Perturbative calculations of entanglement entropy

Pouria Dadras and Alexei Kitaev

California Institute of Technology,
Pasadena, CA 91125, U.S.A.

E-mail: pdadras@caltech.edu, kitaev@caltech.edu

ABSTRACT: This paper is an attempt to extend the recent understanding of the Page curve for evaporating black holes to more general systems coupled to a heat bath. Although calculating the von Neumann entropy by the replica trick is usually a challenge, we have identified two solvable cases. For the initial section of the Page curve, we sum up the perturbation series in the system-bath coupling κ; the most interesting contribution is of order $2s$, where s is the number of replicas. For the saturated regime, we consider the effect of an external impulse on the entropy at a later time and relate it to OTOCs. A significant simplification occurs in the maximal chaos case such that the effect may be interpreted in terms of an intermediate object, analogous to the branching surface of a replica wormhole.

Contents

1 Introduction

While not an observable quantity, entropy is useful as an abstract measure of active degrees of freedom and correlations in the system. Quantum correlations can be elusive, particularly in black holes, where the classical space-time picture is incomplete. There has been a long but ultimately successful chase of correlations in the Hawking radiation. If a black hole forms from an object in a pure quantum state and then evaporates, the resulting radiation must also be in a pure state. Thus, it is strongly (albeit nonlocally) correlated. The general form of such correlations was predicted by Don Page [1], who considered a black hole as a generic quantum system. Still, it long remained unclear how such correlations could emerge in semiclassical gravity. Some important works that contributed to the solution include the Dray-'t Hooft mechanism of gravitational interaction between infalling matter and subsequent radiation [2, 3] and the calculation of out-of-time-order correlators (OTOCs) in the black hole setting [4]. However, the OTOC physics is relevant on short time scales and explains correlations that are present not in the radiation itself but relative to a purifying system [5] (that is, under the assumption that the black hole is part of a thermofield double and that we have unrestricted access to the other part). The recent breakthrough in understanding the correlations developing over the Page time [6, 7] required a careful formulation of the problem, which we will now summarize.

The problem at hand is a semiclassical one. We do not have a complete theory of quantum gravity, nor should it be required. When working at the semiclassical level, it is not possible to derive long-term evolution from the short-term one. Rather, one should look for a global solution, which may depend on the quantity of interest. We consider the

entanglement entropy between the black hole and the emitted radiation at a particular time t. So let $\rho = \rho(t)$ be the black hole's density matrix; we want to compute its von Neumann entropy, $S(\rho) = -\operatorname{Tr}(\rho \ln \rho)$. The latter is expressed as the $s \to 1$ limit of the s-Renyi entropy,

$$S_s(\rho) = \frac{1}{1-s} \ln \operatorname{Tr} \rho^s. \tag{1.1}$$

For integer s, the expression $\operatorname{Tr} \rho^s$ may be interpreted as the partition functions of s replicas of the system.

Now, it turns out that the transition from the early phase of the black hole evaporation (when the radiation is uncorrelated as the naive theory predicts) to the later phase (when the entanglement entropy equals the black hole's coarse-grained entropy) is first order. The later phase is described by a new type of space-time geometry, the replica wormhole [6, 7]. Although choosing the correct solution of the two is a global problem, each of them can be examined locally. We will study some properties of both solutions for general many-body systems, where the geometric description is not applicable.

The von Neumann and Renyi entropies are nonlinear functions of the quantum state, which is why they are not observables. However, the logarithmic nonlinearity is mild, such that in the thermodynamic limit, $S(\rho)$ is determined by typical microstates that contribute to the mixed state ρ. In contrast, Renyi entropies are often dominated by a fraction of microstates of tiny overall weight. This distinction is also evident from the replica wormhole picture. The s-Renyi entropy is related to an s-fold cover of space-time, whose metric is different from the physical one. But when we analytically continue the solution in s and take s to 1, we get the standard metric with an additional piece of data, the branching surface. Thus, the $s \to 1$ limit is essential for compatibility with the usual (non-entropic) physics. Our main technical advance is how to take this limit in some specific cases.

2 Early phase of entanglement growth

We adopt a simpler variant of the evaporation problem, where instead of radiating energy, the system comes into contact with a heat bath at the same temperature. Turning the system-bath interaction on represents a slight change in the Hamiltonian and results in a brief period of non-equilibrium dynamics. Then a steady state is achieved such that all simple correlation functions are thermal. However, if the system's initial state was pure (though mimicking the thermal state), its von Neumann entropy will grow at a constant rate. We focus on this regime as well as the very beginning of quantum evolution. The entropy growth eventually saturates at the thermal (i.e. coarse-grained) entropy, but that is not captured by our method.

Our calculation is perturbative in the system-bath coupling strength κ. Note that the von Neumann entropy has a logarithmic singularity at the unperturbed state, which is pure. This is reflected by the fact that in addition to terms of order κ^2 (or any constant power of κ), terms of order κ^{2s} (where s is the number of replicas) play an important role.

2.1 The model and general formulas

Let us consider a quantum system (meant to represent a black hole) with some Hilbert space $\mathcal{H}_B$ and Hamiltonian H_B. For an exact analogy with the evaporation problem, we would have to pick a pure state that looks like thermal to all simple measurements. Instead, we double the system and postulate that its initial state is the thermofield double, $|\text{TFD}_B\rangle \in \mathcal{H}_B^* \otimes \mathcal{H}_B$. Only the right part is coupled to the heat bath, but we are interested in the von Neumann entropy of the double system as its density matrix ρ_{B^*B} evolves in time. Likewise, the bath is also doubled, so that the initial state of the world is

$$|\Psi_0\rangle = |\text{TFD}_B\rangle \otimes |\text{TFD}_b\rangle \in \mathcal{H}_B^* \otimes \mathcal{H}_B \otimes \mathcal{H}_b \otimes \mathcal{H}_b^*. \tag{2.1}$$

A similar, but not identical,[1] setting was used in [8, 9], where the s-Renyi entropy for integer $s > 1$ was calculated.

The full Hamiltonian $H = H_B + H_b + H_{Bb}$ acts only on the two objects in the middle, i.e. on $\mathcal{H}_B \otimes \mathcal{H}_b$. We assume that the interaction term has the form

$$H_{Bb} = \kappa \sum_{j=1}^{N} O_B^j O_b^j \tag{2.2}$$

with some bosonic operators O_B^j and O_b^j. In the case of fermionic systems like the SYK model [10–13], we should multiply the coupling parameter κ by i. For simplicity, we will do the computation for a bosonic system, but the final answer will equally be applicable to fermionic systems.

Thus, the evolution of the world in the interaction picture takes the form

$$\rho_{B^*Bbb^*}(t) = U(t)|\Psi_0\rangle\langle\Psi_0|U^{-1}(t), \qquad U(t) = \mathbf{T}\left(e^{-i\int_0^t H_{Bb}(u)du}\right), \tag{2.3}$$

where $\mathbf{T}$ stands for time ordering. We also assume that $\langle O_B^j\rangle = \langle O_b^j\rangle = 0$.[2] In the rest of the section, we will compute the s-Renyi entropy of the system's density matrix $\rho_{B^*B}(t)$ after tracing out the bath. It is given by the perturbative expansion

$$\rho_{B^*B}(t) = \sum_{n,m} \frac{(i\kappa)^n(-i\kappa)^m}{n!m!} \int_0^t \Big[\mathbf{T}\{O_B^{j_n}(u_n)\cdots O_B^{j_1}(u_1)\}|\text{TFD}_B\rangle\langle\text{TFD}_B|\widetilde{\mathbf{T}}\{O_B^{j_1'}(u_1')\cdots O_B^{j_m'}(u_m')\}$$

$$\times \Big\langle \widetilde{\mathbf{T}}\{O_b^{j_1'}(u_1')\cdots O_b^{j_m'}(u_m')\}\mathbf{T}\{O_b^{j_n}(u_n)\cdots O_b^{j_1}(u_1)\}\Big\rangle \Big] du\,du', \tag{2.4}$$

where the expectation value $\langle\cdots\rangle$ is with respect to the bath's thermal state and $\widetilde{\mathbf{T}}$ denotes reverse time ordering. (If $u_1 < \cdots < u_n$ and $u_1' < \cdots < u_m'$, then the operators are already ordered.) There is also an implicit sum over repeated indices, with each index going from 1 to N. Note that operators with same indices have the same time argument.

