them out could yield a strongly non-local field theory on a fixed background, similar to an exact QET.[48] An instance of this phenomenon has been explicitly demonstrated recently in [158], where the process of integrating out genus wormholes in JT gravity indeed results in an exact QET for the dilaton field with a strongly non-local potential.

This non-locality sourced by gravitational wormholes provides a failure mode for local approximations, similar to what occurs in exact QET. In particular, building on the ideas of [79–84], it predicts an uncertainty in the couplings derived from local approximations, providing a concrete picture of how these approximations inherently lead to ensemble descriptions in local effective theory. It would be interesting to identify which low-energy but high-complexity observables probe this coupling indeterminacy and illuminate the UV completion of a local effective theory, especially in the presence of IR/UV mixing.

6.2.2 Kaluza-Klein renormalization

Here we comment on some standard procedures within GET interpreted as a gravitational path integral, and the ways in which they are comparable to very coarse truncations in QET, and nowhere close to exact or even approximate QET manipulations.

We will take the gravitational path integral as our starting point. Such an object is formally understood as path integrating not only over all possible metrics one can endow a given manifold with, but also over manifolds of all possible topologies. In the semiclassical limit of string theory though, it is customary to begin by discriminating large, extended dimensions from small, compact ones, and fixing the topology of the latter. From the point of view of a gravity, this is already in tension with diffeomorphism invariance of the theory. Moreover, if the complete theory is to allow for all possible metrics and topologies, this also constitutes a dramatic truncation of the phase space of quantum gravity.

Once the topology of the compact space is fixed, Kaluza-Klein dimensional reduction encodes all metric fluctuations of the compact dimensions into matter fields propagating in

[48] I thank Daniel Harlow for bringing up this possibility.

the extended spacetime. If the Kaluza-Klein tower allows a consistent truncation, one often discards all but the lightest modes, effectively freezing or eliminating the compact dimensions. This procedure again does not consist of integrating out some sector of heavy Kaluza-Klein modes, but rather of drastically truncating away infinitely many quantum fields. While this may seem natural from the perspective of decoupling in Kaluza-Klein renormalization, it is a rather radical operation from a spacetime point of view: this truncation effectively removes regions of spacetime (entire dimensions, actually). Were one to actually integrate out the tower of Kaluza-Klein modes, this would correspond to tracing out spacetime dimensions, a procedure which from a field theory perspective should lead to very significant mixing in the states that the resulting GET would prepares. Put differently, an actual trace over Kaluza-Klein modes would result in a strongly non-local QET, but this is not what is often done. We exemplify in section 6.2.3 why the mass hierarchy of the tower of Kaluza-Klein modes is not obviously helpful for performing this integration.

The truncations above are guided by a Wilsonian intuition for low energies, where it is typically assumed that the lowest truncated Kaluza-Klein mode sets the breakdown scale for the resulting GET. However, as discussed in section 6.2.1, the IR/UV mixing phenomenon suggests that this assumption may not always hold. After all, it may not be so surprising if a GET constructed this way exhibits non-localities in certain regimes which are a lot more dramatic than expected.

6.2.3 A truncation gone wrong

This section presents a simple example of how a naïve mode truncation inside a path integral can lead to a completely wrong result as compared to what exact integration actually gives.

Consider a setting like in section 2.5, involving a theory on a space M with an action given by eqs. (2.5) and (2.8), and with the interaction term in eq. (2.10). We would like to construct a QET for the φ fields. We already know the form of the exact QET action one obtains by exact integration, which is given in eq. (2.13). From eq. (2.15) and below, we also know how to obtain approximations to this result in a reliable way. Here we will attempt (and fail) to find a shortcut to these results by trying to avoid exact integration of a full tower of modes.

Suppose the Laplacian on M admits a complete set of orthonormal eigenfunctions $\{\Phi_k\}$ with nonnegative eigenvalues $\{\lambda_k^2\}$, i.e.,[49]

$$-\nabla^2\Phi_k = \lambda_k^2\Phi_k, \tag{6.11}$$

where completeness and orthonormality respectively mean

$$\sum_{k=0}^{\infty}\Phi_k^\dagger(x)\Phi_k(y) = \delta(x-y), \qquad \int_M d^dx\,\Phi_k^\dagger(x)\Phi_l(x) = \delta_{kl}. \tag{6.12}$$

One can decompose Φ into eigenfunctions of the Laplacian (cf. a momentum representation),

$$\Phi(x) = \sum_{k=0}^{\infty}\alpha_k\Phi_k(x), \qquad \alpha_k \equiv \int d^dx\,\Phi_k^\dagger(x)\Phi(x), \tag{6.13}$$

[49]Similar expressions apply to cases in which this set is a continuous family of eigenfunctions, for which sums become integrals. We are using a discrete notation just for the sake of clarity.

and also easily construct the propagator for Φ satisfying eq. (2.12) as

$$G(x,y) = \sum_{k=0}^{\infty} \frac{\Phi_k^\dagger(x)\Phi_k(y)}{\lambda_k^2 + m_\Phi^2}. \tag{6.14}$$

In the QET, suppose we are only interested in probing φ at energies much smaller than some scale Λ. In Wilsonian renormalization this allows for some nice simplifications if $\Lambda \ll m_\Phi$, as carried out in eq. (2.15) and below, which eventually take one from a local complete theory to a local QET. Alternatively, one could also do this using perturbation theory, but we would like to proceed within the framework of exact integration.

