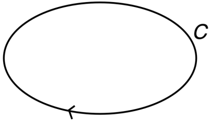
Algebras and Entropies For Black Holes

Edward Witten, IAS

I will talk about the paper "Gravity and the Crossed Product" (arXiv:2112.12828) and "An Algebra of Observables for de Sitter Space" (arXiv:2206.10780) with V. Chandrasekharan, R. Longo, and G. Penington.

I will start with a motivating example from the field of affine Lie algebras. An affine Lie algebra is a central extension of the Lie algebra of maps of a circle C to an ordinary Lie algebra $\mathfrak{g}$ (usually simple or semi-simple)

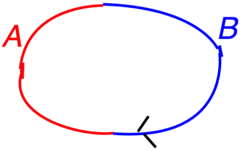


If a Hilbert space $\mathcal{H}$ furnishes a representation of an affine Lie algebra $\widehat{\mathfrak{g}}$, then to every $\mathfrak{g}^*$-valued function $f : S^1 \to \mathfrak{g}^*$, there is an operator $J(f)$ on $\mathcal{H}$ that physicists denote as

$$J(f) = \oint_C (J, f),$$

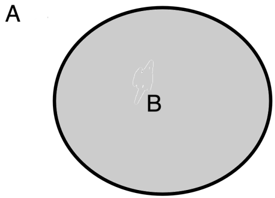
where J is the "'current." Bounded functions of operators $J(f)$ generate the Type I von Neumann algebra of all bounded operators on $\mathcal{H}$.

Suppose however that we divide the circle into two pieces A and B



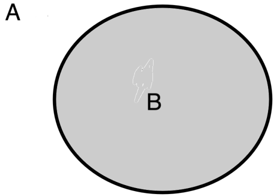
and consider only functions f with support in, say, region A. Operators $J(f)$ for such f generate what is called a von Neumann algebra of Type III. In case this notion is not familiar, I will explain enough later to make the talk understandable.

This is actually a special case of a much more general phenomenon, which holds for quantum field theory in any dimension. A quantum field theory in D spacetime dimensions – therefore $d = D - 1$ space dimensions – has a Hilbert space $\mathcal{H}$, and if we consider all of the quantum fields anywhere in space, they generate a Type I von Neumann algebra of all bounded operators on $\mathcal{H}$. But if we divide space into two disjoint regions A and B



and consider only operators in region A, then we get an algebra of Type III. This was shown by H. Araki in the 1960's (for the case of free field theory).

Why would one care about this as a physicist? The basic motivation comes from black holes. In this application



region A corresponds to the region of space that is "outside the horizon," causally accessible to an outside observer, and region B is the "interior" of the black hole, the region "behind the horizon." Thus, the operators available to us if we live outside the black hole horizon are the operators in region A. As I have just explained, in ordinary quantum field theory, this is a Type III algebra.

Gravity, however, does not fit into the framework of ordinary quantum field theory. Trying to provide a framework that in some sense extends or refines quantum field theory and in which gravity can fit is the main goal of string theory. The main point that I want to convey in this talk is that there are reasons to believe that when gravity is taken into account, the operators outside the black hole horizon actually generate an algebra of Type II, and that this helps resolve some of the puzzles about quantum black holes. Again, I will try to say a little about what is a Type II algebra. First we have to review the idea of "black hole thermodynamics," a subject that is by now 50 years old.

Jacob Bekenstein (1972), inspired by questions from his advisor John Wheeler, asked what the Second Law of Thermodynamics means in the presence of a black hole.

The Second Law says that, for an ordinary system, the “entropy” can only increase. However, if we toss a cup of tea into a black hole, the entropy seems to disappear. Bekenstein wanted to “generalize” the concept of entropy so that the Second Law would hold even in the presence of a black hole. For this, he wanted to assign an entropy to the black hole.

