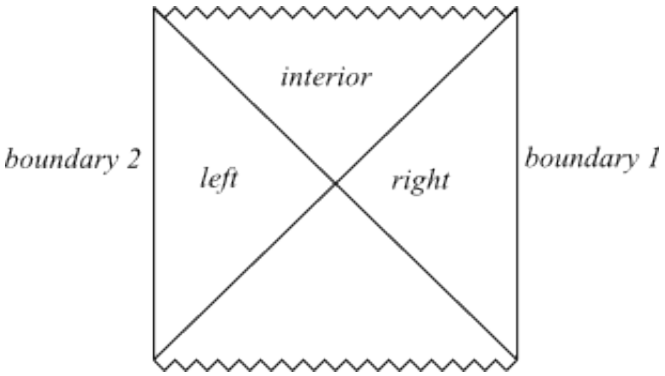
17 Eternal Black Holes and Entanglement

References: This section is based mostly on Maldacena hep-th/0106112; see also the relevant section of Harlow's review lectures, 1409.1231.

An *eternal* black hole is the black hole with the full, two-sided Penrose diagram. It has a past singularity, a future singularity, and two asymptotic regions:



(17.1)

This is to be distinguished from a black hole that forms from gravitational collapse, which has no past singularity and no second asymptotic region on the 'left' of the Penrose diagram. Although we often use the maximally extended Penrose diagram to discuss all sorts of black holes, it is only in the eternal black hole that we should really take the left side of the Penrose diagram seriously.

An eternal black hole in AdS — the maximally extended AdS-Schwarzschild spacetime — has two boundaries. This means that it is dual to two copies of the CFT. In fact, the connection between thermal field theory and this 'doubling' of degrees of freedom was well known long ago, and is called the thermofield double formalism. First we will describe this formalism in QFT, then we'll make the connection to AdS black holes.

17.1 Thermofield double formalism

Consider any QFT, with Hamiltonian H and complete set of eigenstate $|n\rangle$,

$$H|n\rangle = E_n|n\rangle \ . \tag{17.2}$$

The thermofield double formalism is a trick to treat the thermal, mixed state $\rho = e^{-\beta H}$ as a pure state in a bigger system. First we double the degrees of freedom, *i.e.,* we consider a new QFT which is two copies of the original QFT. If the theory is defined by a Lagrangian, then for every field ϕ in the original QFT, there are two fields $\phi_1(x_1)$ and $\phi_2(x_2)$ in the doubled QFT. These two fields live in different spacetimes x_1 and x_2, and are not coupled in the Lagrangian at all. The states of the doubled QFT are

$$|m\rangle_1 |n\rangle_2 \ . \tag{17.3}$$

Now in this doubled system we consider the *thermofield double state*:

$$|TFD\rangle = \frac{1}{\sqrt{Z(\beta)}} \sum_n e^{-\beta E_n/2} |n\rangle_1 |n\rangle_2 \ . \tag{17.4}$$

This is a particular pure state in the doubled system. The density matrix of the doubled QFT in this state is

$$\rho_{total} = |TFD\rangle\langle TFD| \ . \tag{17.5}$$

The reduced density matrix of system 1 is

$$\begin{aligned} \rho_1 &= \mathrm{tr}_2 \, \rho_{total} \\ &= \sum_m {}_2\langle m| \left(\sum_{n,n'} e^{-\beta E_n/2} |n\rangle_1 |n\rangle_2 \, {}_2\langle n'| \, {}_2\langle n'| e^{-\beta E_{n'}/2} \right) |m\rangle_2 \\ &= \sum_n e^{-\beta E_n} |n\rangle_1 \, {}_1\langle n| \\ &= e^{-\beta H_1} \end{aligned} \tag{17.6}$$

Therefore, *if we restrict our attention to system 1, this pure state in the doubled system is indistinguishable from a thermal state.* For example, if O_1 is made of local operators acting on system 1, $O_1 = \phi_1(x_1)\chi_1(y_1)\cdots$, then

$$\langle TFD|O_1|TFD\rangle = \frac{1}{Z(\beta)} \ \mathrm{Tr}_{\ \mathcal{H}_1} \, e^{-\beta H_1} O_1 \ . \tag{17.7}$$

This procedure is called *purifying* the thermal state. In fact, any mixed state can be purified by adding enough auxiliary states and tracing them out.