Unfortunately, the path integral over Φ is often too difficult to compute directly. In such cases, an exact QET expression, like the one in eq. (2.11), is unavailable as a starting point. Instead, one must begin with the complete theory and look for possible simplifications. Inspired by Wilsonian renormalization, an approach that seems reasonable is the following.[50]

Without loss of generality, let $\lambda_{k+1}^2 \geq \lambda_k^2$ for all k. Because we are only interested in probing φ at scales much smaller than some energy Λ, one may expect the dynamics of heavy modes of the Φ field with λ_k^2 exceeding Λ^2 to play a negligible role. Hence consider truncating the mode expansion in eq. (6.13) to some $k = N$ such that $\lambda_N^2 \approx \Lambda^2$. This truncation may be as low as to $N = 0$ if already $\lambda_1 \gg \lambda_0$, as one can easily engineer by making M a parametrically small compact space. Then the free action of Φ in terms of the α_k coefficients is simply

$$I_\Phi[\Phi] = \frac{1}{2}\sum_{k=0}^{N} \alpha_k^2(\lambda_k^2 + m_\Phi^2). \tag{6.15}$$

The path integral over Φ becomes a multi-integral over every mode α_k, the measure of which will for convenience be simply denoted by $\mathcal{D}\alpha$. The resulting expression takes the form of a standard Hubbard-Stratonovich transformation,

$$\int_M \mathcal{D}\Phi\, e^{-I[\varphi,\Phi]} \propto e^{-I_\varphi[\varphi]} \int \mathcal{D}\alpha \,\exp\left[-\sum_{k=0}^{N}\left(\frac{\alpha_k^2}{2}(\lambda_k^2 + m_\Phi^2) + \lambda\, \alpha_k \int_M d^dx\, f_\varphi(x)\Phi_k(x)\right)\right] \tag{6.16}$$

$$\propto \exp\left[-I_\varphi[\varphi] + \frac{\lambda^2}{2}\sum_{k=0}^{N}\int_{M^2} d^dx\, d^dy\, \frac{f_\varphi(x)\Phi_k(x)\, f_\varphi(y)\Phi_k^\dagger(y)}{\lambda_k^2 + m_\Phi^2}\right]. \tag{6.17}$$

In the limit $N \to \infty$ this consistently reproduces eq. (2.11) after the mode expansion is resummed. However, at any finite N, this calculation emphasizes what a grave error it would have been to truncate the mode expansion of Φ: non-localities in eq. (6.17) do not die off with distance, but exhibit a periodicity of order Λ^{-1}.[51] An extreme case corresponds to truncating to the lowest $k = 0$ mode only, which would have seemed reasonable if $\lambda_1 \gg \lambda_0$, whether or

[50] The reader is encouraged to read what follows with skepticism; we are going to illustrate an argument that leads to a catastrophically wrong result.

[51] This is akin to a Fourier series where the highest frequency is of order Λ.

JHEP05(2025)024

not m_Φ is large. This corresponds to constant eigenfunctions for which $\lambda_0 = 0$ and one gets

$$\int_M \mathcal{D}\Phi\, e^{-I[\varphi,\Phi]} \propto \exp\left[-I_\varphi[\varphi] + \frac{\lambda^2}{2}\frac{\Phi_0^2}{m_\Phi^2}\int_{M^2} d^dx\, d^dy\, f_\varphi(x) f_\varphi(y)\right]. \tag{6.18}$$

This bilocal interaction is independent of the separation between points, unlike the exact result in eq. (2.11), which decays with distance as determined by the Green's function. The insensitivity arises because truncating Φ to its lowest, position-independent mode effectively correlates its value throughout the entire space. The correlations φ inherits are thus equally scale-independent, and would be able to break factorization at the level of partition functions. This is the same effect that a dynamical cosmological constant would have, as discussed at the end of section 2.4.

In conclusion, truncating a field to its lowest energy modes within the path integral leads to unphysical and strongly non-local consequences for the resulting effective theory. Interestingly, the bilocal coupling that eq. (6.18) gives is the type of interaction α-parameters are often considered to give when integrated out in the context of gravitational wormholes.