He needed a property of a black hole that can only increase. It is actually not true that the mass or energy of a black hole can only increase. But at the time that Bekenstein was working, Stephen Hawking had just proved the "area theorem," which says that in classical General Relativity, the area of a black hole horizon can only increase. It was therefore fairly natural for Bekenstein to propose that the *area* A *of the black hole horizon* should represent, in some sense, a contribution to the entropy. To be more exact, since entropy is dimensionless but A of course has dimensions of area (or length^2 if we are in three space dimensions), Bekenstein proposed that a multiple of

$$\frac{\mathsf{A}}{G\hbar}$$

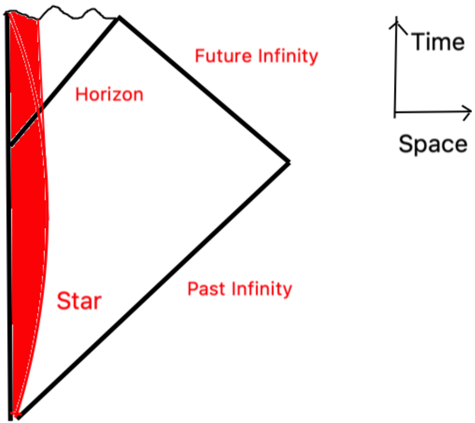
should be viewed as a contribution to the entropy (here G is Newton's constant and $\hbar$ is Planck's constant). Hawking later showed that this should actually be $A/4G\hbar$.

Bekenstein proposed that the quantity that satisfies the second law and always increases is not the ordinary entropy of matter and radiation outside a black hole, which I will call $S_{\rm out}$, but rather a "generalized entropy" which is the sum of $A/4G\hbar$ and $S_{\rm out}$:

$$S_{\rm gen} = \frac{\mathsf{A}}{4G\hbar} + S_{\rm out}.$$

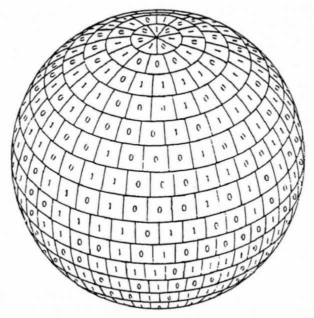
The idea is that the "correct" quantity to which the second law applies should really be the generalized entropy. When we toss a cup of tea into a black hole $S_{\rm out}$ goes down but $\mathsf{A}/4G\hbar$ goes up by more.

Supposedly, Stephen Hawking was skeptical of Bekenstein's idea and set out to disprove it by studying the behavior of a quantum field interacting with a black hole.



But he ended up proving that Bekenstein was right, by finding that at the quantum level a black hole is not really black but has a temperature of order $\hbar$.

For an ordinary system, the "entropy" is a measure of the number of degrees of freedom – or more precisely, the number of degrees of freedom that are relevant at a given temperature. Many researchers have thought that, somehow, the entropy $S = \mathrm{A}/4G$ means that the black hole horizon can be described by some sort of degrees of freedom that live at its surface – one bit or qubit for every $4G$ of area. For example, in a famous article in 1992, John Wheeler illustrated that idea with this picture:



Black hole thermodynamics has been spectacularly successful – it turns out that subtle properties of classical General Relativity work out in such a way as to ensure that the generalized entropy does behave like a thermodynamic entropy. For example, the Hawking area theorem motivated Bekenstein's idea in a way I already explained, and is a key step in proving that S_{gen} (if S_{out} is properly defined) does obey the second law

$$\frac{\mathrm{d}S_{\mathrm{gen}}}{\mathrm{d}t} \geq 0.$$

(The most complete proof is by A. Wall (2011) and makes use of Tomita-Takesaki theory of von Neumann algebras.) Other subtle properties of classical General Relativity work out in such a way that the first law of thermodynamics is also satisfied

$$\mathrm{d}E = T\mathrm{d}S + \Phi\mathrm{d}Q + \Omega\mathrm{d}J.$$

There is much more besides. There is abundant evidence that $S_{\rm gen}$ behaves as an entropy. But is it an entropy of something? This has been a mystery since the early days of black hole thermodynamics. In today's talk, I will explain a slightly abstract answer: with gravity taken into account, the operators outside a black hole horizon form a Type II algebra, and generalized entropy is the entropy of a state of this algebra.