---

[1]In refs. [8, 9], the initial state is taken to be the thermofield double of two *interacting* subsystems rather than the product of two thermofield doubles.

[2]In case that $\langle O^j\rangle \neq 0$, one can work with $O^j - \langle O^j\rangle$.

Since the combinatorics might soon get complicated, let us introduce some simplifying graphic notation:

$$|\text{TFD}_\text{B}\rangle = \quad \overset{\text{B}}{\underset{\text{B}^*}{\Big\rangle}}, \qquad \langle\text{TFD}_\text{B}| = \quad \overset{\text{B}^*}{\underset{\text{B}}{\Big\langle}}. \tag{2.5}$$

Then each term in the expansion (2.4) will look like this (where $\mathbf{T}$, $\widetilde{\mathbf{T}}$, and the indices are omitted):

$$O(u_2)O(u_1)|\text{TFD}_\text{B}\rangle \langle O(u_1')O(u_2)O(u_1)\rangle_\text{b} \langle\text{TFD}_\text{B}|O(u_1') = \quad \text{for } u_1 < u_2. \tag{2.6}$$

The diagram element in the middle is the Keldysh contour for the heat bath. It consists of a circle at the bottom representing imaginary-time evolution and a stem corresponding to the real-time evolution; the time goes up. For integer s, $\text{Tr}\,(\rho_{\text{B}^*\text{B}}(t))^s$ can be represented by gluing s such diagrams (describing different replicas of the density matrix) in the cyclic order — see figure 1, where the replicas are depicted with different colors. The expectation values should be independently computed for each closed contour, whether it corresponds to the system or the bath.

Let us further assume that the system-bath coupling is sufficiently weak. Then the "radiation quanta" emanating from the system are sparse, which means that dominant diagrams have at most two operators with close times per contour. Therefore, the calculation can be done using Wick contraction. Of course, if the fields O_B^j, O_b^j are Gaussian, then no sparseness condition is necessary.

An example of a (subleading) Wick pairing contributing to $\text{Tr}\,(\rho_{\text{B}^*\text{B}}(t))^s$ is as follows:

$$\langle \overbrace{O(v_1')O(v_2')}\rangle_\text{b}\langle\overbrace{O(v_1')\overbrace{O(v_2')O(u_2)}O(u_1)}\rangle_\text{B}\langle\overbrace{O(u_2)O(u_1)}\rangle_\text{b} \quad \text{for } u_1 < u_2, \quad v_1' < v_2'. \tag{2.7}$$

It corresponds to the black loop at the bottom of figure 1 (a). In general, a Wick contraction diagram is a disjoint union of loops that consist of alternating solid and dotted lines. Solid lines represent contractions of fields on the same contour, whereas dotted lines correspond to interaction terms such as $O_\text{B}^j(u)O_\text{b}^j(u)$. Each contraction comes with a Kronecker delta identifying the indices of the contracted fields. The result is nonzero if all the indices on each loop are the same. Therefore, each loop with d solid line segments evaluates to $N\kappa^d$ multiplied by the product of two-point functions.

The diagrams for $\ln\big(\text{Tr}\,(\rho_{\text{B}^*\text{B}}(t))^s\big)$ are connected, i.e. contain a single loop. The factor of N that appears here is usual for extensive thermodynamic quantities; thus, the intensive parameter is κ^d. If κ is small, we should only keep loops with $d = 2$ contractions (in one replica) and $d = 2s$ contractions (traversing all the replicas). Both types of loops are shown in purple in figure 1.

There are three types of two-point functions for the system,

$$\langle\mathbf{T}\{O_\text{B}^j(u)O_\text{B}^l(v)\}\rangle = iG_{\mathbf{T}\text{B}}(u-v)\delta_{jl}, \quad \langle\widetilde{\mathbf{T}}\{O_\text{B}^j(u)O_\text{B}^l(v)\}\rangle = iG_{\widetilde{\mathbf{T}}\text{B}}(u-v)\delta_{jl}, \tag{2.8}$$

$$\langle O_\text{B}^{j'}(u')O_\text{B}^j(u)\rangle = iG_\text{B}(u'-u)\delta_{j'j}, \tag{2.9}$$
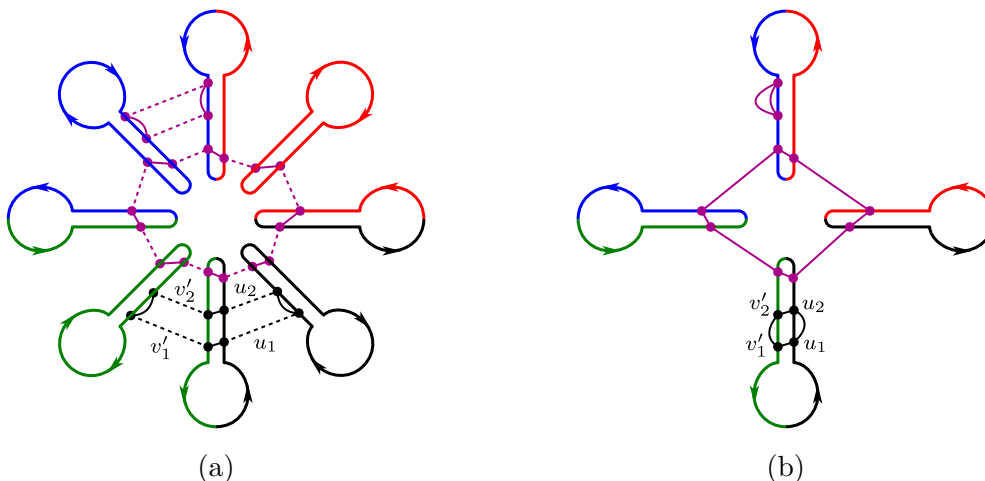
Figure 1. (a) A diagram for $\text{Tr}\left(\rho_{\text{B}*\text{B}}(t)\right)^s$, where $\rho_{\text{B}*\text{B}}(t)$ is given by (2.4) and $s = 4$. The thick solid lines make up $2s$ Keldysh contours for the system and the bath in alternating order. Each half of a contour represents a thermofield-double. Lorentzian time is directed toward the center of the diagram. Different replicas are depicted in different colors. Wick contractions of fields are depicted by thin solid lines, while the dotted lines represent system-bath coupling. The leading contributions are due to purple loops; the shorter and the longer loops correspond to equations (2.11) and (2.12), respectively. (b) The simplified diagram obtained by omitting the bath's replicas.

and similarly for the bath. The expressions for closed paths will be simplified if we think of a two-point function as the matrix element of a bilocal operator $\widehat{G}$,

$$\langle u'|\widehat{G}|u\rangle = f_t(u')G(u'-u)f_t(u), \qquad f_t(u) \equiv \theta(u) - \theta(u-t), \tag{2.10}$$

where $f_t(u)$ is a time window function that vanishes outside the interval $(0, t)$. This way, we can extend the time domain to $(-\infty, \infty)$ and avoid putting limits on the integrals. We also define the transpose of the operator $\widehat{G}$, denoted by $\widehat{G}^{\mathsf{T}}$, with the matrix element $\langle t_1|\widehat{G}^{\mathsf{T}}|t_2\rangle = G(t_2 - t_1)$. Note that $\widehat{G}_{\mathbf{T}}$ and $\widehat{G}_{\widetilde{\mathbf{T}}}$ for bosons are symmetric, i.e. equal to their transpose. To illustrate this notation, the expression (2.7) involves a loop with four Wick contractions, $v'_1 \overset{\text{b}}{\leftarrow} v'_2 \overset{\text{B}}{\leftarrow} u_2 \overset{\text{b}}{\leftarrow} u_1 \overset{\text{B}}{\leftarrow} v'_1$; hence, the result takes the form $\text{Tr}(\widehat{G}_{\widetilde{\mathbf{T}}\,\text{b}} \circ \widehat{G}_{\text{B}} \circ \widehat{G}_{\mathbf{T}\,\text{b}} \circ \widehat{G}_{\text{B}}^{\mathsf{T}})$.