Acknowledgments

It is a pleasure to acknowledge very insightful conversations and invaluable feedback from Netta Engelhardt, Daniel Harlow, Tom Hartman, Mukund Rangamani, Marija Tomašević, and Wayne Weng. I would also like to thank Ning Bao, Mike Blake, Nikolay Bobev, Elliott Gesteau, Aidan Herderschee, Gary Horowitz, Veronika Hubeny, Hong Liu, Alex Maloney, Don Marolf, Jake McNamara, Steve Shenker, Jon Sorce, and Evita Verheijden for stimulating discussions. Finally, I am grateful to Elliott Gesteau, Mukund Rangamani, and Marija Tomašević for very helpful comments on an earlier draft of this paper

This research was supported by the U.S. Department of Energy (award DE-SC0021886), by the Templeton Foundation via the Black Hole Initiative (award 62286), and by the MIT Center for Theoretical Physics. I also thank the Kavli Institute for Theoretical Physics (KITP), supported by grant NSF PHY-2309135, for its hospitality during the completion of this work. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

Data Availability Statement. This article has no associated data or the data will not be deposited.

Code Availability Statement. This article has no associated code or the code will not be deposited.

References

[1] J.M. Maldacena, *The large N limit of superconformal field theories and supergravity*, *Adv. Theor. Math. Phys.* **2** (1998) 231 [hep-th/9711200] [INSPIRE].

[2] E. Witten, *Anti de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150] [INSPIRE].

[3] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Gauge theory correlators from noncritical string theory*, *Phys. Lett. B* **428** (1998) 105 [hep-th/9802109] [INSPIRE].

[4] O. Aharony et al., *Large N field theories, string theory and gravity*, *Phys. Rept.* **323** (2000) 183 [hep-th/9905111] [INSPIRE].

[5] J.M. Maldacena and L. Maoz, *Wormholes in AdS*, *JHEP* **02** (2004) 053 [hep-th/0401024] [INSPIRE].

[6] P. Saad, S.H. Shenker and D. Stanford, *A semiclassical ramp in SYK and in gravity*, arXiv:1806.06840 [INSPIRE].

[7] D. Harlow and D. Jafferis, *The Factorization Problem in Jackiw-Teitelboim Gravity*, *JHEP* **02** (2020) 177 [arXiv:1804.01081] [INSPIRE].

[8] J. Chandra and T. Hartman, *Coarse graining pure states in AdS/CFT*, *JHEP* **10** (2023) 030 [arXiv:2206.03414] [INSPIRE].

[9] T. Van Riet, *Instantons, Euclidean wormholes and AdS/CFT*, *PoS* **CORFU2019** (2020) 121 [arXiv:2004.08956] [INSPIRE].

[10] D. Astesiano and F.F. Gautason, *Supersymmetric Wormholes in String Theory*, *Phys. Rev. Lett.* **132** (2024) 161601 [arXiv:2309.02481] [INSPIRE].

[11] A. Hebecker, T. Mikhail and P. Soler, *Euclidean wormholes, baby universes, and their impact on particle physics and cosmology*, *Front. Astron. Space Sci.* **5** (2018) 35 [arXiv:1807.00824] [INSPIRE].

[12] N. Arkani-Hamed, J. Orgera and J. Polchinski, *Euclidean wormholes in string theory*, *JHEP* **12** (2007) 018 [arXiv:0705.2768] [INSPIRE].

[13] D. Marolf and J.E. Santos, *AdS Euclidean wormholes*, *Class. Quant. Grav.* **38** (2021) 224002 [arXiv:2101.08875] [INSPIRE].

[14] A. Belin and J. de Boer, *Random statistics of OPE coefficients and Euclidean wormholes*, *Class. Quant. Grav.* **38** (2021) 164001 [arXiv:2006.05499] [INSPIRE].

[15] N. Afkhami-Jeddi, H. Cohn, T. Hartman and A. Tajdini, *Free partition functions and an averaged holographic duality*, *JHEP* **01** (2021) 130 [arXiv:2006.04839] [INSPIRE].

[16] A. Maloney and E. Witten, *Averaging over Narain moduli space*, *JHEP* **10** (2020) 187 [arXiv:2006.04855] [INSPIRE].

[17] J. Cotler and K. Jensen, *AdS_3 gravity and random CFT*, *JHEP* **04** (2021) 033 [arXiv:2006.08648] [INSPIRE].

[18] N. Benjamin, C.A. Keller, H. Ooguri and I.G. Zadeh, *Narain to Narnia*, *Commun. Math. Phys.* **390** (2022) 425 [arXiv:2103.15826] [INSPIRE].

[19] S. Collier and A. Maloney, *Wormholes and spectral statistics in the Narain ensemble*, *JHEP* **03** (2022) 004 [arXiv:2106.12760] [INSPIRE].

[20] J. Chandra, S. Collier, T. Hartman and A. Maloney, *Semiclassical 3D gravity as an average of large-c CFTs*, *JHEP* **12** (2022) 069 [arXiv:2203.06511] [INSPIRE].