Let me first explain a little about the meaning of “entropy” in quantum mechanics. When we are making an observation or analyzing an experiment, we usually study not the whole universe but a small subsystem, consisting possibly of the experimental apparatus or possibly (if we are doing astronomy) the Milky Way galaxy. Let me refer generically to the system we are studying as system A, and refer to the rest of the universe as B. So the combined system AB is the whole universe.

Usually in quantum mechanics, one can assume that the subsystems A and B can be described by Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$; the Hilbert space of the combined system is then a tensor product

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B.$$

In such a situation, one can define an algebra $\mathcal{A}$ of operators of system A – that is, operators on $\mathcal{H}_A$ – and an algebra $\mathcal{B}$ of operators of system B – that is, operators on $\mathcal{H}_B$. These are Type I von Neumann algebras (the algebra of all bounded operators on a Hilbert space, here $\mathcal{H}_A$ or $\mathcal{H}_B$). The property of a Type I algebra that we will use is that there is a nondegenerate bilinear form on $\mathcal{A}$ given by

$$(\mathsf{a}, \mathsf{a}') = \mathrm{Tr}_{\mathcal{H}_A}\, \mathsf{a}\mathsf{a}'.$$

Consider any state of the whole universe, meaning any vector $\Psi \in \mathcal{H}_{AB}$. We want to define the entropy of system A when the whole universe is in the state Ψ. Consider the linear function $F : \mathcal{A} \to \mathbb{C}$ defined by

$$F(\mathsf{a}) = \langle \Psi | \mathsf{a} | \Psi \rangle.$$

Remembering that algebra $\mathcal{A}$ has a nondegenerate bilinear form, we see that the function $F(\mathsf{a})$ is

$$F(\mathsf{a}) = \mathrm{Tr}\, \mathsf{a} \rho$$

for some unique $\rho \in \mathcal{A}$. It is not hard to show that ρ is self-adjoint and nonnegative. Moreover, if Ψ is a unit vector, meaning that $\langle \Psi | \Psi \rangle = 1$, then

$$\mathrm{Tr}\, \rho = \mathrm{Tr}\, 1 \cdot \rho = \langle \Psi | 1 | \Psi \rangle = \langle \Psi | \Psi \rangle = 1.$$

An element $\rho \in \mathcal{A}$ that is self-adjoint, non-negative and has trace 1 is called a density matrix. So we have learned that any state Ψ of the whole universe determines a density matrix for the subsystem A. Conversely, if $\rho \in \mathcal{A}$ is a density matrix, it is not hard to show that it is the density matrix of some Ψ (assuming that system B is big enough).

When system A is described by a density matrix ρ, its entropy is defined to be

$$S(\rho) = -\mathrm{Tr}\,\rho \log \rho.$$

This formula is due to von Neumann and is called the von Neumann entropy; in the limit of classical mechanics, it goes over to a classical formula for entropy due to Gibbs (extending earlier work of Boltzmann). It is not hard to prove that $S(\rho)$ is zero if and only if ρ is of rank 1; otherwise it is positive. Under appropriate definitions, when thermodynamics is valid, the von Neumann entropy is equivalent to thermodynamic entropy; however, it is defined universally.

For any density matrix ρ, the function

$$F(\mathsf{a}) = \mathrm{Tr}\, \mathsf{a}\rho$$

satisfies

$$F(\mathsf{a}^\dagger \mathsf{a}) \geq 0, \quad \mathsf{a} \in \mathcal{A}.$$

It defines what is called a "state" of the algebra $\mathcal{A}$. A state is called "pure" if it is not a convex linear combination of other states

$$\rho \neq t\rho_1 + (1-t)\rho_2$$

where ρ_1, ρ_2 are density matrices and $0 < t < 1$. It is not hard to prove that ρ is a pure state by that definition if and only if it has rank 1, equivalently, if it has entropy 0. Thus pure states have entropy 0 and "mixed" states (states that are not pure) have positive entropy.