As already mentioned, there are two types of loops that contribute to $\ln\left(\text{Tr}\left(\rho_{\text{B}*\text{B}}(t)\right)^s\right)$ to leading order. The loops of length $d = 2$ can themselves be of two forms, one of which appears in figure 1. They give the following contributions:

$$P_{2,\mathbf{T}}(t) = N\kappa^2\,\text{Tr}(\widehat{G}_{\mathbf{T}\,\text{B}} \circ \widehat{G}_{\mathbf{T}\,\text{b}}^{\mathsf{T}}), \qquad P_{2,\widetilde{\mathbf{T}}}(t) = N\kappa^2\,\text{Tr}(\widehat{G}_{\widetilde{\mathbf{T}}\,\text{B}} \circ \widehat{G}_{\widetilde{\mathbf{T}}\,\text{b}}^{\mathsf{T}}). \tag{2.11}$$

Here we have used the fact that $(\pm i)^2$ from (2.4) cancels i^2 from (2.8). (The extra factors present in the fermionic case also cancel each other.) There are also loops of length $d = 2s$, which we say to have winding number 1 because they traverse all replicas. The expression for such a loop takes the following form:

$$P_{2s}(t) = N\kappa^{2s}\,\text{Tr}\left(-\widehat{G}_{\text{B}} \circ \widehat{G}_{\text{b}}^{\mathsf{T}}\right)^s. \tag{2.12}$$

Since in the end we are interested in the limit $s \to 1$, the quantities (2.12) and (2.11) are of the same order in the coupling.

As an exercise, let us sum up the leading diagrams in $\mathrm{Tr}\,(\rho_{\mathrm{B}^*\mathrm{B}}(t))^s$ — we should get the exponential of a sum of single loops with certain coefficients. It is sufficient to only keep track of the system's replica, leaving the bath implicit as in figure 1 (b). Let $m_1, \ldots, m_s$ and $n_1, \ldots, n_s$ be the numbers of fields in the time ordered and anti-time ordered branches of the Keldysh contours. The number of ways to break them into k loops of length $2s$ with winding number 1 and some loops of length 2 (with winding number 0) is given by

$$(k!)^{2s-1} \prod_{r=1}^{s} \binom{m_r}{k}\binom{n_r}{k} \frac{(m_r-k)!}{\left(\frac{m_r-k}{2}\right)! \, 2^{\frac{m_r-k}{2}}} \frac{(n_r-k)!}{\left(\frac{n_r-k}{2}\right)! \, 2^{\frac{n_r-k}{2}}} \tag{2.13}$$

Defining $m_r - k = 2p_r$ and $n_r - k = 2q_r$, after manipulation we will get

$$\mathrm{Tr}\,(\rho_{\mathrm{B}^*\mathrm{B}}(t))^s = \sum_{k=0}^{\infty} \sum_{p_1,\ldots,p_s} \sum_{q_1,\ldots,q_s} \frac{1}{k!} \prod_{r=1}^{s} \frac{1}{(p_r)!(q_r)! \, 2^{p_r} 2^{q_r}} \, (P_{2s}(t))^k \, (P_{2,\mathbf{T}}(t))^{p_r} (P_{2,\widetilde{\mathbf{T}}}(t))^{q_r}$$

$$= \exp\left(\frac{1}{2}\left(P_{2,\mathbf{T}}(t) + P_{2,\widetilde{\mathbf{T}}}(t)\right) + P_{2s}(t)\right) \tag{2.14}$$

Using the fact that $\mathrm{Tr}(\widehat{G}_{\mathbf{T}\mathrm{B}} \circ \widehat{G}_{\mathbf{T}\mathrm{b}}^{\mathsf{T}}) + \mathrm{Tr}(\widehat{G}_{\widetilde{\mathbf{T}}\mathrm{B}} \circ \widehat{G}_{\widetilde{\mathbf{T}}\mathrm{b}}^{\mathsf{T}}) = 2\,\mathrm{Tr}(\widehat{G}_{\mathrm{B}} \circ \widehat{G}_{\mathrm{b}}^{\mathsf{T}})$, the final answer is as follows:

$$\ln\left(\mathrm{Tr}\,(\rho_{\mathrm{B}^*\mathrm{B}}(t))^s\right) = N\left(s\kappa^2\,\mathrm{Tr}(\widehat{G}_{\mathrm{B}} \circ \widehat{G}_{\mathrm{b}}^{\mathsf{T}}) + \kappa^{2s}\,\mathrm{Tr}\left((-\widehat{G}_{\mathrm{B}} \circ \widehat{G}_{\mathrm{b}}^{\mathsf{T}})^s\right)\right). \tag{2.15}$$

While the above equation was derived for bosonic systems, it is the same for fermionic systems like the SYK model.

2.2 Short initial period vs. linear growth

Equation (2.15) allows one to compute the von Neumann entropy $S(\rho_{\mathrm{B}^*\mathrm{B}}(t))$ for $t < t_{\mathrm{Page}}$. Although the exact answer is model-dependent, there are two universal regimes: very early times, just after the system-bath coupling is turned on, and intermediate times, when the entropy grows linearly.

Very early times. Let $t \ll t_{\mathrm{UV}}$ such that the effect of the Hamiltonians H_{B} and H_{b} is negligible. For example, $t_{\mathrm{UV}} = J^{-1}$ for the SYK model. More exactly, we assume that the Green functions $G_{\mathrm{B}}(t)$ and $G_{\mathrm{b}}(t)$ may be approximated by some constants. Then the expression (2.15) and the von Neumann entropy take this form:

$$\ln\left(\mathrm{Tr}\,(\rho_{\mathrm{B}^*\mathrm{B}}(t))^s\right) \approx N\left(-s\,c\kappa^2 t^2 + (c\kappa^2 t^2)^s\right), \tag{2.16}$$

$$S(\rho_{\mathrm{B}^*\mathrm{B}}(t)) \approx c\kappa^2 t^2 \left(-\ln(c\kappa^2 t^2) + 1\right), \quad \text{where} \quad c = -G_{\mathrm{B}}(0)G_{\mathrm{b}}(0). \tag{2.17}$$

Intermediate times. For systems with continuous excitation spectrum, connected correlators decay in time. Exponential decay is typical; for example, in a conformal system at finite temperature, the correlator of fields with scaling dimension Δ decays as $\exp\left(-\frac{2\pi\Delta}{\beta}t\right)$ if $t \gg \frac{\beta}{\Delta}$. Let us assume that both $G_{\mathrm{B}}(t)$ and $G_{\mathrm{b}}(t)$ decay exponentially at $t \gg t_*$.

