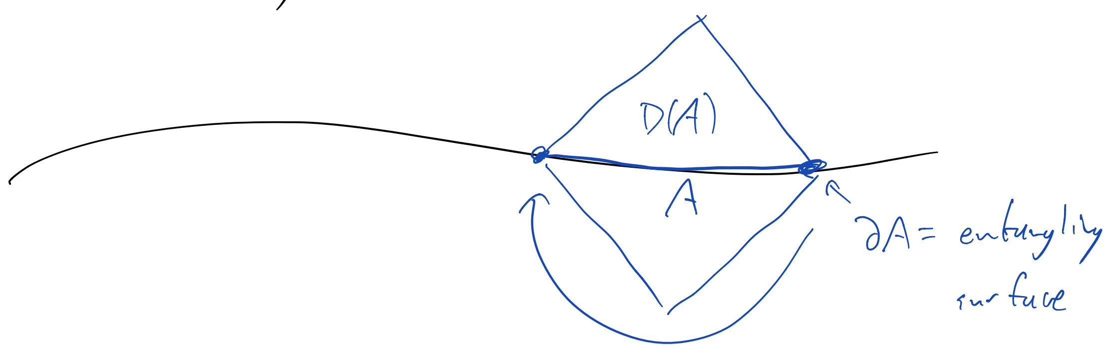
Holographic Entanglement Entropy, PiTP 2018

Matthew Headrick

I. EEs in general QFTs — basic facts + examples

- Natural "subsystem" in relativistic QFT:

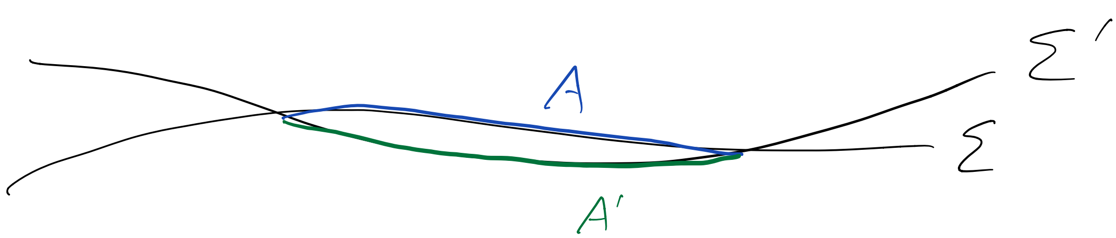
  causal domain $D(A)$ of spatial region A on Cauchy slice



Every causal curve in $D(A)$ intersects A

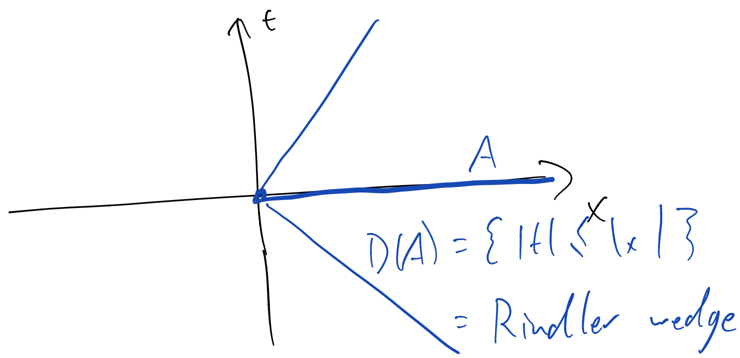
Every op $\mathcal{O}$ in $D(A)$ can be written by Heisenberg EOM in terms of ops in A

$\Rightarrow \langle \mathcal{O} \rangle$ determined by ρ_A



$$D(A') = D(A) \Rightarrow \rho_{A'} = \rho_A \Rightarrow S(A') = S(A)$$

- Half-line



Vacuum $\rho = |0\rangle\langle 0|$

$$\rho_A = \frac{1}{Z} e^{-2\pi K}$$

K = generator of boosts in x-t plane within $D(A)$

$$= \int_0^\infty dx \; x \; T_{tt}$$

Thermal state at temp $\frac{1}{2\pi}$ wrt Hamiltonian K

Local physical temp $T(x) = \frac{1}{2\pi x}$ (Unruh)

Estimate of $S(A)$ for field of mass m:

Entropy density

$$s(T) \simeq \begin{cases} 0, & T < m \\ \# T, & T \gg m \end{cases}$$

$$S(A) \simeq \int_0^\infty dx \; s(T(x)) = \int_0^\infty dx \; s\left(\frac{1}{2\pi x}\right) \simeq \int_0^\xi dx \; \frac{1}{x}$$

$\xi = \frac{1}{m}$ = correlation length

Only fields within distance ξ of entangling surface are significantly entangled

diverges $\Rightarrow$ cut off at $x = \varepsilon$

fields close to entangling surface are <u>very</u> entangled!

$$S(A) \approx \int_\varepsilon^\xi dx \, \frac{1}{x} = \ln \frac{\xi}{\varepsilon}$$

If the theory has UV fixed pt that is a CFT, then for $T \gg m$, $s(T) = \frac{2\pi c}{6} T$, so

$$S(A) = \frac{c}{6} \ln \frac{\xi}{\varepsilon} + \varepsilon\text{-indep.}$$

In higher D, w/ $x^2, \dots, x^{D-1}$ in torus of area σ, $A = \{x^1 \geq 0\}$,

$$S(A) \propto \frac{\sigma}{\varepsilon^{D-2}}$$

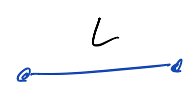
coeff. is not meaningful

Near any smooth ent. surf., A looks like half-space

$\Rightarrow$ $S(A)$ UV divergent like $\frac{\sigma}{\varepsilon^{D-2}} \leftarrow$ area of ∂A