In case ρ is a state of system A which arises as the density matrix of a vector

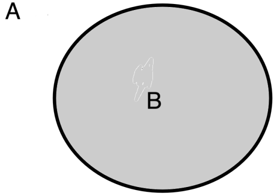
$$\Psi \in \mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B,$$

it is not hard to prove that ρ is a pure state of system A if and only if the vector Ψ is a tensor product

$$\Psi = \Psi_A \otimes \Psi_B, \quad \Psi_A \in \mathcal{H}_A, \ \ \Psi_B \in \mathcal{H}_B.$$

Such a Ψ is called a product state. Thus, system A has zero entropy if and only if the state of the whole universe is the tensor product of a state Ψ_A of system A and a state Ψ_B of the rest of the universe. This is a possible but atypical state of affairs.

The idea that the Bekenstein-Hawking entropy of a black hole should be understood in terms of von Neumann entropy was apparently first put forward by R. Sorkin in 1983 (in a paper that attracted only modest attention at the time). The idea was just the following. In a quantum field theory, divide space into two regions A and B

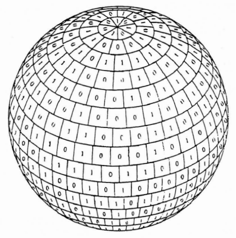


Let Ψ be a state of the system, and ρ the corresponding density matrix for the algebra $\mathcal{A}$ of operators in region region A. (This is a naive formulation and we will be more critical later.) One can try to calculate the entropy $-\mathrm{Tr}\,\rho\log\rho$. One finds that it is infinite: it is ultraviolet divergent (regardless of Ψ) and the coefficient of the leading divergence is proportional to the area A of the boundary between regions A and B.

Sorkin's idea, in modern language, was that somehow gravity cuts off the ultraviolet divergence, leaving an entanglement entropy in the vacuum between the two regions that is the Bekenstein-Hawking entropy $A/4G$, where A is the area of the boundary between them. This makes a lot of intuitive sense, as it matches two ideas:

(1) $A/4G$ is the irreducible entropy of the system for someone who has access only to the region outside the horizon

(2) the divergence in the entanglement entropy is proportional to A because it comes from short wavelength modes near the "horizon," as if (after cutting off the divergence) the density of quantum degrees of freedom on the horizon per unit area is $4G$ as in Wheeler's picture:



Susskind and Uglum (1993) made a simple observation that strongly supports this point of view. If we interpret $S_{\rm out}$ as von Neumann entropy, then the generalized entropy is better defined than either term on the right hand side is separately:

$$S_{\rm gen} = \frac{\mathsf{A}}{4G\hbar} + S_{\rm out}.$$

The second term has an ultraviolet divergence that Sorkin noted. The first term has a similar problem, because there is an ultraviolet divergence in the relation between the bare Newton constant G_0 and the physical, observed Newton constant G:

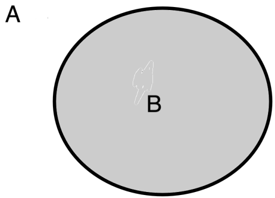
$$\frac{1}{G\hbar} = \frac{1}{G_0\hbar} + c\Lambda^2 + \cdots .$$

Here Λ is an ultraviolet cutoff and c is a constant (at 1-loop level, c is independent of $\hbar$). Susskind and Uglum argued that the ultraviolet divergences in $S_{\rm out}$ cancel those in $1/G$ (and these arguments were refined later).