JHEP05(2025)024

[21] S. Collier, L. Eberhardt and M. Zhang, *Solving 3d gravity with Virasoro TQFT*, *SciPost Phys.* **15** (2023) 151 [arXiv:2304.13650] [INSPIRE].

[22] S. Collier, L. Eberhardt and M. Zhang, *3d gravity from Virasoro TQFT: Holography, wormholes and knots*, *SciPost Phys.* **17** (2024) 134 [arXiv:2401.13900] [INSPIRE].

[23] R. Jackiw, *Lower Dimensional Gravity*, *Nucl. Phys. B* **252** (1985) 343 [INSPIRE].

[24] C. Teitelboim, *Gravitation and Hamiltonian Structure in Two Space-Time Dimensions*, *Phys. Lett. B* **126** (1983) 41 [INSPIRE].

[25] A. Almheiri and J. Polchinski, *Models of AdS$_2$ backreaction and holography*, *JHEP* **11** (2015) 014 [arXiv:1402.6334] [INSPIRE].

[26] P. Saad, S.H. Shenker and D. Stanford, *JT gravity as a matrix integral*, arXiv:1903.11115 [INSPIRE].

[27] D. Stanford and E. Witten, *JT gravity and the ensembles of random matrix theory*, *Adv. Theor. Math. Phys.* **24** (2020) 1475 [arXiv:1907.03363] [INSPIRE].

[28] C.V. Johnson, *Nonperturbative Jackiw-Teitelboim gravity*, *Phys. Rev. D* **101** (2020) 106023 [arXiv:1912.03637] [INSPIRE].

[29] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, *The Page curve of Hawking radiation from semiclassical geometry*, *JHEP* **03** (2020) 149 [arXiv:1908.10996] [INSPIRE].

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

[31] 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].

[32] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, *JHEP* **03** (2022) 205 [arXiv:1911.11977] [INSPIRE].

[33] A. Almheiri et al., *Replica Wormholes and the Entropy of Hawking Radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].

[34] A. Almheiri et al., *The entropy of Hawking radiation*, *Rev. Mod. Phys.* **93** (2021) 035002 [arXiv:2006.06872] [INSPIRE].

[35] D.N. Page, *Average entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291 [gr-qc/9305007] [INSPIRE].

[36] D.N. Page, *Black hole information*, in the proceedings of the *5th Canadian Conference on General Relativity and Relativistic Astrophysics (5CCGRRA)*, Waterloo, Canada, May 13–15 (1993) [hep-th/9305040] [INSPIRE].

[37] S.D. Mathur, *The Information paradox: A Pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001 [arXiv:0909.1038] [INSPIRE].

[38] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **02** (2013) 062 [arXiv:1207.3123] [INSPIRE].

[39] D. Harlow, *Jerusalem Lectures on Black Holes and Quantum Information*, *Rev. Mod. Phys.* **88** (2016) 015002 [arXiv:1409.1231] [INSPIRE].

[40] D. Marolf, *The Black Hole information problem: past, present, and future*, *Rept. Prog. Phys.* **80** (2017) 092001 [arXiv:1703.02143] [INSPIRE].

JHEP05(2025)024

[41] J.S. Cotler et al., *Black Holes and Random Matrices*, *JHEP* **05** (2017) 118 [*Erratum ibid.* **09** (2018) 002] [arXiv:1611.04650] [INSPIRE].

[42] J. Liu, *Spectral form factors and late time quantum chaos*, *Phys. Rev. D* **98** (2018) 086026 [arXiv:1806.05316] [INSPIRE].

[43] D. Stanford, *More quantum noise from wormholes*, arXiv:2008.08570 [INSPIRE].

[44] K. Okuyama, *Eigenvalue instantons in the spectral form factor of random matrix model*, *JHEP* **03** (2019) 147 [arXiv:1812.09469] [INSPIRE].

[45] P. Saad, *Late Time Correlation Functions, Baby Universes, and ETH in JT Gravity*, arXiv:1910.10311 [INSPIRE].

[46] M. Winer and B. Swingle, *Hydrodynamic Theory of the Connected Spectral form Factor*, *Phys. Rev. X* **12** (2022) 021009 [arXiv:2012.01436] [INSPIRE].

[47] Y.-N. Zhou, T.-G. Zhou and P. Zhang, *General properties of the spectral form factor in open quantum systems*, *Front. Phys. (Beijing)* **19** (2024) 31202 [arXiv:2303.14352] [INSPIRE].

[48] J.S. Schwinger, *Brownian motion of a quantum oscillator*, *J. Math. Phys.* **2** (1961) 407 [INSPIRE].

[49] L.V. Keldysh, *Diagram technique for nonequilibrium processes*, *Zh. Eksp. Teor. Fiz.* **47** (1964) 1515 [INSPIRE].