If $t \gg t_*$, then $\mathrm{Tr}(\widehat{G}_{\mathrm{B}} \circ \widehat{G}_{\mathrm{b}}^{\mathsf{T}})$ can be approximated as follows. This expression is as integral over $u, u' \in (0, t)$, but the integrand is negligible unless $|u' - u| \lesssim t_*$. Therefore, we may remove the limits on u', and then use the Fourier transform:

$$\mathrm{Tr}(\widehat{G}_{\mathrm{B}} \circ \widehat{G}_{\mathrm{b}}^{\mathsf{T}}) \approx \int_0^t du \int_{-\infty}^\infty du' \, G_{\mathrm{B}}(u', u) G_{\mathrm{b}}(u', u) = t \int \tilde{G}_{\mathrm{B}}(\omega) \tilde{G}_{\mathrm{b}}(-\omega) \frac{d\omega}{2\pi}, \qquad (2.18)$$

where $\tilde{G}(\omega) = \int_{-\infty}^\infty G(t) e^{i\omega t} \, dt$. The second term in (2.15) can also be approximated in such a way. Thus,

$$\ln\Big(\mathrm{Tr}(\rho(t)_{\mathrm{B}^*\mathrm{B}})^s \Big) \approx N A(s) \, t, \qquad (2.19)$$

where

$$A(s) = s\kappa^2 \int \tilde{G}_{\mathrm{B}}(\omega) \tilde{G}_{\mathrm{b}}(-\omega) \frac{d\omega}{2\pi} + \kappa^{2s} \int \big(-\tilde{G}_{\mathrm{B}}(\omega) \tilde{G}_{\mathrm{b}}(-\omega)\big)^s \frac{d\omega}{2\pi}. \qquad (2.20)$$

After analytically continuing to $s = 1$, the von Neumann entropy will be given by

$$S\big(\rho_{\mathrm{B}^*\mathrm{B}}(t)\big) \approx -N A'(1) \, t, \qquad (2.21)$$

$$A'(1) = \kappa^2 \int \tilde{G}_{\mathrm{B}}(\omega) \tilde{G}_{\mathrm{b}}(-\omega)\Big(-\ln\big(-\kappa^2 \tilde{G}_{\mathrm{B}}(\omega) \tilde{G}_{\mathrm{b}}(-\omega)\big) + 1\Big) \frac{d\omega}{2\pi}. \qquad (2.22)$$

The integrals in $A(s)$ and $A'(1)$ converge because a possible peak at $\omega = 0$ is broadened to have a width $\omega_{\min} = t_*^{-1}$. There is also a natural UV cutoff at $\omega_{\max} = t_{\mathrm{UV}}^{-1}$. An interesting case is where both the system and the bath are conformal at $\omega \ll \omega_{\max}$. Let us first assume that the temperature is zero; then $\tilde{G}_{\mathrm{B}}(\omega) \propto \omega^{2\Delta_{\mathrm{B}}-1}$ and $\tilde{G}_{\mathrm{b}}(\omega) \propto \omega^{2\Delta_{\mathrm{b}}-1}$ for $\omega > 0$, but both $\tilde{G}_{\mathrm{B}}$ and $\tilde{G}_{\mathrm{b}}$ vanish at $\omega < 0$. (Recall that these are Wightman functions.) Hence, $\tilde{G}_{\mathrm{B}}(\omega) \tilde{G}_{\mathrm{b}}(-\omega)$ is zero for all $\omega \neq 0$. At finite temperature, the integral in (2.22) is dominated by the region $\omega \sim \beta^{-1}$, where $i\tilde{G}_{\mathrm{B}}(\omega) \sim t_{\mathrm{UV}}(t_{\mathrm{UV}}/\beta)^{2\Delta_{\mathrm{B}}-1}$ and $i\tilde{G}_{\mathrm{b}}(\omega) \sim t_{\mathrm{UV}}(t_{\mathrm{UV}}/\beta)^{2\Delta_{\mathrm{b}}-1}$. It follows that

$$\frac{dS(\rho_{\mathrm{B}^*\mathrm{B}}(t))}{N \, dt} = -A'(1) \sim \frac{-x \ln x}{\beta}, \qquad \text{where} \quad x = (\beta\kappa)^2 \Big(\frac{\beta}{t_{\mathrm{UV}}}\Big)^{-2(\Delta_{\mathrm{B}}+\Delta_{\mathrm{b}})}. \qquad (2.23)$$

A good example is two SYK models at large βJ. (A bath with $\Delta_{\mathrm{b}} = \frac{1}{2}$ can also be realized by a critical Majorana chain.) The Renyi entropies in this case were studied in [9] using the effective action method, which is generally more powerful than perturbation theory. However, the analytic continuation to $s = 1$ was not obtained. The computed growth rate of the s-Renyi entropy for integer $s > 1$ is consistent with our estimate,

$$\frac{dS_s(\rho_{\mathrm{B}^*\mathrm{B}}(t))}{N \, dt} = \frac{A(s)}{1-s} \sim \beta\kappa^2 (\beta J)^{-2(\Delta_{\mathrm{B}}+\Delta_{\mathrm{b}})}. \qquad (2.24)$$

3 Perturbations to the saturated phase

We now consider the system at later times, such that its von Neumann entropy has reached the coarse-grained (thermodynamic) entropy. The entanglement entropy in this phase shows interesting behavior under perturbations. For example, a short impulse increasing the system's energy (similar to throwing a rock into a black hole) will cause a resurgence
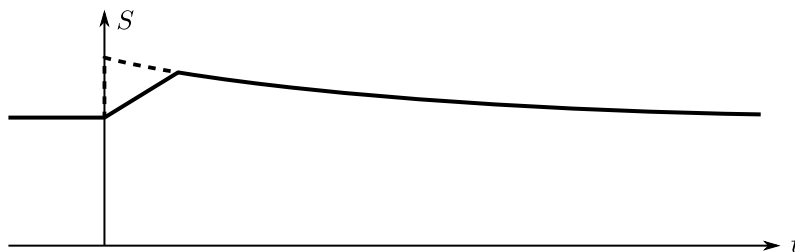
Figure 2. Qualitative plot of the system's coarse-grained entropy (dashed line) and the entanglement entropy (solid line) in the presence of an instantaneous perturbation.

of entropy growth. Indeed, such an action can be described by some unitary operator V. It increases the coarse-grained entropy and effective temperature, though the true microscopic entropy does not change. Letting the system interact with the bath, we should see a behavior similar to the cusp in the Page curve. Specifically, we expect the von Neumann entropy to grow until it becomes equal to the coarse-grained entropy. Since the growth rate is constant while the perturbation can be arbitrarily weak, the resurgence can be short — just slightly longer than the scrambling time. It will be followed by a thermal equilibration period, when both the coarse-grained and microscopic entropies decrease, see figure 2. Thus, the Page curve cusp is accessible in this setting. However, to actually produce a cusp, the perturbation should be sufficiently strong, likely beyond the Taylor expansion. We will study a simpler problem, calculating the effect in the lowest order.

While our formal goal is to compute the von Neumann entropy in a rather general setting, the key result pertains to systems that saturate the chaos bound [14]. The expression obtained in this case admits a holographic interpretation, which will be discussed section 4.

3.1 Statement of the problem

For the study of the saturated phase, it is sufficient to consider one copy of the system and the bath rather than the thermofield double. The bath can be integrated out, giving rise to the interaction function $\sigma(\tau_1, \tau_2) = \kappa^2 G_{\mathrm{b}}(\tau_1, \tau_2)$ on the Keldysh contour. Let us simplify the model a bit and replace the interaction σ, which is constantly on, with a superoperator R acting at a specific time. (The results we will obtain in this setting can be easily generalized to the original model.) So, the exact problem involves a quantum system at thermal equilibrium subjected to a sequence of two instantaneous perturbations, V and R.

Let $V = e^{-ixX}$, where X is a Hermitian operator and x is a small parameter. We consider the action of V on the thermal state ρ_0 and expand the resulting density matrix $\rho_1(x)$ to the second order in x:

$$\rho_1(x) = V\rho_0 V^\dagger \approx \rho_0 - ix[X, \rho_0] + x^2 \left(X\rho_0 X - \frac{1}{2}(X^2\rho_0 + \rho_0 X^2) \right). \qquad (3.1)$$

In fact, a nontrivial effect will be seen in the second order, and only when combined with a subsequent interaction with the environment. The latter is described by a physically realizable (i.e. completely positive, trace-preserving) superoperator R. Suppose that R is

close to the identity such that it can be expanded to the first order in some parameter ϵ:

$$R \approx 1 \cdot 1 + \epsilon L, \qquad L = -i(C \cdot 1 - 1 \cdot C) + \sum_j \left(A_j \cdot A_j^\dagger - \frac{1}{2}(A_j^\dagger A_j \cdot 1 + 1 \cdot A_j A_j^\dagger) \right), \quad (3.2)$$

where $A \cdot B$ stands for the superoperator that takes ρ to $A\rho B$. The first term in L (which involves a Hermitian operator C and act as $\rho \mapsto -i[C, \rho]$) may be neglected because it represents an infinitesimal unitary transformation, and thus, does not change the entropy. The sum over j (known as Lindbladian) corresponds to tracing out the environment. As will be justified later, we may replace L with $\sum_j A_j \cdot A_j^\dagger$ so that the final density matrix becomes

$$\rho(x, \epsilon) = R(\rho_1(x)) = (1 - ixX)\rho_0(1 + ixX) + \epsilon \sum_j A_j(1 - ixX)\rho_0(1 + ixX)A_j^\dagger$$
$$+ \text{ unimportant terms.} \tag{3.3}$$

Our goal is to compute $\frac{\partial^2}{\partial x^2} \frac{\partial}{\partial \epsilon} S(\rho(x, \epsilon))$, where $S(\rho) = -\operatorname{Tr}(\rho \ln \rho)$.