Although systems 1 and 2 are not coupled in the Lagrangian of the doubled system,

they are correlated because we are in this particular entangled state. For example, if O_1 is built from operators acting on system 1 and O_2 is built from operators acting on system 2, then

$$\langle TFD|O_1O_2|TFD\rangle \tag{17.8}$$

can be non-zero.

The Hamiltonian

The choice of Hamiltonian acting on the doubled system is up to us. Two convenient choices are

$$H_{tot} = H_1 - H_2 \qquad \text{and} \qquad \tilde{H}_{tot} = H_1 + H_2 \ . \tag{17.9}$$

For our purposes, we will just use H_{tot}, but $\tilde{H}_{tot}$ is also useful in other contexts. Under H_{tot}, the TFD state is time-independent, since the phases cancel:

$$|TFD(t)\rangle \equiv e^{-iH_{tot}}|TFD\rangle = \sum_n e^{-\beta E_n/2}e^{-i(H_1-H_2)}|n\rangle_1\,|n\rangle_2 = |TFD\rangle \ . \tag{17.10}$$

17.2 Holographic dual of the eternal black hole

The statement

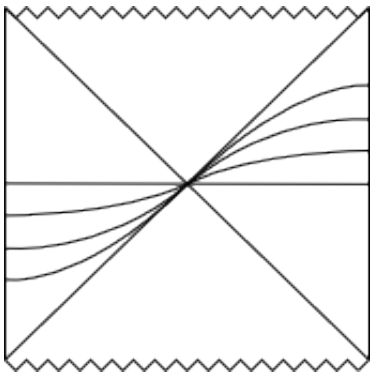
Maldacena's proposal is that the eternal black hole depicted in (17.1) is dual to two copies of the CFT, in the thermofield double state $|TFD\rangle$. Each asymptotic boundary of AdS is a copy of the original dual CFT. So, for example, to compute correlation functions like

$$\langle TFD|\phi_1(x_1)\chi_2(x_2)|TFD\rangle \tag{17.11}$$

we would use Witten diagrams with χ inserted on the left boundary, and ϕ inserted on the right boundary. Note that the local bulk fields are not doubled: there is just one bulk field Φ dual to the boundary operators ϕ_1 and ϕ_2, but this makes sense because we have to specify double the boundary conditions for Φ. The boundary condition for Φ on the left acts like a source for ϕ_2, and the boundary condition for Φ on the right acts like a source for ϕ_1.

The Hamiltonian

The Hamiltonian H_{tot} in (17.9) has a natural bulk interpretation. It is dual to the bulk Hamiltonian that generates time evolution along the isometry ∂_t, where t is the usual Schwarzschild coordinate. Recall (or look back at a textbook on the Kruskal coordinate change) that the Schwarzschild t coordinate runs 'backwards' on the left side of the Penrose diagram. That is, all of the spatial slices drawn in this figure are equivalent under the ∂_t isometry:



(17.12)

This corresponds to the minus sign in $H_{tot} = H_1 - H_2$.

Derivation

To justify the claim that the eternal black hole is dual to the TFD state, we will apply the AdS/CFT dictionary (14.2), in the form

$$Z_{gravity}[\partial M = \Sigma] = Z_{cft}[\Sigma] \ . \tag{17.13}$$

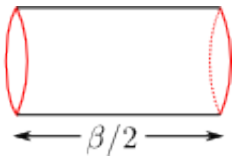
(Here M is the bulk manifold, and the meaning of the lhs is the gravity path integral with boundary condition $\partial M = \Sigma$.)