- 2d CFT: 1 interval

$\xi \to \infty \Rightarrow$ IR divergence

Cut off by considering interval 

2 endpoints $\Rightarrow$ $S(A) = \frac{c}{3} \ln \frac{1}{\epsilon} + \cdots$

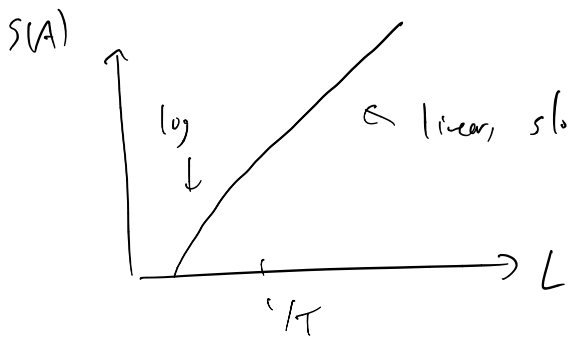
dim. analysis

$\Rightarrow$ $S(A) = \frac{c}{3} \ln \frac{L}{\epsilon}$ + non-universal const.

(can be checked by various honest calcs., which give more info, allow generalizations)

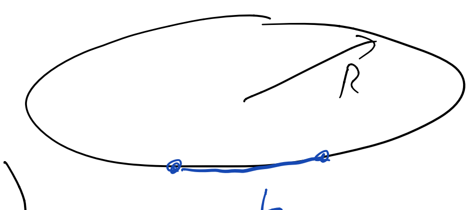
Now suppose CFT is at temp T

$$S(A) = \frac{c}{3} \ln \frac{\sinh \pi T L}{\pi T \epsilon}$$

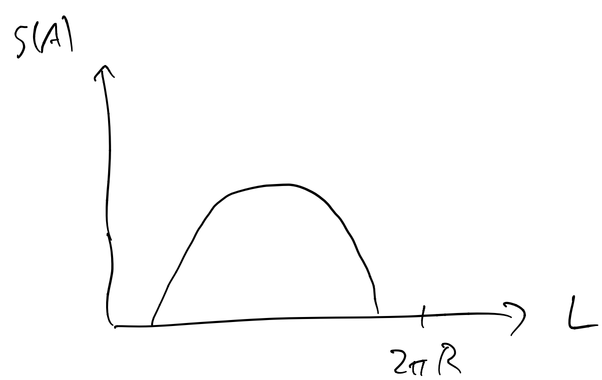


$\leftarrow$ linear, slope $= \frac{2\pi}{6} c T L = s(T)$

CFT on S^1 in vac:



$$S(A) = \frac{c}{3} \ln\left(\frac{2R}{\epsilon} \sin \frac{L}{2R} \right)$$



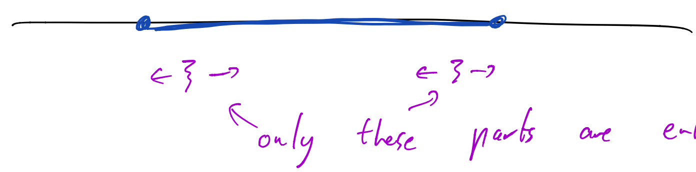
Note: • $S(A^c) = S(A)$ (pure state)

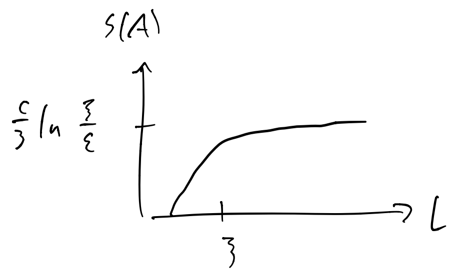
• $\frac{dS(A)}{dL} < 0$ for $L > \pi R$

$\Rightarrow$ cond. ent. < 0

- Capped theory: interval

Expect
$$S(A) \approx \begin{cases} \frac{c}{3} \ln \frac{L}{\varepsilon} & L \ll \xi \\ \frac{c}{3} \ln \frac{\xi}{\varepsilon} & L \gg \xi \end{cases}$$

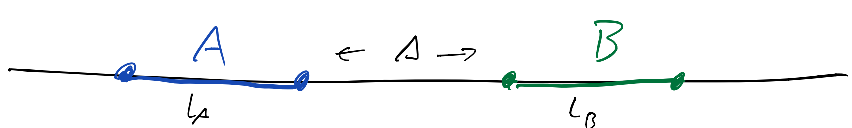




(checked semi-analytically for free bosons + fermions)

(Note that all curves so far have been concave ⇐ SSA)

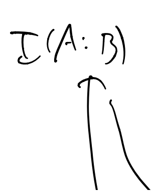
- CFT: 2 intervals

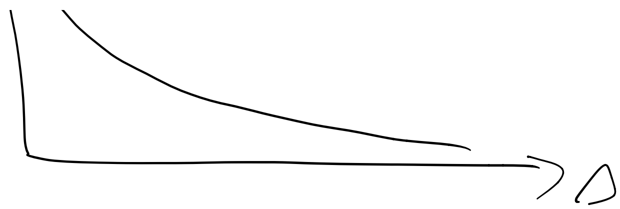


$$I(A:B) = S(A) + S(B) - S(AB)$$

Only calculated exactly for free fermions:

$$I(A:B) = \frac{1}{2} \ln \frac{(L_A + \Delta)(L_B + \Delta)}{\Delta (L_A + L_B + \Delta)}$$