We assume that V acts at time 0, whereas R acts at a later time t. Thus, A_j is understood as $A_j(t) = e^{iH_0 t}A_j(0)e^{-iH_0 t}$, where $A_j(0)$ is some simple (e.g. one- or two-body) operator. The calculation will be done by the replica method for a general large N system in the early time regime, i.e. before the scrambling time. However t is taken to be sufficiently large such that OTOCs are parametrically greater than correlators with non-alternating times. Note that $\rho(x, \epsilon)$ involves only non-alternating operators such as $A_j X\rho_0 X A_j^\dagger$. However, OTOCs appear due to the use of replicas. The "unimportant terms" in (3.3) are exactly those that do not generate any OTOCs.

In the next section, we study partial derivatives of $S(\rho)$, assuming that ρ depends on parameters in some particular way. This setting does not directly include the function $\rho(x, \epsilon)$ given by equation (3.3). To cover this case, we will use a trick called "locking two operators in the same replica", see section 3.5.

3.2 Thermodynamic response theory for the replicated system

Let us recall the standard definition of connected correlators. We begin with the partition function $Z = \operatorname{Tr} W$, where W is the imaginary-time evolution operator:

$$W = \mathbf{T} \exp\left(-\int_0^\beta H(\tau)\, d\tau \right). \tag{3.4}$$

Without perturbation, we have $H(\tau) = H_0$. The insertion of operators $X_1, \ldots, X_n$ at times $\tau_1, \ldots, \tau_n$ is described by perturbing the Hamiltonian:

$$H(\tau) = H_0 - \sum_{j=1}^n x_j \delta(\tau - \tau_j)\, X_j, \qquad \beta \geq \tau_n \geq \cdots \geq \tau_1 \geq 0, \tag{3.5}$$

where x_j are infinitesimal numbers. We generally assume that the operators X_j are bosonic. (If any of them is fermionic, the corresponding variable x_j should be anti-commuting.) Thus,

$$W(\beta, x_n, \ldots x_1) = e^{-(\beta - \tau_n)H_0}(1 + x_n X_n)e^{-(\tau_n - \tau_{n-1})H_0} \cdots (1 + x_1 X_1)e^{-\tau_1 H_0} \tag{3.6}$$

and $Z(\beta, x_n, \ldots x_1) = \operatorname{Tr} W(\beta, x_n, \ldots x_1)$. The full correlator is simply

$$\langle X_n(\tau_n) \cdots X_1(\tau_1) \rangle = Z^{-1} \frac{\partial^n Z}{\partial x_1 \cdots \partial x_n} \Big|_{x_1 = \cdots = x_n = 0} . \tag{3.7}$$

The corresponding connected correlator is defined as follows:[3]

$$\langle X_n(\tau_n), \ldots, X_1(\tau_1) \rangle = \frac{\partial^n \ln Z}{\partial x_1 \cdots \partial x_n} \Big|_{x_1 = \cdots = x_n = 0} . \tag{3.8}$$

For example, $\langle X, Y \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle$ and

$$\langle X, Y, Z \rangle = \langle XYZ \rangle - \langle XY \rangle \langle Z \rangle - \langle XZ \rangle \langle Y \rangle - \langle YZ \rangle \langle X \rangle + 2 \langle X \rangle \langle Y \rangle \langle Z \rangle. \tag{3.9}$$

Now, let us introduce s replicas of the system, such that the partition function becomes

$$Z(s, \beta, x_n, \ldots, x_1) = \operatorname{Tr} \big(W(\beta, x_n, \ldots, x_1) \big)^s. \tag{3.10}$$

We may think of the parameter s as being associated with a *branching operator* $\mathbf{B}$, which commutes with everything. It is not defined by itself but only through its connected correlators:

$$\langle \mathbf{B}, X_n(\tau_n), \cdots, X_1(\tau_1) \rangle = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} \Big|_{x_1 = \cdots = x_n = 0} \frac{\partial \ln Z}{\partial s} \Big|_{s=1} . \tag{3.11}$$

The branched correlator (3.11) is related to the entropy $S = S(\rho)$ of the density matrix $\rho = Z^{-1} W$ at $s = 1$ because

$$\big(\partial_s (\ln Z) \big) \big|_{s=1} = \ln Z - S. \tag{3.12}$$

Thus, the entropy derivative with respect to $x_1, \ldots, x_n$ is given by $-\langle \mathbf{B}, X_n(\tau_n), \cdots, X_1(\tau_1) \rangle + \langle X_n(\tau_n), \cdots, X_1(\tau_1) \rangle$. It is usually the easiest to compute the derivative of the relative entropy, $S(\rho \| \rho_0) = \operatorname{Tr}(\rho(\ln \rho - \ln \rho_0))$:

$$\frac{\partial^n S(\rho \| \rho_0)}{\partial x_1 \cdots \partial x_n} \Big|_{x_1 = \cdots = x_n = 0} = \langle (\mathbf{B} + \beta H_0), X_n(\tau_n), \cdots, X_1(\tau_1) \rangle - \langle X_n(\tau_n), \cdots, X_1(\tau_1) \rangle \tag{3.13}$$

$$= \frac{\partial^n}{\partial x_1 \cdots \partial x_n} \Big|_{x_1 = \cdots = x_n = 0} \big(\partial_s (s^{-1} \ln \operatorname{Tr} (W(\beta/s, \ldots))^s) \big) \big|_{s=1} .$$

For integer s, the expression $\operatorname{Tr} (W(\beta/s, \ldots))^s$ may be interpreted in terms of gluing s intervals of length β/s to make a circle of length β. The operators $X_n(\tau_n), \ldots, X_1(\tau_1)$ are distributed along that circle. Thus, the number in question is, essentially, a correlation function at the given β.

An important caveat is that there is no natural definition of the full correlator $\langle \mathbf{B} \, A \rangle$ as a function of A such that one could compute $\langle \mathbf{B} \, YZ \rangle$ by substituting YZ for A. If such a function (with the usual relation to the connected correlator) existed, we would have this corollary of equation (3.9): $\langle \mathbf{B}, Y, Z \rangle = \langle \mathbf{B}, YZ \rangle - \langle \mathbf{B}, Y \rangle \langle Z \rangle - \langle Y \rangle \langle \mathbf{B}, Z \rangle$. But this last identity is false because in the expression for $\langle \mathbf{B}, Y, Z \rangle$, the operators Y and Z can occur in different replicas, but in $\langle \mathbf{B}, YZ \rangle$, they cannot.

---

[3]We use commas instead of double brackets because the usual notation (without commas) has some ambiguity.

3.3 Branched two-point correlator

Suppose the ordinary correlation function $\langle Y(\tau), X(0) \rangle$ is known on the imaginary axis, $\tau = it$, and let us use its Fourier transform in t. In these terms,

$$\langle Y(\tau), X(0) \rangle = \int_{-\infty}^{\infty} F_{Y,X}(\omega) e^{-\omega\tau} \frac{d\omega}{2\pi}. \tag{3.14}$$

The corresponding branched correlator is expected to have a similar form,

$$\langle \mathbf{B} + \beta H_0, Y(\tau), X(0) \rangle - \langle Y(\tau), X(0) \rangle = \int_{-\infty}^{\infty} h_{Y,X}(\omega) e^{-\omega\tau} \frac{d\omega}{2\pi}. \tag{3.15}$$

The goal of this section is to find the function $h_{Y,X}$.