Twenty-first century developments have strongly supported these ideas, though leaving us with plenty of mysteries. In the available time, I am just going to talk about one aspect of the story. Why is it that the entropy of the region outside the horizon is ill-defined in quantum field theory (so that S_{out} has that quadratic divergence that Sorkin pointed out) but well-defined once gravity is included? I will explain a somewhat abstract answer.

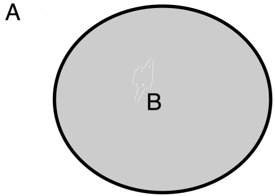
First of all, as I already explained, in ordinary quantum mechanics, when one considers a system AB made from subsystems A and B, one normally assumes at the start that each system has its own Hilbert space $\mathcal{H}_A$ or $\mathcal{H}_B$. If the combined system is in a state $\Psi = \Psi_{AB}$, the density matrix ρ of system A generically has a high rank and a positive entropy, but ρ *might* – if Ψ is a product state – be a pure state density matrix with zero entropy. Thus in ordinary quantum mechanics, whether a subsystem A of a larger system AB has a positive entropy or not depends on the choice of a state $\Psi \in \mathcal{H}_{AB}$. In case of a generic Ψ, subsystem A has a positive entropy, but if Ψ is a product state, the entropy vanishes.

That is not the situation for the entropy of a region of space in quantum field theory.



The divergence found by Sorkin was an ultraviolet divergence, so it does not depend on the state: every state looks like the vacuum at short distances.

The root of the problem is that it is not true



that there are separate Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ for the "outside" and "inside" regions. There is only a combined Hilbert space $\mathcal{H}$ for the whole system. What the separate regions A and B have are not Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, but only algebras of observables $\mathcal{A}$ and $\mathcal{B}$. These algebras act on $\mathcal{H}$ so they can be defined to be von Neumann algebras (a von Neumann algebra is an algebra of bounded operators on a Hilbert space that is closed under weak limits).

There are three types of von Neumann algebra:

(I) A Type I algebra is the algebra of all operators on a Hilbert space. In ordinary quantum mechanics, when we discuss a system A, it has a Hilbert space $\mathcal{H}_A$ and the algebra of observables of the system is the algebra of all (bounded) operators on $\mathcal{H}_A$. One can define density matrices and entropies for a system that has such an algebra of observables and there are pure states that have zero entropy.

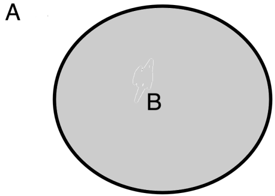
The other types may be less familiar. But first the bottom line:

(II) A Type II algebra does not have pure states, but there is a notion of density matrix and entropy for a system in which the algebra of observables is of Type II.

(iii) A Type III algebra is the "worst" type – a system whose observables form a Type III algebra does not have pure states and also does not have density matrices or entropies.

By now you might anticipate the bad news:

In quantum field theory, the algebra of observables of a region of spacetime



is always of Type III. So to a region, one can never associate a pure state, or a density matrix or entropy. The Type III nature of the algebra is the "reason" for the universal ultraviolet divergence of the entanglement entropy.

However, it turns out that including gravity in a semiclassical way changes the picture: it changes the algebra of the region outside the horizon from Type III to Type II. So when gravity is turned on (even semiclassically), the region outside the black hole horizon is described by an algebra in which the notion of entropy is well-defined, though there is no notion of a quantum mechanical microstate (a pure state of the algebra). We can interpret that as a somewhat abstract answer to the question of why including gravity suddenly enabled us to convert the ill-defined (divergent) $S_{\rm out}$ into the better defined

$$S_{\rm gen} = \frac{\mathsf{A}}{4G} + S_{\rm out}.$$

But to understand how this works, we need to understand something about these von Neumann algebras of Types II and III. A Type II algebra was originally constructed by Murray and von Neumann in the following way. A "qubit" is just a quantum system described by a two-dimensional Hilbert space. Let A and B be systems consisting of countably many qubits. Keeping only N qubits of system A and N more of system B, consider the state

$$\Psi = \frac{1}{2^{N/2}} \bigotimes_{n=1}^{N} \left(\sum_{i=1,2} |i\rangle_{A,n} \otimes |i\rangle_{B,n} \right).$$