First, the CFT: The Euclidean path integral that prepares the TFD state is a path integral on an interval of length $\beta/2$, times a circle:

$$\Sigma = \text{Interval}_{\beta/2} \times S^{d-1} \ . \tag{17.14}$$

Pictorially,

$$|TFD\rangle = \quad \beta/2 \tag{17.15}$$



This path integral has two open cuts (red), at the ends of the interval. We interpret the left cut as defining a state in system 2, and the right cut as defining a state in system 1. That is, this picture should be interpreted as a rule for computing the transition amplitude with field data φ_1 and φ_2 specified at the ends of the interval. To confirm that this path integral really prepares the TFD state, all we need to do is check that it computes the correct transition amplitudes. The path integral with these boundary conditions is[73]

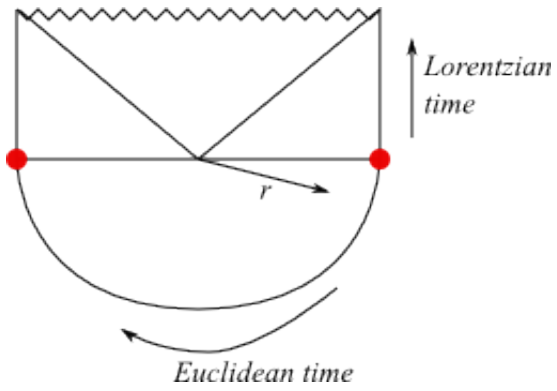
$$\begin{aligned} {}_1\langle\varphi_1|\,{}_2\langle\varphi_2|TFD\rangle &= \langle\varphi_1|e^{-\beta H/2}|\varphi_2^*\rangle && (17.16)\\ &= \sum_n \varphi_1|n\rangle\langle n|\tilde{\varphi}_2\rangle e^{-\beta E_n/2} && (17.17)\\ &= \sum_n e^{-\beta E_n/2}\langle\varphi_1|n\rangle_1\langle\varphi_2|n\rangle_2 && (17.18) \end{aligned}$$

These are precisely the matrix elements of the state $|TFD\rangle$ defined in (17.4). So, as claimed, this is the Euclidean path integral that prepares $|TFD\rangle$.

Now that we've produced this state from a Euclidean path integral on the manifold Σ, we can apply (17.13). We must find a Euclidean gravity solution with conformal boundary condition $\partial M = \text{Interval}_{\beta/2} \times S^{d-1}$. In fact, *half* of the Euclidean black hole has precisely this boundary condition. That is, we consider the Euclidean Scharzschild-AdS solution and restrict to $t_E \in [0, \beta/2]$ instead of the full range $t_E \in [0, \beta]$. The (t_E, r) portion of this Eulidean spacetime makes a half-disk; the boundary of the half-disk is $\text{Interval}_{\beta/2} \times S^{d-1}$. The half-disk is cut down the middle; this cut is interpreted as the time=0 surface of the Lorentzian spacetime. Pictorially, the bulk spacetime has a Euclidean piece that prepares the state, then a Lorentzian piece describing the time

[73] The tildes indicate the conjugate state.

evolution in Minkowski signature:



(17.19)

The red blobs in this picture denote the S^{d-1}'s at the end of the interval on the boundary, the red circles in (17.15).

17.3 ER=EPR

Let's describe this result in words. The left side of the Penrose diagram is dual to CFT_2, and the right side is dual to CFT_1. The Einstein-Rosen bridge connecting to the two sides, the black hole interior itself, are somehow 'created' by entangling CFT_1 with CFT_2. In fact this is a precise statement: In CFT language, correlators between the two CFTs like $\langle TFD|O_1O_2|TFD\rangle$ are non-zero only because we are in the entangled state $|TFD\rangle$. After all, the two CFTs are not coupled. In the bulk, these correlations are nonzero because we can draw Witten diagrams going through the interior. For very massive fields or high energies, the left-right correlation functions can be approximated by geodesics that pass through the black hole interior. Without a wormhole connecting the two sides, there would be no such correlations.