Let us consider the Fourier modes of the operators Y and X, for example, $Y_\omega = \int Y(it) e^{i\omega t} \, dt$. Their connected correlator is

$$\langle Y_\omega, X_{\omega'} \rangle = F_{Y,X}(\omega) \cdot 2\pi\delta(\omega + \omega'), \tag{3.16}$$

and we also have

$$Y(\tau) = \int \underbrace{Y_\omega e^{-\omega\tau}}_{Y_\omega(\tau)} \frac{d\omega}{2\pi}, \qquad X(0) = \int \underbrace{X_\omega}_{X_\omega(0)} \frac{d\omega}{2\pi}. \tag{3.17}$$

We now calculate the branched correlator of Y_ω and $X_{\omega'}$, equal to $h_{Y,X}(\omega) \cdot 2\pi\delta(\omega + \omega')$. When the number of replicas s is a positive integer, each of the operators in question can be inserted in any replica, so the calculation involves a double sum. Since each replica's length is β/s, putting Y_ω in the k-th replica is described by $Y_\omega(k\beta/s) = Y_\omega e^{-k\beta\omega/s}$. With this in mind, we get:

$$\langle \mathbf{B} + \beta H_0, Y_\omega, X_{\omega'} \rangle - \langle Y_\omega, X_{\omega'} \rangle \tag{3.18}$$

$$= \left(\partial_s \left(s^{-1} \sum_{k=0}^{s-1} \sum_{l=0}^{s-1} \mathbf{T} \left\langle Y_\omega \left(\frac{k\beta}{s} \right), X_{-\omega'} \left(\frac{l\beta}{s} \right) \right\rangle \right) \right) \Bigg|_{s=1} \tag{3.19}$$

$$= \left(\partial_s \sum_{k=0}^{s-1} \left\langle Y_\omega \left(\frac{k\beta}{s} \right), X_{-\omega'}(0) \right\rangle \right) \Bigg|_{s=1} \tag{3.20}$$

$$= \langle Y_\omega, X_{\omega'} \rangle \left(\partial_s \underbrace{\sum_{k=0}^{s-1} e^{-k\beta\omega/s}}_{\frac{1-u}{1-u^{1/s}} \text{ for } u = e^{-\beta\omega}} \right) \Bigg|_{s=1} = 2\pi\delta(\omega + \omega') \cdot F_{Y,X}(\omega) \frac{\beta\omega}{e^{\beta\omega} - 1}. \tag{3.21}$$
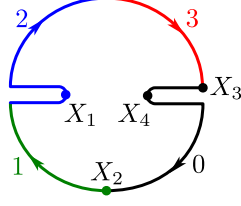
Thus,

$$h_{Y,X}(\omega) = F_{Y,X}(\omega) \frac{\beta\omega}{e^{\beta\omega} - 1}. \tag{3.22}$$

3.4 Branched correlator related to early-time OTOCs

Let us recall the original problem of computing $S(\rho(x, \epsilon))$ with $\rho(x, \epsilon)$ given by equation (3.3). In this section, we calculate an analogous branched correlator $\langle \mathbf{B}, A_j(\beta + it), X(\beta), X(0), A_k^\dagger(it) \rangle$ and, more generally,

$$\langle \mathbf{B}, X_4(\beta + it_4), X_3(\beta + it_3), X_2(it_2), X_1(it_1) \rangle \qquad \text{for } t_1, t_4 \approx t, \quad t_2, t_3 \approx 0. \tag{3.23}$$

Figure 3. Graphic representation of a single term in (3.25). In this example, $s = 4$ (with the replicas labeled $0, 1, 2, 3$), $k = 1$, $j = 0$, $l = 2$, and $t_2 = t_3 = 0$.

One can eliminate β from the time arguments by cyclically permuting $X_4, \ldots, X_1$. As already mentioned, the replica calculation involves OTOCs, which are dominant for sufficiently large t. Neglecting all terms with non-alternating times, we get:

$$\mathcal{B} := \langle \mathbf{B}, X_4(\beta + it_4), X_3(\beta + it_3), X_2(it_2), X_1(it_1) \rangle = \langle \mathbf{B}, X_2(it_2), X_1(it_1), X_4(it_4), X_3(it_3) \rangle$$
$$\approx \left(\partial_s (\mathcal{B}_+(s) + \mathcal{B}_-(s)) \right) \Big|_{s=1}, \qquad (3.24)$$

where

$$\mathcal{B}_+(s) = \sum_{k=0}^{s-1} \sum_{j=0}^{k} \sum_{l=k+1}^{s-1} \left\langle X_1 \left(it_1 + \frac{l\beta}{s} \right), X_2 \left(it_2 + \frac{k\beta}{s} \right), X_4 \left(it_4 + \frac{j\beta}{s} \right), X_3(it_3) \right\rangle, \quad (3.25)$$

$$\mathcal{B}_-(s) = \sum_{k=0}^{s-1} \sum_{j=0}^{k} \sum_{l=k+1}^{s-1} \left\langle X_2(it_2), X_1 \left(it_1 - \frac{j\beta}{s} \right), X_3 \left(it_3 - \frac{k\beta}{s} \right), X_4 \left(it_4 - \frac{l\beta}{s} \right) \right\rangle. \quad (3.26)$$

(Equation (3.25) is illustrated by figure 3.) In order to make further progress, we will use the single-mode ansatz for early-time OTOCs [13],

$$\langle X_1(\tau_1), X_2(\tau_2), X_4(\tau_4), X_3(\tau_3) \rangle \approx -C^{-1} e^{-i\varkappa(\tau_1 + \tau_4 - \tau_2 - \tau_3 - \beta/2)/2}$$
$$\times \Upsilon^{\mathrm{R}}_{X_1, X_4}(\tau_1 - \tau_4) \Upsilon^{\mathrm{A}}_{X_2, X_3}(\tau_2 - \tau_3), \qquad (3.27)$$

combined with the Fourier representation

$$\Upsilon^{\mathrm{R}}_{X_1, X_4}(\tau) = \int \tilde{\Upsilon}^{\mathrm{R}}_{X_1, X_4}(\omega) e^{-\omega\tau} \frac{d\omega}{2\pi}, \qquad \Upsilon^{\mathrm{A}}_{X_2, X_3}(\tau) = \int \tilde{\Upsilon}^{\mathrm{A}}_{X_2, X_3}(\omega) e^{-\omega\tau} \frac{d\omega}{2\pi}. \qquad (3.28)$$

The result has this general form:

$$\mathcal{B} = -C^{-1} e^{\varkappa(t_1 + t_4 - t_2 - t_3)/2} \int \tilde{\mathcal{B}}(\omega_{14}, \omega_{23}) e^{-i\omega_{14}(t_1 - t_4)} e^{-i\omega_{23}(t_2 - t_3)} \frac{d\omega_{23}}{2\pi} \frac{d\omega_{14}}{2\pi}, \quad (3.29)$$

$$\tilde{\mathcal{B}}(\omega_{14}, \omega_{23}) = \tilde{\Upsilon}^{\mathrm{R}}_{X_1, X_4}(\omega_{14}) \tilde{\Upsilon}^{\mathrm{A}}_{X_2, X_3}(\omega_{23}) f(\omega_{14}, \omega_{23}), \qquad (3.30)$$

so the task is to compute $f(\omega_{14}, \omega_{23})$.