That is, the n^{th} qubit of the system A is completely entangled with the n^{th} qubit of system B. Let a, a' be operators that act only on the first k spins of system A, for some $k \leq N$. Define a function

$$F(\mathrm{a}) = \langle \Psi | \mathrm{a} | \Psi \rangle.$$

The state Ψ was constructed so that the corresponding density matrix is a multiple of the identity, $\rho = 2^{-N} \cdot 1$. So

$$F(\mathsf{a}) = \langle \Psi | \mathsf{a} | \Psi \rangle = \mathrm{Tr}\, \mathsf{a} \rho = 2^{-N} \mathrm{Tr}\, \mathsf{a}$$

from which we see that $F(\mathsf{a})$ satisfies

$$F(1) = 1,$$

and

$$F(\mathsf{a}\mathsf{a}') = F(\mathsf{a}'\mathsf{a}).$$

For any given a, $F(\mathsf{a})$ is defined and independent of N as soon as N is big enough (as soon as we include all qubits on which a acts) so $F(\mathsf{a})$ has a large N limit.

For $N \to \infty$, the function $F(\mathsf{a})$ can be defined for any operator a that acts on any finite set of qubits in system A and of course it still satisfies

$$F(1) = 1$$

and

$$F(\mathsf{a}\mathsf{a}') = F(\mathsf{a}'\mathsf{a}).$$

F is also positive in the sense that

$$F(\mathsf{a}^\dagger \mathsf{a}) > 0 \text{ for all } \mathsf{a} \neq 0.$$

So far we have defined F on the whole algebra $\mathcal{A}_0$ of all operators that act on only finitely many qubits in system A. By taking weak limits, we can complete $\mathcal{A}_0$ to a von Neumann algebra $\mathcal{A}$, still with a function $F(\mathsf{a})$ that has the same properties I've stated. Since $F(\mathsf{a}\mathsf{a}') = F(\mathsf{a}'\mathsf{a})$ this function is usually called a trace: We formally define

$$F(\mathsf{a}) = \mathrm{Tr}\,\mathsf{a}$$

but $\mathrm{Tr}\,\mathsf{a}$ is *not* the trace of a in any Hilbert space representation. In the language of physicists, it is a renormalized trace with an infinite factor $2^{-N}\big|_{N\to\infty}$ removed.

There is a more elementary example of an infinite-dimensional algebra with a trace – the Type I algebra $\mathcal{B}$ of all operators on an infinite-dimensional Hilbert space $\mathcal{H}$. In this example, however, while we can define a trace on elements of $\mathcal{B}$, it is not defined for all elements of $\mathcal{B}$, only for those that are "trace class." For example, the identity element of $\mathcal{B}$ does not have a trace (unless one wants to allow $\mathrm{Tr}\,1 = \infty$). By contrast, from the infinite system of qubits, we constructed an algebra $\mathcal{A}$ in which every element has a trace. Clearly then it is an essentially new type of algebra. This is, in fact, the simplest example of a Type II algebra – it is the Type II_1 factor of Murray and von Neumann.

Small generalizations of this construction lead to algebras of Type III (as shown by Powers, Araki, and Wood in the 1960's). We modify the previous construction a little bit by putting the n^{th} qubit of system A and the n^{th} qubit of system B in the state

$$\frac{1}{(1+e^{-\beta/2})^{1/2}}\left(|\uparrow\rangle_A|\uparrow\rangle_B + e^{-\beta/2}|\downarrow\rangle_A|\downarrow\rangle_B\right).$$

We define a state Ψ in which, for large N, this is done for the n^{th} pair for $n = 1, 2, \cdots, N$. Then we can define the function $F(\mathrm{a}) = \langle\Psi|\mathrm{a}|\Psi\rangle$ and as before it has an $N \to \infty$ limit. The important difference is that now $F(\mathrm{a}\mathrm{a}') \neq F(\mathrm{a}'\mathrm{a})$. For $N \to \infty$, we can define an algebra $\mathcal{A}_0$ consisting of operators that act on any finite set of qubits of the A system, and its completion is now a von Neumann algebra of Type III.