Qualitative features hold in any CFT (+ more generally)

- Finite (indep. of ε) $\Leftarrow$ divergences are local at endpoints
- > 0 (makes sense: ≥ 0 by subadditivity, if 0 then $\rho_{AB} = \rho_A \otimes \rho_B \Rightarrow$ no correlations)
- Conformally invariant i.e. fn of cross-ratio of endpoints ($\Leftarrow$ finite)
- Non-decreasing fn of L_B ($\Leftarrow$ SSA, $I(A:BC) \geq I(A:B)$)
  $\Rightarrow$ non-increasing fn of Δ (makes sense, since $I(A:B)$ measures correlations)
- $\to \infty$ as $\Delta \to 0$
- $\to 0$ as $\Delta \to \infty$ (rate determined by lightest non-identity op)

II. Ryu-Takayanagi formula

- Motivate (ahistorically) from an example we understand

Thermal state: $\rho_A = \frac{1}{Z} e^{-\beta H} = \frac{1}{Z} \sum_i e^{-\beta E_i} |i\rangle_A \langle i|_A$

TFD: $|TFD\rangle = \frac{1}{\sqrt{Z}} \sum_i e^{-\beta E_i/2} |i\rangle_A \, |i\rangle_{A^c}$

A is another copy of same system

Unique purification of ρ_A invariant under

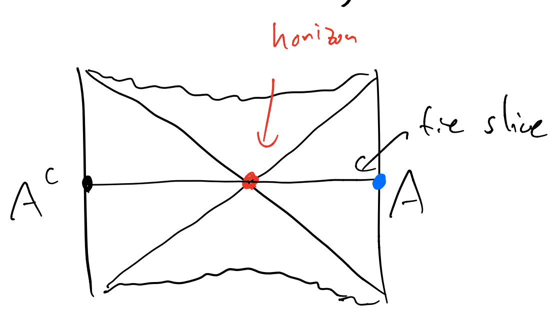
$A \leftrightarrow A^c$ exchange × T-reversal

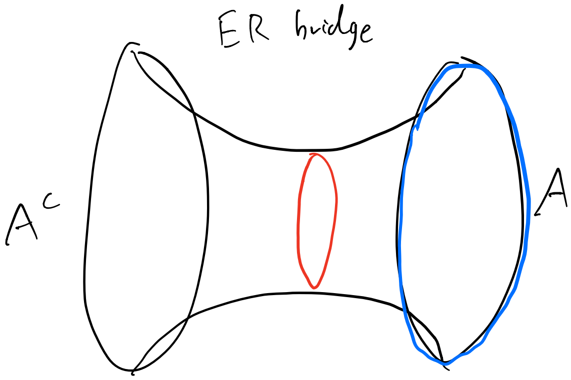
By construction, $S(A) = S_{thermo}$

In holographic QFT, ρ_A is described for $T > T_{HP}$

by black hole $S(A) = S_{BH} = \frac{1}{4G_N} \text{area}(\text{horizon}) + \mathcal{O}(G_N^0)$

$|TFD\rangle$ described by max. analytic extension:





So in this case,

$$S(A) = \frac{1}{4G_N}\,\text{area}(\text{hor})\left[+\mathcal{O}(G_N^0)\right]$$

What distinguishes horizon?

Fixed point of Killing sym.

But if we want to generalize to examples without Killing symmetries, we need weaker condition

Horizon = minimal-area surface between A + A^c (on time-slice)

What picks out time-slice? Symmetry: invariant under time-reflection

Want to, + will, get rid of this symmetry assumption as well

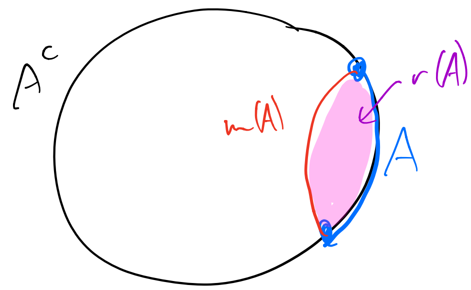
So we guess:

$$S(A) = \frac{1}{4G_N}\,\text{area}(m(A)) + \mathcal{O}(G_N^0)$$

$m(A)$ = min-area surf (on t-ref. invariant time-slice) between A + A^c

What does "between" mean? It means $\exists$ bulk region $r(A)$ that interpolates between A and $m(A)$,

i.e. such that $\partial r(A) = A \cup -m(A)$



Mathematically, $m(A)$ is homologous to A

$r(A)$ is called "homology region"

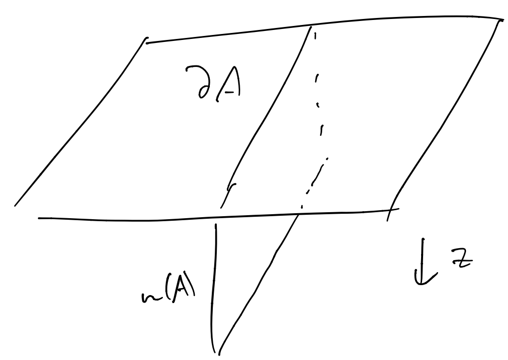
(Its causal domain is "entanglement wedge")

In black hole example, $\partial A = \phi$

In general, homology $\Rightarrow$ $\partial m(A) = \partial A$

Given $\frac{1}{z^2}$ factor in metric, min. surf. ends $\perp$ on bdy

Div. area



$$\sigma \int_\varepsilon dz \frac{1}{z^{D-1}}$$

$$= \frac{\sigma}{\varepsilon^{D-2}} \quad (D>2)$$

or $\ln \varepsilon \quad (D=2)$

Just as we saw in field theory!