First, we find the similar function $f_+(s; \omega_{14}, \omega_{23})$ related to $\mathcal{B}_+(s)$. Let

$$u = e^{-\beta\omega_{23}}, \qquad v = e^{-\beta\omega_{14}}, \qquad w = e^{-i\beta\varkappa/2}. \qquad (3.31)$$

Then

$$f_+(s; \omega_{14}, \omega_{23}) = \sum_{k=0}^{s-1} \sum_{j=0}^{k} \sum_{l=k+1}^{s-1} u^{k/s} v^{(l-j)/s} w^{(j+l-k)/s-1/2} \tag{3.32}$$

$$= \frac{w^{-1/2}}{(1-(w/v)^{1/s})(1-(wv)^{1/s})} \left((vw)^{1/s} \frac{1-u/v}{1-(u/v)^{1/s}} - vw \frac{1-u/w}{1-(u/w)^{1/s}} \right.$$
$$\left. - w^{2/s} \frac{1-uw}{1-(uw)^{1/s}} + v^{1-1/s} w^{1+1/s} \frac{1-u/v}{1-(u/v)^{1/s}} \right), \quad \tag{3.33}$$

and hence,

$$(\partial_s f_+(s; \omega_{14}, \omega_{23}))|_{s=1} = -\frac{1}{(1-uw)(1-uw^{-1})} \left(\frac{w^{-1/2}}{1-v^{-1}u^{-1}} + \frac{w^{1/2}}{v^{-1}-u^{-1}} \right) \ln u$$
$$+ \frac{1}{(1-v^{-1}w)(1-v^{-1}w^{-1})} \left(\frac{w^{-1/2}}{1-uv} + \frac{w^{1/2}}{u-v} \right) \ln v \tag{3.34}$$
$$+ \frac{w^{-1/2}(1+uv^{-1}) - w^{1/2}(u+v^{-1})}{(1-uw)(1-uw^{-1})(1-v^{-1}w)(1-v^{-1}w^{-1})} \ln w.$$

The function f_- is obtained from f_+ by replacing w with w^{-1}. Adding both terms together, we get:

$$f(\omega_{14}, \omega_{23}) = \frac{(w^{-1/2} + w^{1/2})(u^{-1} - 1)(1 + v^{-1})}{(1-uw)(1-uw^{-1})(1-u^{-1}v^{-1})(v^{-1}-u^{-1})} \ln u$$
$$+ \frac{(w^{-1/2} + w^{1/2})(1+u)(1-v)}{(1-v^{-1}w)(1-v^{-1}w^{-1})(1-uv)(u-v)} \ln v \tag{3.35}$$
$$+ \frac{(w^{-1/2} - w^{1/2})(1+u)(1+v^{-1})}{(1-uw)(1-uw^{-1})(1-v^{-1}w)(1-v^{-1}w^{-1})} \ln w,$$

where $u = e^{-\beta\omega_{23}}$, $v = e^{-\beta\omega_{14}}$, and $w = e^{-i\beta\varkappa/2}$.

A great simplification occurs in the maximal chaos case:

$$f(\omega_{14}, \omega_{23}) = \frac{2\pi}{(1+e^{-\beta\omega_{23}})(1+e^{\beta\omega_{14}})} \qquad \text{if} \quad \varkappa = \frac{2\pi}{\beta}. \tag{3.36}$$

Importantly, the function $f(\omega_{14}, \omega_{23})$ splits into two factors. They may be interpreted in terms of interaction of the fluctuating horizon (which corresponds to **B**) with incoming and outgoing radiation, see section 4.

3.5 Locking two operators in the same replica

We now adapt the obtained result to express the entropy of the density matrix $\rho(x, \epsilon)$. The latter is a normalized version of the operator

$$\overline{W}(\beta, x, \epsilon) = (1-ixX)e^{-\beta H_0}(1+ixX) + \epsilon \sum_j A_j(t)(1-ixX)e^{-\beta H_0}(1+ixX) A_j^\dagger(t). \tag{3.37}$$

Note that we have made the time explicit and will follow the convention that $A_j = A_j(0)$. The corresponding entropy could be calculated as in the previous section while restricting

$X_1 = A_j^\dagger$ and $X_4 = A_j$ to be in the same replica. However, instead of imposing this restriction by hand, we will modify the problem so that it fits the branched correlator setting. First, we replace the set of operators A_j with a single operator Y. This is achieved by extending the physical system with an auxiliary one, comprising a ground state $|0\rangle$ with zero energy and a set of excited stated state $|j\rangle$ with energy Ω. We denote the Hamiltonian of the extended system by $H(\Omega)$ and set

$$Y = \sum_j A_j \otimes |0\rangle\langle j|, \qquad Y^\dagger = \sum_j A_j^\dagger \otimes |j\rangle\langle 0|. \tag{3.38}$$

Although the transformation just described alters the operator $\overline{W}(\beta, x, \epsilon)$ in a nontrivial way, we will find an agreement in the $\Omega \to \infty$ limit. For the time being, let us construct some operators acting on the extended system that correspond to the two terms in (3.37) as closely as possible:

$$(1 - ixX)e^{-\beta H(\Omega)}(1 + ixX) = (1 - ixX)e^{-\beta H_0}(1 + ixX) \tag{3.39}$$

$$\otimes \left(|0\rangle\langle 0| + e^{-\beta\Omega} \sum_j |j\rangle\langle j| \right),$$

$$e^{\beta\Omega} Y(t) (1 - ixX)e^{-\beta H(\Omega)}(1 + ixX) Y^\dagger(t) \tag{3.40}$$

$$= \left(\sum_j A_j(t) (1 - ixX)e^{-\beta H_0}(1 + ixX) A_j^\dagger(t) \right) \otimes |0\rangle\langle 0|.$$

Now let

$$W(\Omega, \beta, x, y) = (1 + e^{\beta\Omega/2}yY(t))(1 - ixX)e^{-\beta H(\Omega)}(1 + ixX)(1 + e^{\beta\Omega/2}yY^\dagger(t)). \tag{3.41}$$

Then

$$\lim_{\Omega\to\infty} W(\Omega, \beta, x, y) = \overline{W}(\beta, x, y^2) \otimes |0\rangle\langle 0|, \tag{3.42}$$

and hence,

$$\lim_{\Omega\to\infty} \mathrm{Tr}\big(W(\Omega, \beta, x, y)\big)^s = \mathrm{Tr}\big(\overline{W}(\beta, x, y^2)\big)^s \tag{3.43}$$

for any s. The last equation can be interpreted as the operators $Y(t)$ and $Y^\dagger(t)$ in the expansion of $\mathrm{Tr}\, W^s$ being locked in the same replica. We now take the s derivative of both sides at $s = 1$ and consider the x^2y^2 term in the Taylor expansion. Thus, equation (3.43) becomes

$$\lim_{\Omega\to\infty} \left(e^{\beta\Omega} \Big(-\langle \mathbf{B}, Y(\beta + it), X(\beta), X(0), Y^\dagger(it)\rangle_\Omega + \langle Y(\beta + it), X(\beta), X(0), Y^\dagger(it)\rangle_\Omega \Big) \right)$$

$$= \frac{1}{2} \left(\frac{\partial^2}{\partial x^2} \frac{\partial}{\partial\epsilon} S(\rho(x, \epsilon)) \right)\Bigg|_{x=\epsilon=0} . \tag{3.44}$$

Finally, let us use the results of the previous section. Part of the correspondence is obvious: $X_1 = Y^\dagger$, $X_2 = X_3 = X$, and $X_4 = Y$. In the OTOC-based approximation, the branched correlator

$$\mathcal{B} = \langle \mathbf{B}, Y(\beta + it), X(\beta), X(0), Y^\dagger(it)\rangle_\Omega = \langle \mathbf{B}, X(0), Y^\dagger(it), Y(it), X(0)\rangle_\Omega \tag{3.45}$$

dominates the left-hand side of equation (3.44) and is given by (3.29), (3.30) with $t_1 = t_4 = t$ and $t_2 = t_3 = 0$. Note that if Ω is large, then

$$\Upsilon^{\mathrm{R}}_{Y^\dagger, Y}(\tau) \approx e^{-\Omega(\beta - \tau)} \sum_j \Upsilon^{\mathrm{R}}_{A_j^\dagger, A_j}(\tau) \tag{3.46}$$

by analogy with the rather obvious equation $\langle Y^\dagger(\tau) Y(0) \rangle \approx e^{-\Omega(\beta - \tau)} \sum_j \langle A_j^\dagger(\tau) A_j(0) \rangle$. Hence,

$$\tilde{\Upsilon}^{\mathrm{R}}_{Y^\dagger, Y}(\omega) \approx e^{-\beta \Omega} \sum_j \tilde{\Upsilon}^{\mathrm{R}}_{A_j^\dagger, A_j}(\omega + \Omega). \tag{3.47}$$

On the other hand, the function $\tilde{\mathcal{B}}(\omega_{14}, \omega_{23})$ in equation (3.29) may be replaced with $\tilde{\mathcal{B}}(\omega_{14} - \Omega, \omega_{23})$ without affecting the result.[4] Combining (3.30) with (3.47), we get

$$e^{\beta \Omega} \tilde{\mathcal{B}}(\omega_{14} - \Omega, \omega_{23}) = \sum_j \tilde{\Upsilon}^{\mathrm{R}}_{A_j^\dagger, A_j}(\omega_{14}) \tilde{\Upsilon}^{\mathrm{A}}_{X, X}(\omega_{23}) f(\omega_{14} - \Omega, \omega_{23}), \tag{3.48}$$

where the prefactor $e^{\beta \Omega}$ is used to match the left-hand side of (3.44).