[50] F.M. Haehl, R. Loganayagam and M. Rangamani, *Schwinger-Keldysh formalism. Part I: BRST symmetries and superspace*, *JHEP* **06** (2017) 069 [arXiv:1610.01940] [INSPIRE].

[51] F.M. Haehl, R. Loganayagam and M. Rangamani, *Schwinger-Keldysh formalism. Part II: thermal equivariant cohomology*, *JHEP* **06** (2017) 070 [arXiv:1610.01941] [INSPIRE].

[52] P. Pelliconi and J. Sonner, *The influence functional in open holography: entanglement and Rényi entropies*, *JHEP* **06** (2024) 185 [arXiv:2310.13047] [INSPIRE].

[53] R.P. Feynman and F.L. Vernon Jr., *The theory of a general quantum system interacting with a linear dissipative system*, *Annals Phys.* **24** (1963) 118 [INSPIRE].

[54] F. Lombardo and F.D. Mazzitelli, *Coarse graining and decoherence in quantum field theory*, *Phys. Rev. D* **53** (1996) 2001 [hep-th/9508052] [INSPIRE].

[55] U. Weiss, *Quantum Dissipative Systems*, World Scientific (2008) [DOI:10.1142/8334].

[56] G. Lindblad, *On the Generators of Quantum Dynamical Semigroups*, *Commun. Math. Phys.* **48** (1976) 119 [INSPIRE].

[57] H.-P. Breuer and F. Petruccione, *The theory of open quantum systems*, Oxford University Press (2002) [DOI:10.1093/acprof:oso/9780199213900.001.0001].

[58] H. Grabert, P. Schramm and G.L. Ingold, *Quantum Brownian motion: The functional inegral approach*, *Phys. Rept.* **168** (1988) 115 [INSPIRE].

[59] E. Aurell, R. Kawai and K. Goyal, *An operator derivation of the Feynman-Vernon theory, with applications to the generating function of bath energy changes and to an-harmonic baths*, *J. Phys. A* **53** (2020) 275303 [arXiv:1907.02671].

[60] R. LaFlamme and J. Louko, *Reduced density matrices and decoherence in quantum cosmology*, *Phys. Rev. D* **43** (1991) 3317 [INSPIRE].

[61] A.O. Barvinsky and N. Kolganov, *Nonequilibrium Schwinger-Keldysh formalism for density matrix states: Analytic properties and implications in cosmology*, *Phys. Rev. D* **109** (2024) 025004 [arXiv:2309.03687] [INSPIRE].

[62] S. Goldman, N. Lashkari and R.G. Leigh, *A Lindbladian for exact renormalization of density operators in QFT*, arXiv:2410.16582 [INSPIRE].

[63] A.O. Caldeira and A.J. Leggett, *Influence of dissipation on quantum tunneling in macroscopic systems*, Phys. Rev. Lett. **46** (1981) 211 [INSPIRE].

[64] P. Betzios, E. Kiritsis and O. Papadoulaki, *Euclidean Wormholes and Holography*, JHEP **06** (2019) 042 [arXiv:1903.05658] [INSPIRE].

[65] L. Eberhardt, *Partition functions of the tensionless string*, JHEP **03** (2021) 176 [arXiv:2008.07533] [INSPIRE].

[66] L. Eberhardt, *Summing over Geometries in String Theory*, JHEP **05** (2021) 233 [arXiv:2102.12355] [INSPIRE].

[67] E. Gesteau, M. Marcolli and J. McNamara, *Wormhole Renormalization: The gravitational path integral, holography, and a gauge group for topology change*, arXiv:2407.20324 [INSPIRE].

[68] J. Polchinski, *Renormalization and Effective Lagrangians*, Nucl. Phys. B **231** (1984) 269 [INSPIRE].

[69] T.R. Morris, *The exact renormalization group and approximate solutions*, Int. J. Mod. Phys. A **9** (1994) 2411 [hep-ph/9308265] [INSPIRE].

[70] C. Wetterich, *Exact evolution equation for the effective potential*, Phys. Lett. B **301** (1993) 90 [arXiv:1710.05815] [INSPIRE].

[71] O.J. Rosten, *Fundamentals of the Exact Renormalization Group*, Phys. Rept. **511** (2012) 177 [arXiv:1003.1366] [INSPIRE].

[72] D. Skinner, *Lecture Notes on Advanced Quantum Field Theory. Part III*, https://www.damtp.cam.ac.uk/user/dbs26/AQFT.html (2017).

[73] V. Balasubramanian, M.B. McDermott and M. Van Raamsdonk, *Momentum-space entanglement and renormalization in quantum field theory*, Phys. Rev. D **86** (2012) 045014 [arXiv:1108.3568] [INSPIRE].

[74] J.B. Hartle and S.W. Hawking, *Wave Function of the Universe*, Phys. Rev. D **28** (1983) 2960 [INSPIRE].