Algebras of Type II or Type III do not have an irreducible representation in a Hilbert space; whenever such an algebra acts on a Hilbert space $\mathcal{H}$, it always commutes with another algebra of the same type. For example, we constructed our Type II and Type III algebras as algebras of operators on the "A" part of a bipartite system AB, so in that construction they commute with an identical algebra that acts on system B.

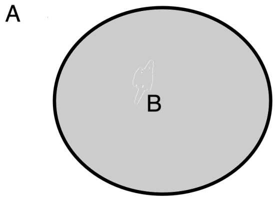
The difference between a Type II algebra and a Type III algebra is that a Type II algebra has a trace, and a Type III algebra does not.

Moreover, in a Type II algebra, the trace is nondegenerate in the sense that $(\mathsf{a}, \mathsf{a}') = \mathrm{Tr}\,\mathsf{a}\mathsf{a}'$ is a nondegenerate (and positive-definite) bilinear form on the algebra (this follows from our earlier result that $\mathrm{Tr}\,\mathsf{a}^\dagger\mathsf{a} > 0$ for all $\mathsf{a} \neq 0$). Hence if $F(\mathsf{a})$ is any linear function of $\mathsf{a} \in \mathcal{A}$, we have

$$F(\mathsf{a}) = \mathrm{Tr}\,\mathsf{a}\mathsf{a}'$$

for some unique $\mathsf{a}' \in \mathcal{A}$.

Now let us go back to the situation considered by Sorkin:



We consider some state Ψ of the whole universe. *Suppose* it were true that physics in region A is described by a Type II algebra $\mathcal{A}$. Then the linear function $\mathsf{a} \to \langle\Psi|\mathsf{a}|\Psi\rangle$ would be equal to $\mathrm{Tr}\,\mathsf{a}\rho_A$ for some $\rho_A \in \mathcal{A}$:

$$\langle\Psi|\mathsf{a}|\Psi\rangle = \mathrm{Tr}\,\mathsf{a}\rho_A.$$

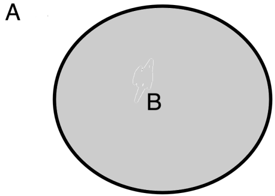
If the algebra were Type I, we would use this condition to define the density matrix ρ_A of state Ψ for measurements in region A. So it makes sense to call ρ_A the density matrix also in the Type II situation. (I. Segal, 1962; R. Longo and EW, 2021.)

Once we have density matrices, we can define entropies as well:

$$S_A = -\mathrm{Tr}\, \rho_A \log \rho_A.$$

So *if* the region outside the horizon is described by a Type II algebra, then we can define an entropy for this region.

As I have already explained, in ordinary quantum field theory the algebras



are Type III. But it turns out that when we include gravity, things are different: gravitational effects even for very weak coupling convert the Type III algebras into Type II algebras.

The mathematical mechanism leading to this is quite simple and was developed by Connes and Takesaki in the 1970's; what is new is only that this mechanism is actually implemented by gravity in the field of a black hole. The details are different in the two cases and I really will only have time for very brief explanations.

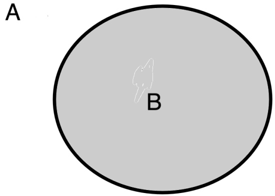
The motivation of Connes and Takesaki was simply that Type III algebras are difficult to study. It turns out that if $\mathcal{A}$ is a Type III_1 algebra (the generic Type III algebra is of this type) then there is an associated Type II_∞ algebra $\widehat{\mathcal{A}}$ that can be canonically constructed from $\mathcal{A}$ and from which $\mathcal{A}$ can be reconstructed ("up to multiplicity"). The existence of $\widehat{\mathcal{A}}$ proved to be useful as a tool for studying $\mathcal{A}$.