Thus, we have arrived at the conclusion that the replica locking amounts to replacing the function $f(\omega_{14}, \omega_{23})$ in (3.30) with $f(-\infty, \omega_{23})$. Using the explicit formulas (3.35), (3.36), we get:

$$f(-\infty, \omega_{23}) = \frac{\cos \frac{\beta \varkappa}{4} \cdot 2\beta \omega_{23}(1 - e^{-\beta \omega_{23}}) + \sin \frac{\beta \varkappa}{4} \cdot \beta \varkappa (1 + e^{-\beta \omega_{23}})}{(1 - e^{-\beta(\omega_{23} + i\varkappa/2)})(1 - e^{-\beta(\omega_{23} - i\varkappa/2)})} \tag{3.49}$$

$$= \frac{2\pi}{1 + e^{-\beta \omega_{23}}} \qquad \text{if} \quad \varkappa = \frac{2\pi}{\beta}. \tag{3.50}$$

4 Summary and discussion

We have computed the von Neumann entropy in certain many-body settings by combining time-dependent perturbation theory and the replica trick. In particular, we have considered a system made of two parts that are prepared in the thermofield double state, and then one part is coupled to a heat bath. The von Neumann entropy of the two double system grows as $S(\rho(t)) \approx at \ln t^{-1}$ at very short times and as $S(\rho(t)) \approx bt$ at longer times (but less than the Page time); the coefficients a and b have been calculated. These calculations involve unusual terms, proportional to κ^{2s}, in the perturbative expansion in the system-bath coupling strength κ, where s is the number of replicas.

In the second part of the paper, we developed a variant of thermodynamic response theory. In general, thermodynamic response is concerned with perturbing the system with operators $X_1(\tau_1), \ldots, X_n(\tau_n)$ taken with coefficients $x_1, \ldots, x_n$. The standard theory uses the generating function $\ln Z$, whereas our modified version is based on $\ln Z - S$. Thus, the connected correlator turns into a so-called branched correlator,

$$\langle \mathbf{B}, X_n(\tau_n), \cdots, X_1(\tau_1) \rangle = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} \Big|_{x_1 = \cdots = x_n = 0} \left(\ln Z - S \right). \tag{4.1}$$

---

[4]This is because $t_1 = t_4$. Note, however, that the same condition was implicitly used in (3.40). A more general model of system-bath coupling involves $A_j(t_4)$ and $A_j^\dagger(t_1)$ so that the additional factor $e^{-i\Omega(t_1 - t_4)}$ has to be added on the left-hand side of (3.40). To reproduce this factor, one *should* replace $\tilde{\mathcal{B}}(\omega_{14}, \omega_{23})$ with $\tilde{\mathcal{B}}(\omega_{14} - \Omega, \omega_{23})$ in (3.29).

(**B** is not an independently defined object but rather, part of the notation.) While the calculation of branched correlators is rather involved, a simplification occurs for maximally chaotic systems. We now argue that this special case is consistent with a holographic picture in a very broad sense.

For comparison, consider a Euclidean black hole in a hyperbolic space (say, in two dimensions). The replica geometry involves an s-fold cover of both the circle and the disk it bounds, with a branching point at the center. Inserting boundary fields slightly deforms the space. In the $s \to 1$ limit, the geometry is given by a smooth metric on the disk and the position of the branching point. The branching point is a special case of a quantum extremal surface [15] (where "surface" means a codimension 2 submanifold). Its position is determined by an extremum of entropy. Instead of the entropy S, we may consider $\ln Z - S$. Indeed, the partition function Z depends only on the space-time metric, which should be fixed before finding the extremal surface.

In the Lorentzian case, the branching point is described by null coordinates (u_+, u_-). (We set aside the ambiguity in the choice of origin due to the deformation of space-time relative to AdS$_2$.) The entropy can be expanded to the second order in u_+, u_-:

$$S(u_+, u_-) = S_0 + p_+ u_+ + p_- u_- - C u_+ u_-, \qquad (4.2)$$

where p_+ and p_- depend on the inserted field.[5] Solving the extremum problem, we get

$$S_{\text{ext}} = S_0 + C^{-1} p_+ p_-. \qquad (4.3)$$

Now, let us forget about geometry. The only property we need is that if there is large time separation, then p_+ and p_- depend only on the fields inserted in the past and the future, respectively. Thus, the change in the entropy should factor into two quantities dependent on the corresponding fields. This is exactly what we observed in the maximal chaos case, see equation (3.36). We leave the interpretation of these quantities to future research.

Acknowledgments

---

[5]This expression is similar to 't Hooft's effective action [3] for the fluctuating horizon, where p_+ and p_- are null energies. In our case, they are just abstract coefficients.

References

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

[2] T. Dray and G. 't Hooft, *The gravitational shock wave of a massless particle*, *Nucl. Phys. B* **253** (1985) 173 [INSPIRE].

[3] G. 't Hooft, *The black hole interpretation of string theory*, *Nucl. Phys. B* **335** (1990) 138 [INSPIRE].

[4] S.H. Shenker and D. Stanford, *Stringy effects in scrambling*, *JHEP* **05** (2015) 132 [arXiv:1412.6087] [INSPIRE].

[5] P. Hayden and J. Preskill, *Black holes as mirrors: quantum information in random subsystems*, *JHEP* **09** (2007) 120 [arXiv:0708.4025] [INSPIRE].

[6] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *Replica wormholes and the entropy of Hawking radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].

[7] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, arXiv:1911.11977 [INSPIRE].

[8] Y. Gu, A. Lucas and X.-L. Qi, *Spread of entanglement in a Sachdev-Ye-Kitaev chain*, *JHEP* **09** (2017) 120 [arXiv:1708.00871] [INSPIRE].

[9] Y. Chen, X.-L. Qi and P. Zhang, *Replica wormhole and information retrieval in the SYK model coupled to Majorana chains*, *JHEP* **06** (2020) 121 [arXiv:2003.13147] [INSPIRE].

[10] S. Sachdev and J. Ye, *Gapless spin fluid ground state in a random, quantum Heisenberg magnet*, *Phys. Rev. Lett.* **70** (1993) 3339 [cond-mat/9212030] [INSPIRE].

[11] A. Kitaev, *Hidden correlations in the Hawking radiation and thermal noise*, talk at *KITP*, http://online.kitp.ucsb.edu/online/joint98/kitaev/, University of California, Santa Barbara, CA, U.S.A., 12 February 2015.

[12] J. Maldacena and D. Stanford, *Remarks on the Sachdev-Ye-Kitaev model*, *Phys. Rev. D* **94** (2016) 106002 [arXiv:1604.07818] [INSPIRE].

[13] A. Kitaev and S.J. Suh, *The soft mode in the Sachdev-Ye-Kitaev model and its gravity dual*, *JHEP* **05** (2018) 183 [arXiv:1711.08467] [INSPIRE].

[14] J. Maldacena, S.H. Shenker and D. Stanford, *A bound on chaos*, *JHEP* **08** (2016) 106 [arXiv:1503.01409] [INSPIRE].

[15] N. Engelhardt and A.C. Wall, *Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime*, *JHEP* **01** (2015) 073 [arXiv:1408.3203] [INSPIRE].