First example of how this simple formula <u>geometrizes</u> physics of entanglement in field theory

Each part of minimal surface represents a different aspect of entanglement

Here, part of $m(A)$ near bdy represents UV-divergent entanglement across ∂A

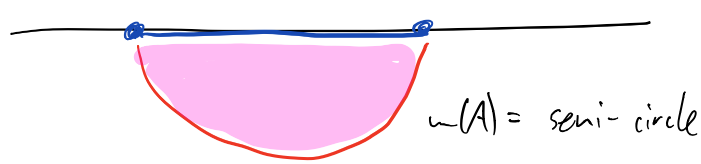
- Examples $(D=2)$

Interval in CFT

$$ds^2 = \frac{\ell^2}{z^2}\left(dx^2 + dz^2\right)$$

AdS radius (ℓ)

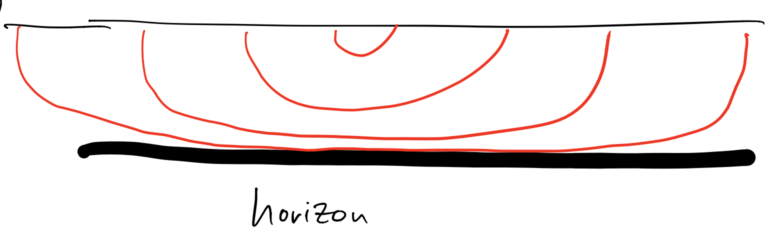
$$\frac{\ell}{G_N} = \frac{2}{3}c$$



$$\text{length} = 2\ell \ln\frac{L}{\varepsilon}$$

$$\Rightarrow \quad S(A) = \frac{c}{3}\ln\frac{L}{\varepsilon}$$

Finite temp:



$$S(A) = \frac{c}{3}\ln\frac{\sinh \pi T L}{\pi T \varepsilon}$$

For $L \ll \frac{1}{T}$, $r(A)$ stays close to bdry, does not "notice" horizon $\Rightarrow$ vac. result

For $L > \frac{1}{T}$, $r(A)$ skirts but does not intersect horizon $\Rightarrow$ linear $s(T)$ growth of $S(A)$

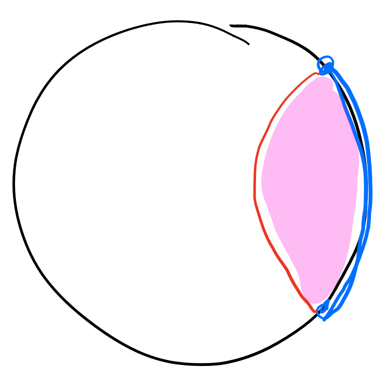
Note that minimal surfaces are nested + do not cross horizon

This nesting is general: $A \subset A' \Rightarrow r(A) \subset r(A')$

This gives us license to ignore what's on other side of horizon. This is good — $S(A)$ should not depend on choice of purification of ρ_A

There are reasons to believe that $r(A)$ is the dual of ρ_A

Circle:



Again, reproduces gen result:

$$S(A) = \frac{c}{3} \ln \frac{2R}{\varepsilon} \sin \frac{L}{2R}$$

[Actually, all of the examples so far are cheats, since they are all black holes in different coordinate systems, so we're just using BH. But all the qualitative features hold in higher D and with less sym]

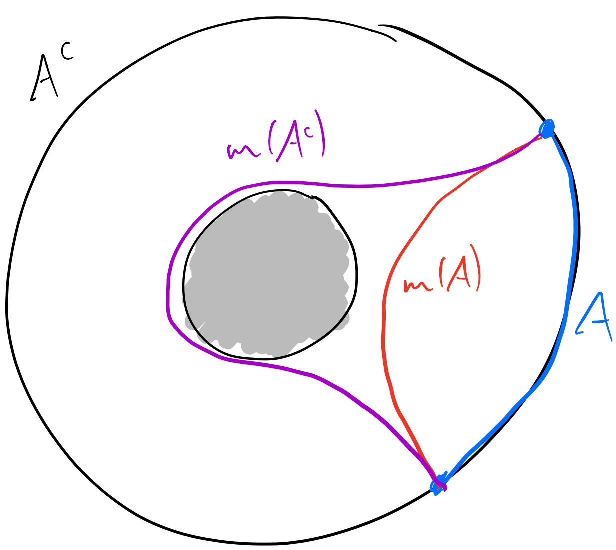
How do we get $S(A^c) = S(A)$ here? Because