The definition is as follows. Suppose that $\mathcal{A}$ is a von Neumann algebra that acts on a Hilbert space $\mathcal{H}$ with a cyclic separating vector Ψ. (A suitable $\mathcal{H}$ and Ψ always exist.) Let Δ_Ψ be the modular operator of Ψ in the sense of Tomita-Takesaki theory, and set $H = -\log \Delta_\Psi$. Let $\widehat{\mathcal{H}} = \mathcal{H} \otimes L^2(\mathbb{R})$, where we think of $L^2(\mathbb{R})$ as the space of square-integrable functions of a "new" real variable X. Then one defines the "crossed product" algebra

$$\widehat{A} = \{\mathcal{A}, H + X\}'',$$

that is, the von Neumann algebra generated by $\mathcal{A}$ and (bounded functions of) $H + X$. (It is called the crossed product of $\mathcal{A}$ with its modular automorphism group.)

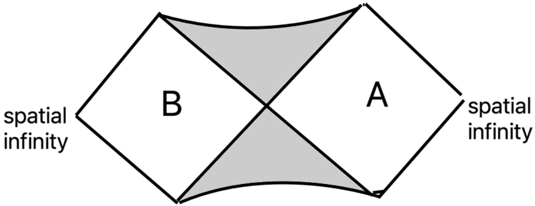
This construction has many remarkable properties. The definition of $\widehat{\mathcal{A}}$ made use of a cyclic separating vector $\Psi \in \mathcal{H}$, but one can show that $\widehat{\mathcal{A}}$ is independent of Ψ, up to a canonical isomorphism. If $\mathcal{A}$ is of Type III_1 – the usual situation in quantum field theory –



then $\widehat{\mathcal{A}}$ is of Type II_∞ and there is an explicit formula for the trace function on $\widehat{\mathcal{A}}$.

So to get from ordinary quantum field theory where we cannot define the entropy of a region (or we can define it and say that it is $+\infty$) to gravity where we can define such an entropy and get a finite answer, we just need to know that gravity adds one variable in the construction of the Hilbert space, namely what I called X, and one generator of the algebra of operators outside the black hole, namely $H + X$.

Here is a Penrose diagram of the maximally extended Schwarzshild black hole in asymptotically flat spacetime:



The black hole is a “wormhole” that connects two asymptotically flat universes, which are our two systems A and B. The extra operator that is accessible to the observer on the right and that corresponds to what I called $H + X$ earlier is H_R, the ADM energy measured at infinity on the right side. What I called X is (up to a scalar multiple) H_L, the ADM energy measured at infinity on the right side.

For the cyclic separating vector Ψ, we can take the Hartle-Hawking state of the black hole. (This state depends on a choice of temperature $1/\beta$ which determines the mass of the black hole we are going to study.) The modular operator Δ_Ψ of this state was determined by Sewell (1982), reinterpreting classic results of Unruh and of Bisognano and Wichman. The result was that $H = -\log \Delta_\Psi$ satisfies

$$\beta H_R - \beta H_L = H.$$

Equivalently

$$\beta H_R = H + \beta H_L.$$

Thus setting $X = \beta H_L$, we see that gravity is making the operator $H + X$ accessible to an observer in the right exterior region outside the horizon. The algebra of observables for this observer is thus not the Type III algebra $\mathcal{A}$ that we would have in ordinary quantum field theory, but is the "crossed product" algebra $\widehat{\mathcal{A}}$ of Type II_∞. This gives a framework for interpreting black hole entropy: it is just the entropy of a state of the Type II_∞ algebra.

There is a similar story for cosmological horizons; this is the topic of the second paper with Chandrasekharan, Longo, and Penington.