This idea and its generalizations have recently been given the slogan 'ER=EPR' (by Maldacena and Susskind): Einstein-Rosen bridges are equivalent to entanglement (as discussed by Einstein-Podolsky-Rosen). This slogan is only entirely precise and well defined in the semiclassical limit, describing the eternal black hole and similar space-times, but the idea is that some more general construction should make sense in the very quantum, non-geometrical limit.

17.4 Comments in information loss in AdS/CFT

Hawking's information loss paradox relied on a black hole that forms from collapse, then evaporates. In AdS, this only happens for small black holes. These black holes are not in thermal equilibrium, and are difficult to address precisely using AdS/CFT. Of course, the CFT is always unitary, so if we believe AdS/CFT (or use AdS/CFT to define a theory of quantum gravity) then obviously this evaporation process, however it is described in CFT, must be unitary. This strongly suggests that unitarity should be preserved, and locality or some other tenet of effective field theory must be violated. However it is not very satisfying, since it does not answer the question of what went wrong with Hawking's calculation. Presumably the answer is that local effective field theory is not quite right in non-perturbative quantum gravity, but we do not really understand how to characterize this breakdown. This is a very important open question in current research.

17.5 Maldacena's information paradox

Maldacena introduced a different version of the information paradox that applies to large, eternal black holes. This version is easier to address in AdS/CFT. The idea is to first perturb the thermal state by inserting an operator O_2 in CFT_2,

$$|TFD\rangle \to |\widetilde{TFD}\rangle = (1 + \epsilon O_2)|TFD\rangle \ . \tag{17.20}$$

This changes the reduced density matrix of system 1,

$$\rho_1 \to \tilde{\rho}_1 = e^{-\beta H_1} + \text{tiny corrections} \ . \tag{17.21}$$

Now, we compute expectation values in CFT_1,

$$\langle \widetilde{TFD}|O_1|\widetilde{TFD}\rangle \tag{17.22}$$

in the perturbed state. To first order in the perturbation, this is the two-sided correlation function

$$\langle O_1\rangle \sim G_{12} \equiv \langle TFD|O_1 O_2|TFD\rangle \ . \tag{17.23}$$

Now we can produce a contradiction by waiting a very long time, so this correlation function decays. On the gravity side, if we hold O_2 at a particular time and send O_1 to very late times, then the geodesic distance between these two points grows linearly with time, forever. Therefore the correlation function must decay as

$$G_{12}^{gravity} \sim e^{-\text{const}\times t/\beta} \tag{17.24}$$

for $t \gg \beta$. This decays exponentially to zero. At very late times, it therefore becomes exactly thermal, with arbitrarily small corrections.

This contradicts unitarity of the CFT. In the CFT, any perturbation of the thermal state should stay forever a perturbation of the thermal state: it will of course become scrambled and appear to thermalize, but it should never forget the initial perturbation completely, so it should never become arbitrarily close to the thermal state. In fact the corrections to the thermal state should be suppressed by the entropy, but finite:

$$G_{12}^{CFT} \sim e^{-\text{const}\times S} \tag{17.25}$$

for $t \gg \beta$. In summary, at very late times, gravity 'forget' the initial perturbation, but a unitary CFT does not:

$$G_{12}^{gravity} \ll G_{12}^{unitary} \ . \tag{17.26}$$

However is this paradox resolved? The answer is that we have neglected non-perturbative contributions of the gravity side of order $e^{-1/G_N} \sim e^{-S}$. For example, there is another saddlepoint (the thermal AdS saddle) and fluctuations around this saddle will also contribution to the two-sided correlation function at this order.

Although this tells us where the gravity derivation went wrong, it does not tell us exactly how to recover the lost information in quantum gravity, *i.e.,* without referring to the dual CFT. Presumably this would require treating the full non-perturbative string theory, which is currently not possible.