$m(A^c) = -m(A)$. $A \sim -A^c$, so if $m \sim A$

then $-m \sim A^c$

But if state is mixed, we should <u>not</u> get

$S(A^c) = S(A)$ always

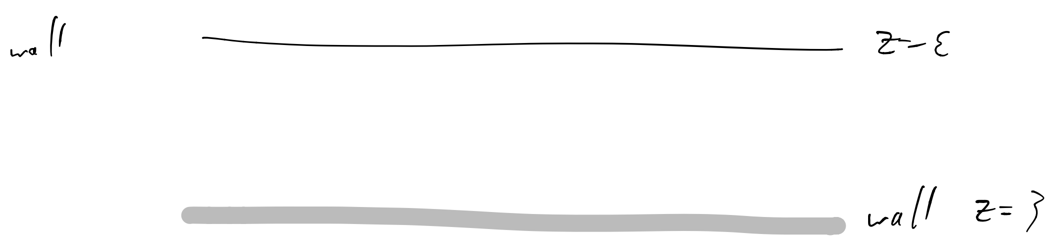


Because of horizon, A is <u>not</u> homologous to A^c

So $m(A)$, $m(A^c)$ are in different topological classes $\Rightarrow$ in general $S(A) \neq S(A^c)$

Capped theory

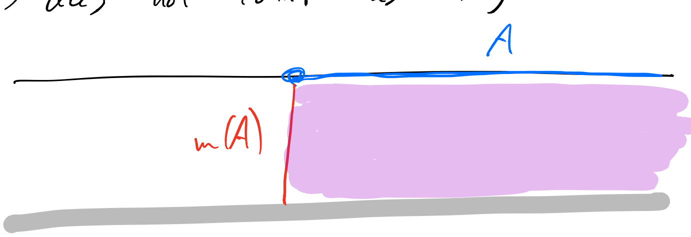
Toy model: cut off z at $z=\xi$ w/ "confining"



does not carry entropy

$\Rightarrow$ does not count as "holy" in homology cond.

Half-line

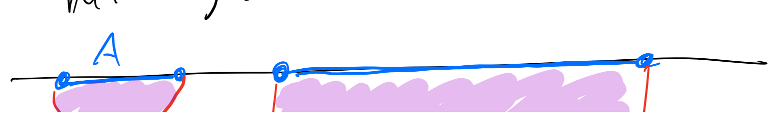


$$S(A) = \frac{1}{4G_N} \int_{\varepsilon}^{\xi} dz \, \frac{\ell}{z} = \frac{c}{6} \ln \frac{\xi}{\varepsilon}$$

as expected from field th.

min. surf. is cut off by IR cutoff of bulk geom.

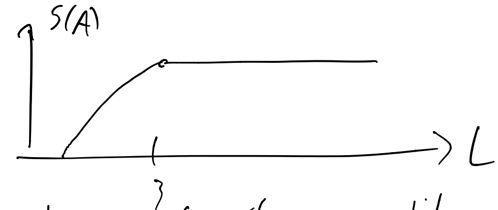
Interval:



$L < \xi$: $S(A) = \frac{c}{3} \ln \frac{L}{\epsilon}$

$L > \xi$: $S(A) = \frac{c}{3} \ln \frac{\xi}{\epsilon}$

Phase transition!



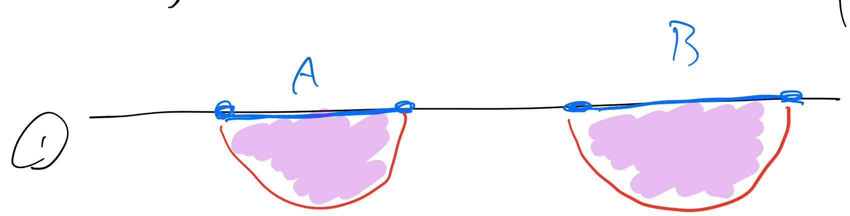
This is a large-c phase transition — like Hawking-Page, finite-volume large N deconfinement

Must be smoothed out by non-perturbative quantum corrections — not understood

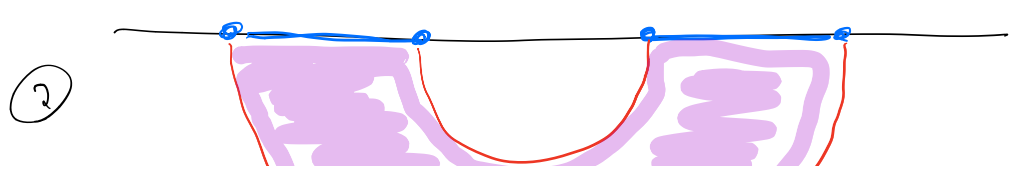
2 intervals in CFT vac

2 locally minimal surfaces for AB:

RT as generalization of BH!



$m(AB) = m(A) \cup m(B)$

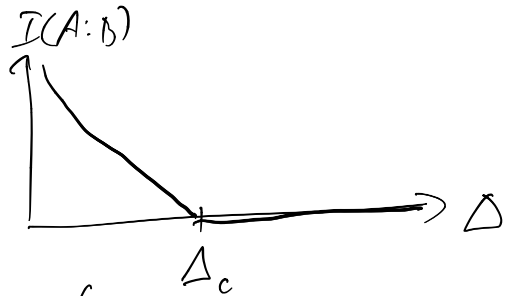


① is shorter when A, B are far apart

② when close together

① $\Rightarrow I(A:B)=0$

② $\Rightarrow I(A:B) > 0$



Again, a large-c phase transition

Obeys all properties of $I(A:B)$ predicted on general grounds before, <u>except</u> $I(A:B) > 0$

$$I(A:B)=0 \Rightarrow \rho_{AB} = \rho_A \otimes \rho_B \Rightarrow \langle O_A O_B \rangle_c = 0$$