[75] S.W. Hawking, *The Quantum State of the Universe*, Nucl. Phys. B **239** (1984) 257 [INSPIRE].

[76] D.N. Page, *Density Matrix of the Universe*, Phys. Rev. D **34** (1986) 2267 [INSPIRE].

[77] T. Anous, J. Kruthoff and R. Mahajan, *Density matrices in quantum gravity*, SciPost Phys. **9** (2020) 045 [arXiv:2006.17000] [INSPIRE].

[78] Y. Chen, V. Gorbenko and J. Maldacena, *Bra-ket wormholes in gravitationally prepared states*, JHEP **02** (2021) 009 [arXiv:2007.16091] [INSPIRE].

[79] S.W. Hawking, *The Unpredictability of Quantum Gravity*, Commun. Math. Phys. **87** (1982) 395 [INSPIRE].

[80] S.B. Giddings and A. Strominger, *Loss of incoherence and determination of coupling constants in quantum gravity*, Nucl. Phys. B **307** (1988) 854 [INSPIRE].

[81] S.R. Coleman, *Black holes as red herrings: Topological fluctuations and the loss of quantum coherence*, Nucl. Phys. B **307** (1988) 867 [INSPIRE].

[82] S.W. Hawking, *The alpha parameters of wormholes*, Phys. Scripta T **36** (1991) 222 [INSPIRE].

JHEP05(2025)024

[83] V. Balasubramanian, J.J. Heckman, E. Lipeles and A.P. Turner, *Statistical Coupling Constants from Hidden Sector Entanglement*, *Phys. Rev. D* **103** (2021) 066024 [arXiv:2012.09182] [INSPIRE].

[84] S.B. Giddings and G.J. Turiaci, *Wormhole calculus, replicas, and entropies*, *JHEP* **09** (2020) 194 [arXiv:2004.02900] [INSPIRE].

[85] J. Pollack, M. Rozali, J. Sully and D. Wakeham, *Eigenstate Thermalization and Disorder Averaging in Gravity*, *Phys. Rev. Lett.* **125** (2020) 021601 [arXiv:2002.02971] [INSPIRE].

[86] B. Freivogel, D. Nikolakopoulou and A.F. Rotundo, *Wormholes from averaging over states*, *SciPost Phys.* **14** (2023) 026 [arXiv:2105.12771] [INSPIRE].

[87] D. Marolf and H. Maxfield, *Observations of Hawking radiation: the Page curve and baby universes*, *JHEP* **04** (2021) 272 [arXiv:2010.06602] [INSPIRE].

[88] D. Marolf and H. Maxfield, *The page curve and baby universes*, *Int. J. Mod. Phys. D* **30** (2021) 2142027 [arXiv:2105.12211] [INSPIRE].

[89] H. Casini, M. Huerta, J.M. Magán and D. Pontello, *Entanglement entropy and superselection sectors. Part I. Global symmetries*, *JHEP* **02** (2020) 014 [arXiv:1905.10487] [INSPIRE].

[90] H. Casini, M. Huerta, J.M. Magán and D. Pontello, *Entropic order parameters for the phases of QFT*, *JHEP* **04** (2021) 277 [arXiv:2008.11748] [INSPIRE].

[91] H. Casini and J.M. Magán, *On completeness and generalized symmetries in quantum field theory*, *Mod. Phys. Lett. A* **36** (2021) 2130025 [arXiv:2110.11358] [INSPIRE].

[92] W. Arveson, *An Invitation to C*-Algebras*, Springer New York (1976) [DOI:10.1007/978-1-4612-6371-5].

[93] S.S. Horuzhy, *Introduction to Algebraic Quantum Field Theory*, Springer Netherlands (1990) [DOI:10.1007/978-94-009-1179-6].

[94] R. Haag, *Local Quantum Physics: Fields, particles, algebras*, Springer Berlin Heidelberg (1996) [DOI:10.1007/978-3-642-61458-3].

[95] E. Witten, *A Mini-Introduction To Information Theory*, *Riv. Nuovo Cim.* **43** (2020) 187 [arXiv:1805.11965] [INSPIRE].

[96] M. Van Raamsdonk, *Building up spacetime with quantum entanglement*, *Gen. Rel. Grav.* **42** (2010) 2323 [arXiv:1005.3035] [INSPIRE].

[97] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781 [arXiv:1306.0533] [INSPIRE].

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

[99] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A covariant holographic entanglement entropy proposal*, *JHEP* **07** (2007) 062 [arXiv:0705.0016] [INSPIRE].

[100] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **08** (2013) 090 [arXiv:1304.4926] [INSPIRE].

[101] X. Dong, A. Lewkowycz and M. Rangamani, *Deriving covariant holographic entanglement*, *JHEP* **11** (2016) 028 [arXiv:1607.07506] [INSPIRE].