which is not true

Have to include quantum corrections

$$I(A:B) = \mathcal{O}(c^0)$$

↑ can be calculated, are $\neq 0$

Notice that existence of config ① enforces

$$I(A:B) \geq 0 \qquad \text{subadditivity}$$

- General properties

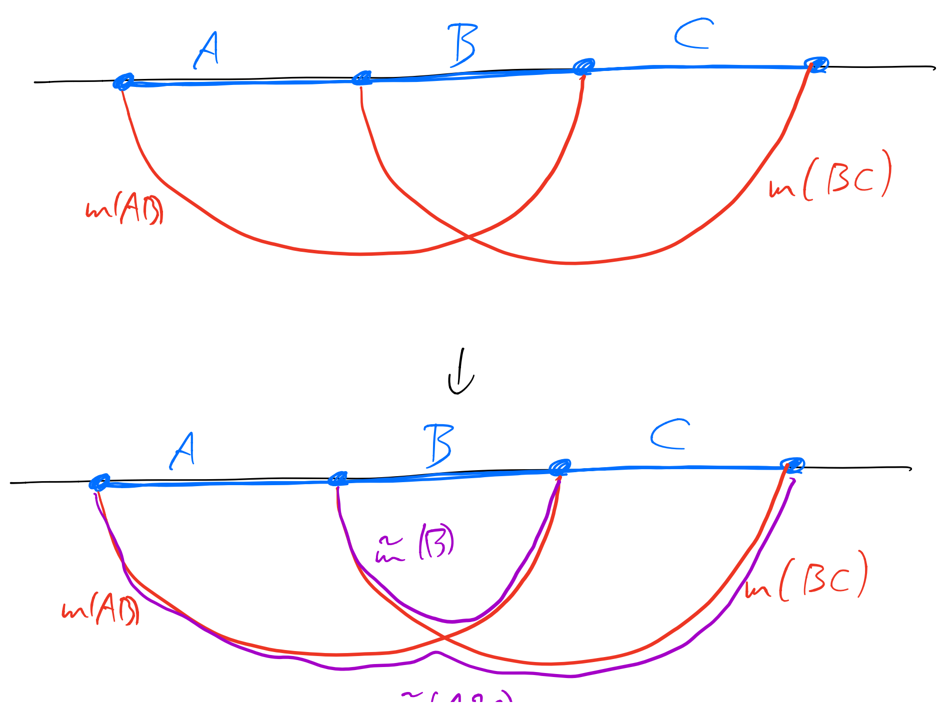
We already saw that the minimal surface automatically realizes two important properties of EE:

$$\rho \text{ pure} \Rightarrow S(A) = S(A^c)$$

$$S(AB) \leq S(A) + S(B)$$

Similarly it automatically realizes the much more difficult <u>strong</u> subadditivity inequality:

$$S(AB) + S(BC) \geq S(B) + S(ABC)$$



$m(ABC)$

$$S(AB) + S(BC) = \frac{1}{4G_N}\left[\text{area}\left(m(AB)\right) + \text{area}\left(m(BC)\right)\right]$$

$$= \frac{1}{4G_N}\left[\text{area}\left(\tilde{m}(B)\right) + \text{area}\left(\tilde{m}(ABC)\right)\right]$$

$$\geq \frac{1}{4G_N}\left[\text{area}\left(m(B)\right) + \text{area}\left(m(ABC)\right)\right]$$

$$= S(B) + S(ABC)$$

Note: $\tilde{m}(B) \sim B$ via $\tilde{r}(B) = r(AB) \cap r(BC)$

$\tilde{m}(ABC) \sim ABC$ via $\tilde{r}(ABC) = r(AB) \cup r(BC)$

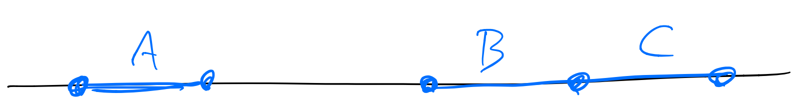
RT also obeys properties that are <u>not</u> general properties of EEs

The simplest of these is superadditivity of mutual info (MMI):

$$S(AB) + S(BC) + S(AC) \geq S(A) + S(B) + S(C) + S(ABC)$$

i.e. $\quad I(A:BC) \geq I(A:B) + I(A:C)$

e.g.



Ex: Can happen that $I(A:B) = I(A:C) = 0$

$$I(A:BC) > 0$$

Does not generally hold for quantum or even classical states, e.g. $\rho_{ABC} = \frac{1}{2}\left(|000\rangle\langle 111| + |111\rangle\langle 111| \right)$

III. Covariant holographic EE

So far we've restricted to T-reversal invariant spacetimes + regions on fixed slice. (Note that this includes all static spacetimes, which is most common case studied.)

In general $\nexists$ special slice Σ on which we can look for a minimal surface. Can try simply

$$m(A) = \text{min. surf.} \sim A$$

without restriction to any slice.

Doesn't work — can make arbitrarily small by wiggling in t-direction

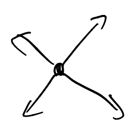
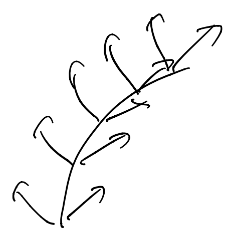
Two possible natural ways to resolve this:

1) $m(A)$ = extremal surf

$:=$ extremum of area functional

if $\exists$ several, take 1 w/ least area

2 other ways to express extremal:

a)  expansion of null geodesics $= 0$ in all 4 directions 

b) extrinsic curvature tensor

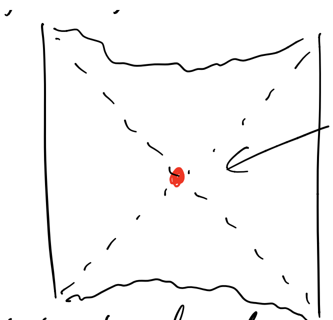
$$K^{\mu}{}_{\nu\lambda} \qquad K^{\mu}{}_{\nu\lambda}\, g^{\nu\lambda} = 0$$

$\mu \in$ normal to m

$\nu\lambda$ tangent to m

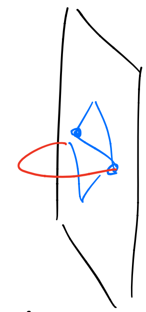
(generalization of $K=0$ for hypersurf.)

Hubeny-Rangamani-Takayanagi (HRT) formula



bifurcate horizon is extremal

so this case works

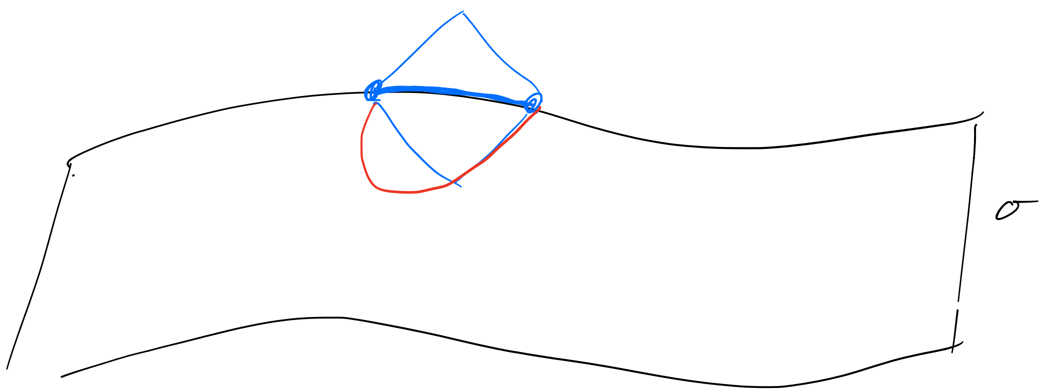


Notice that it depends only on $D(A)$

2) Minimize area over space, maximize over time

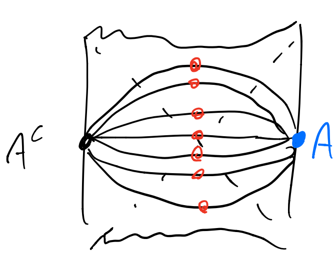
$$\max_{\sigma \supset \partial A} \ \min_{m \sim \text{ on } D(A) \text{ in } \sigma} \ \text{area}(m)$$

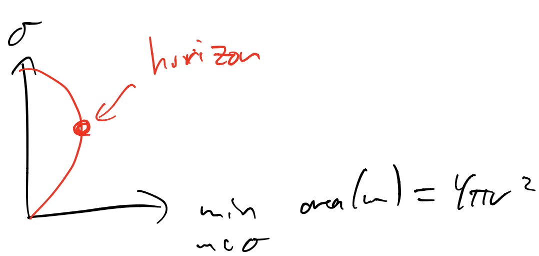
↑ bulk Cauchy slice (again, depends only on $D(A)$)



If σ wiggles in time, $\min_{m \subset \sigma}$ area(m) will be small,

so σ tries to be smooth





"Maximin"

Both seem reasonable. Which is right?

Luckily we don't have to decide — they're equivalent

Specifically, Wall showed

1) A maximin surface exists (under certain conditions)

2) The maximin surf. is the HRT surface

Bonus: HRT surf. exists!

We won't discuss proof of (1)

Proof of (2) is nice; we'll give outline

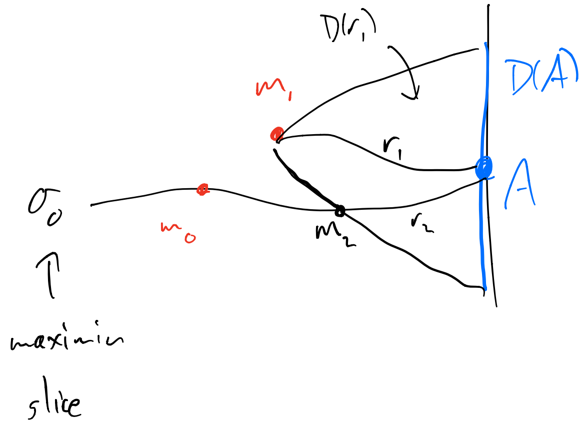
Two parts:

2a) Maximin surf. m_0 is extremal

Intuition: extremum wrt. variations in space + time directions $\Rightarrow$ wrt. variations in all directions