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

JHEP05(2025)024

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

[104] M. Rangamani and T. Takayanagi, *Holographic Entanglement Entropy*, Springer (2017) [DOI:10.1007/978-3-319-52573-0] [INSPIRE].

[105] A.C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, Class. Quant. Grav. **31** (2014) 225007 [arXiv:1211.3494] [INSPIRE].

[106] C. Akers, N. Engelhardt, G. Penington and M. Usatyuk, *Quantum Maximin Surfaces*, JHEP **08** (2020) 140 [arXiv:1912.02799] [INSPIRE].

[107] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Holographic representation of local bulk operators*, Phys. Rev. D **74** (2006) 066009 [hep-th/0606141] [INSPIRE].

[108] B. Czech, J.L. Karczmarek, F. Nogueira and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, Class. Quant. Grav. **29** (2012) 155009 [arXiv:1204.1330] [INSPIRE].

[109] M. Headrick, V.E. Hubeny, A. Lawrence and M. Rangamani, *Causality & holographic entanglement entropy*, JHEP **12** (2014) 162 [arXiv:1408.6300] [INSPIRE].

[110] D.L. Jafferis and S.J. Suh, *The Gravity Duals of Modular Hamiltonians*, JHEP **09** (2016) 068 [arXiv:1412.8465] [INSPIRE].

[111] A. Almheiri, X. Dong and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, JHEP **04** (2015) 163 [arXiv:1411.7041] [INSPIRE].

[112] F. Pastawski, B. Yoshida, D. Harlow and J. Preskill, *Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence*, JHEP **06** (2015) 149 [arXiv:1503.06237] [INSPIRE].

[113] D.L. Jafferis, A. Lewkowycz, J. Maldacena and S.J. Suh, *Relative entropy equals bulk relative entropy*, JHEP **06** (2016) 004 [arXiv:1512.06431] [INSPIRE].

[114] X. Dong, D. Harlow and A.C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, Phys. Rev. Lett. **117** (2016) 021601 [arXiv:1601.05416] [INSPIRE].

[115] D. Harlow, *The Ryu-Takayanagi Formula from Quantum Error Correction*, Commun. Math. Phys. **354** (2017) 865 [arXiv:1607.03901] [INSPIRE].

[116] N. Engelhardt, S. Fischetti and A. Maloney, *Free energy from replica wormholes*, Phys. Rev. D **103** (2021) 046021 [arXiv:2007.07444] [INSPIRE].

[117] V. Chandrasekaran, N. Engelhardt, S. Fischetti and S. Hernández-Cuenca, *A tale of two saddles*, JHEP **11** (2022) 110 [arXiv:2207.09472] [INSPIRE].

[118] S.W. Hawking, *Particle Creation by Black Holes*, Commun. Math. Phys. **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [INSPIRE].

[119] H. Liu and S. Vardhan, *Entanglement Entropies of Equilibrated Pure States in Quantum Many-Body Systems and Gravity*, PRX Quantum **2** (2021) 010344 [arXiv:2008.01089] [INSPIRE].

[120] R. Bousso and M. Tomašević, *Unitarity From a Smooth Horizon?*, Phys. Rev. D **102** (2020) 106019 [arXiv:1911.06305] [INSPIRE].

[121] R. Bousso and E. Wildenhain, *Gravity/ensemble duality*, Phys. Rev. D **102** (2020) 066005 [arXiv:2006.16289] [INSPIRE].

[122] J. de Boer, D. Liska, B. Post and M. Sasieta, *A principle of maximum ignorance for semiclassical gravity*, *JHEP* **02** (2024) 003 [arXiv:2311.08132] [INSPIRE].

[123] H. Verlinde, *ER = EPR revisited: On the Entropy of an Einstein-Rosen Bridge*, arXiv:2003.13117 [INSPIRE].

[124] H. Verlinde, *Wormholes in Quantum Mechanics*, arXiv:2105.02129 [INSPIRE].

[125] M. Blake and S. Hernández-Cuenca, *The Hamiltonian Ensembles and Spectral Statistics in Effective Theory*, work in progress.

[126] C. Bény, A. Kempf and D.W. Kribs, *Generalization of Quantum Error Correction via the Heisenberg Picture*, *Phys. Rev. Lett.* **98** (2007) 100502 [INSPIRE].

[127] C. Bény, A. Kempf and D.W. Kribs, *Quantum error correction of observables*, *Phys. Rev. A* **76** (2007) 042303.

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

[129] L. Susskind, *Computational Complexity and Black Hole Horizons*, *Fortsch. Phys.* **64** (2016) 24 [arXiv:1403.5695] [INSPIRE].

[130] N. Engelhardt and A.C. Wall, *Coarse Graining Holographic Black Holes*, *JHEP* **05** (2019) 160 [arXiv:1806.01281] [INSPIRE].

[131] 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].