2b) Any <u>other</u> extremal surf. m_1 has

$$\mathrm{area}(m_1) \geq \mathrm{area}(m_0)$$



Lemma: $D(r_1)$ meets bdy on $D(A)$

$\Rightarrow m_2 \sim D(A) \cap \sigma_0$ via $r_2 := D(r_1) \cap \sigma_0$

By Raychaudhuri + Einstein + NEC, $\dot{\theta} \leq 0$

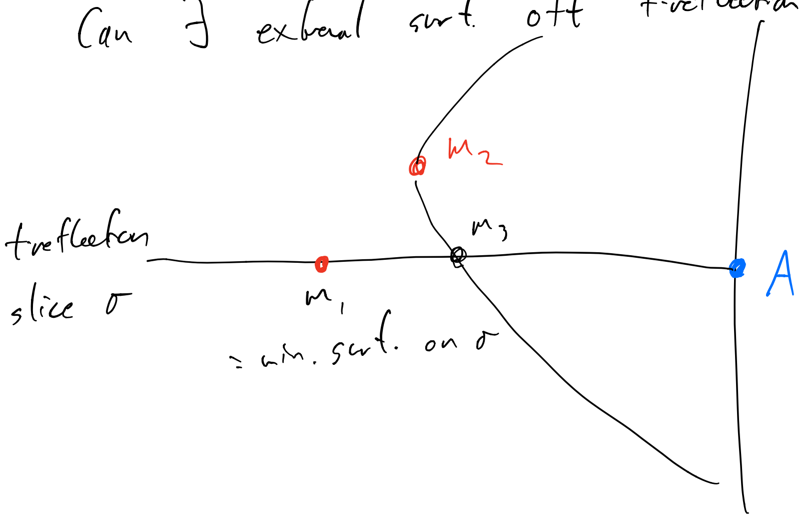
but $\theta = 0$ on $m_1 \Rightarrow \theta \leq 0 \Rightarrow \text{area}(m_2) \leq \text{area}(m_1)$

Since m_0 is minimal on σ, $\text{area}(m_2) \geq \text{area}(m_0)$

By a similar argument, can show flat HRT $\Rightarrow$ RT

RT surface is automatically extremal

Can $\exists$ extremal surf. off t-reflection invariant slice?



$m_3 \sim A$

$\text{area}(m_3) \geq \text{area}(m_1)$

$\text{area}(m_2) \geq \text{area}(m_3)$

$\Rightarrow \text{area}(m_2) \geq \text{area}(m_1)$

SSA: Static proof doesn't apply because

$m(AB)$, $m(BC)$ aren't on same slice

Lemma: $\exists$ common maximin slice σ for

B + ABC (nesting of maximin slices $\Rightarrow$ nesting of entanglement wedges)

On σ, find min. surf. $\tilde{m}(AB)$, $\tilde{m}(BC)$

By RT SSA argument,

$$\text{area}(\tilde{m}(AB)) + \text{area}(\tilde{m}(BC)) \geq S(B) + S(ABC)$$

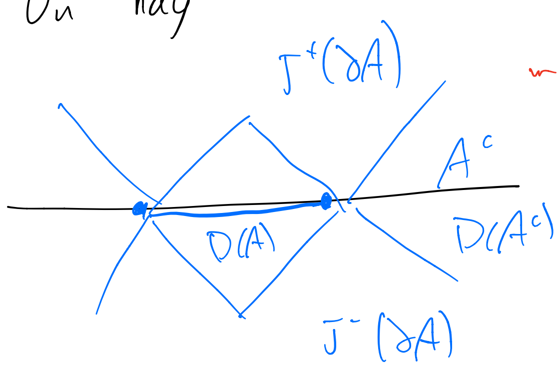
By maximin,

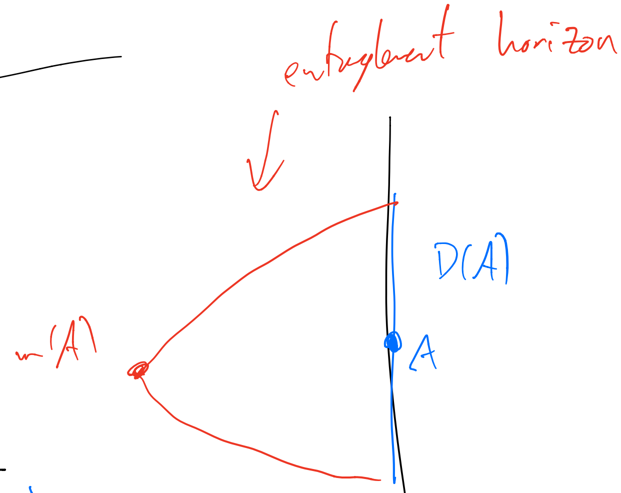
$$S(AB) \geq \text{area}(\tilde{m}(AB))$$

$$S(BC) \geq \text{area}(\tilde{m}(BC))$$

Causality:

On bdy





$$J^{\pm}(m(A)) \cap \text{bdy} = J^{\pm}(\partial A)$$

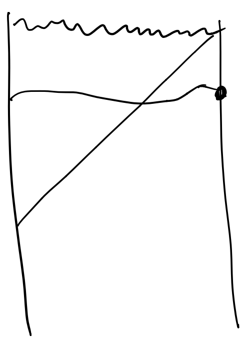
$\Rightarrow$ only bulk signals sent from $J^-(\partial A)$ can reach $m(A)$, possibly changing $S(A)$

So HRT respects bdy causality

(Essentially a version of Gao-Wald for extremal surfaces)

In spacetime w/ black hole formed from collapse,

$S(A^c) = S(A)$: $\quad \Sigma \sim \emptyset$ by bulk Cauchy slice



Σ = bdy Cauchy slice

$\Rightarrow A^c \sim A$

In particular, $S(\Sigma) = 0$

More generally, since $S(A)$ depends only on $D(A)$, if

A = whole connected component, then $S(A)$ is t-indep.

So HRT gives entropy wrt unitarily-evolving "fine-grained" density matrix