[132] I.H. Kim, E. Tang and J. Preskill, *The ghost in the radiation: robust encodings of the black hole interior (invited paper)*, *JHEP* **06** (2020) 031 [arXiv:2003.05451] [INSPIRE].

[133] 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].

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

[135] C. Akers et al., *The black hole interior from non-isometric codes and complexity*, *JHEP* **06** (2024) 155 [arXiv:2207.06536] [INSPIRE].

[136] L. Yang and N. Engelhardt, *The complexity of learning (pseudo)random dynamics of black holes and other chaotic systems*, *JHEP* **03** (2025) 153 [arXiv:2302.11013] [INSPIRE].

[137] N. Engelhardt and S. Hernández-Cuenca, *Complexity Breakdown of Local Effective Theory and the Non-Local Information in Hawking Radiation*, work in progress.

[138] S.R. Coleman, *Why There Is Nothing Rather Than Something: A theory of the Cosmological Constant*, *Nucl. Phys. B* **310** (1988) 643 [INSPIRE].

[139] S.B. Giddings and A. Strominger, *Baby Universes, Third Quantization and the Cosmological Constant*, *Nucl. Phys. B* **321** (1989) 481 [INSPIRE].

[140] J. Preskill, *Wormholes in Space-time and the Constants of Nature*, *Nucl. Phys. B* **323** (1989) 141 [INSPIRE].

[141] I.R. Klebanov, L. Susskind and T. Banks, *Wormholes and the Cosmological Constant*, *Nucl. Phys. B* **317** (1989) 665 [INSPIRE].

[142] A. Lyons and S.W. Hawking, *Wormholes in string theory*, *Phys. Rev. D* **44** (1991) 3802 [INSPIRE].

JHEP05(2025)024

[143] D. Marolf and H. Maxfield, *Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information*, *JHEP* **08** (2020) 044 [arXiv:2002.08950] [INSPIRE].

[144] J. McNamara and C. Vafa, *Baby Universes, Holography, and the Swampland*, arXiv:2004.06738 [INSPIRE].

[145] A. Blommaert, L.V. Iliesiu and J. Kruthoff, *Gravity factorized*, *JHEP* **09** (2022) 080 [arXiv:2111.07863] [INSPIRE].

[146] A. Blommaert, L.V. Iliesiu and J. Kruthoff, *Alpha states demystified — towards microscopic models of AdS_2 holography*, *JHEP* **08** (2022) 071 [arXiv:2203.07384] [INSPIRE].

[147] G.T. Horowitz and J. Polchinski, *A correspondence principle for black holes and strings*, *Phys. Rev. D* **55** (1997) 6189 [hep-th/9612146] [INSPIRE].

[148] S.B. Giddings, D. Marolf and J.B. Hartle, *Observables in effective gravity*, *Phys. Rev. D* **74** (2006) 064018 [hep-th/0512200] [INSPIRE].

[149] W. Donnelly and S.B. Giddings, *Diffeomorphism-invariant observables and their nonlocal algebra*, *Phys. Rev. D* **93** (2016) 024030 [*Erratum ibid.* **94** (2016) 029903] [arXiv:1507.07921] [INSPIRE].

[150] S. Minwalla, M. Van Raamsdonk and N. Seiberg, *Noncommutative perturbative dynamics*, *JHEP* **02** (2000) 020 [hep-th/9912072] [INSPIRE].

[151] M.R. Douglas and N.A. Nekrasov, *Noncommutative field theory*, *Rev. Mod. Phys.* **73** (2001) 977 [hep-th/0106048] [INSPIRE].

[152] R.J. Szabo, *Quantum field theory on noncommutative spaces*, *Phys. Rept.* **378** (2003) 207 [hep-th/0109162] [INSPIRE].

[153] N. Craig and S. Koren, *IR Dynamics from UV Divergences: UV/IR Mixing, NCFT, and the Hierarchy Problem*, *JHEP* **03** (2020) 037 [arXiv:1909.01365] [INSPIRE].

[154] A. Rajaraman and M. Rozali, *Noncommutative gauge theory, divergences and closed strings*, *JHEP* **04** (2000) 033 [hep-th/0003227] [INSPIRE].

[155] Y. Kiem and S. Lee, *UV / IR mixing in noncommutative field theory via open string loops*, *Nucl. Phys. B* **586** (2000) 303 [hep-th/0003145] [INSPIRE].

[156] E. Witten, *Noncommutative Geometry and String Field Theory*, *Nucl. Phys. B* **268** (1986) 253 [INSPIRE].

[157] N. Seiberg and E. Witten, *String theory and noncommutative geometry*, *JHEP* **09** (1999) 032 [hep-th/9908142] [INSPIRE].

[158] S. Hernández-Cuenca, N. Valdes-Meller and W.W.-E. Weng, *A single geometry from an all-genus expansion in quantum gravity*, arXiv:2412.08799 [INSPIRE].