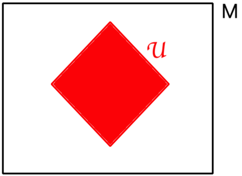
A Background Independent Algebra in Quantum Gravity

Edward Witten

Lecture at Strings 2023

In ordinary quantum field theory without gravity in a spacetime M, we can associate an algebra $\mathcal{A}_{\mathcal{U}}$ of observables to any open set $\mathcal{U} \subset M$:

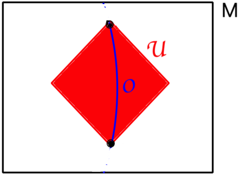


There are a few problems with this notion in the presence of gravity. I will point out three. The most obvious problem is that since spacetime fluctuates, it is in general difficult to describe a spacetime region in quantum gravity. We do not have the same freedom in doing this that we have in the absence of gravity.

A possibly deeper problem concerns background independence. In ordinary quantum field theory, the algebra $\mathcal{A}_{\mathcal{U}}$ that we associate to an open set $\mathcal{U} \subset M$ depends on M and $\mathcal{U}$, of course, but it does not depend on the state of the quantum fields. What would be the analog of that in gravity? In gravity, the spacetime M is part of what the fields determine, so an algebra that doesn't depend on the state of the quantum fields should be defined universally, independently of M. By contrast, anything we define as the algebra of the observables in a region $\mathcal{U} \subset M$ will depend on the choice of M and $\mathcal{U}$.

A third problem concerns the question of why we want to define an algebra in the first place – what is this algebra supposed to mean? In ordinary quantum mechanics, an observer is external to the system and we are quite free to make what assumptions we want about the capability of the observer. In quantum field theory without gravity, we can imagine an observer who can probe a system at will but only in a specified region $\mathcal{U} \subset M$, and that is the context in which it makes sense to consider the algebra $\mathcal{A}_\mathcal{U}$. In gravity, at least in a closed universe or in a typical cosmological model, there is no one who can probe the system from outside so an algebra only has operational meaning if it is the algebra of operators accessible to some observer.

Following Unruh (1976) and many others, I will model an observer by a timelike worldline (which I will take to be a geodesic) and I will assume that what the observer can measure are the quantum fields along the worldline. According to the "timelike tube" theorem (Borchers 1961; Araki 1963; Strohmaier 2000; Strohmaier and EW 2023), in quantum field theory without gravity, the algebra of operators along the worldline is equivalent to the algebra of operators in a certain open set:



So the algebra of operators along a timelike geodesic is a reasonable substitute for what we usually consider in the absence of gravity, and makes more sense when gravity is included.

Some relevant papers on algebras in quantum gravity:

Algebras of operators outside a black hole horizon

Leutheusser and Liu (2021)
EW (2021), Chandrasekharan, Penington, and EW (2022)

Algebra for a static patch in de Sitter space:

(*) Chandrasekharan, Longo, Penington, and EW (2022)

In JT gravity with negative cosmological constant

Penington and EW (2023), Kolchmeyer (2023)

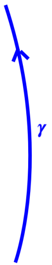
In a general diamond-like region

Jensen, Sorce, and Speranza (2023).

My starting point for today will be to reinterpret the CLPW paper on the static patch in de Sitter space. In that paper, the goal was to define an algebra for the static patch. Because of the symmetries of the static patch, it was necessary to assume that there was an observer in the static patch: then one could define an algebra by "gravitationally dressing" an operator to the observer's worldline. This logic does not apply for more general spacetimes: in a generic spacetime with less symmetry, one could "gravitationally dress" an operator to a feature of the spacetime, or of the state. Today we will consider the same construction with a different motivation: background independence.

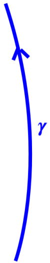
We expect that in a full theory of quantum gravity, an observer cannot be introduced from outside but must be described by the theory. What it means then to assume the presence of an observer is that we define an algebra that makes sense in a subspace of states in which an observer is present. We don't try to define an algebra that makes sense in all states.

First let us describe the situation in the absence of gravity. The observer propagates in a spacetime M on a geodesic γ:



The worldline is parametrized by proper time τ. The observer measures along γ, for example, a scalar field ϕ, or the electromagnetic field $F_{\mu\nu}$, or the Riemann tensor $R_{\mu\nu\alpha\beta}$, as well as their covariant derivatives in normal directions.

Focus on a particular observable, say $\phi(x(\tau))$ for a scalar field ϕ; I will abbreviate this as $\phi(\tau)$. When we take gravity to be dynamical, we have to consider that the same worldline can be embedded in a given spacetime in different ways, differing by $\tau \to \tau + \mathrm{constant}$:



So $\phi(\tau)$ isn't by itself a meaningful observable: we need to introduce the observer's degrees of freedom and define τ relative to the observer's clock.

In a minimal model, we equip the observer with a Hamiltonian $H_{\rm obs} = mc^2 + q$, and a canonical variable $p = -{\rm i}\frac{\rm d}{{\rm d}q}$. However, it turns out that it is better to assume that the observer energy is bounded below, say $q \geq 0$ (so m is the observer's rest mass). We then only allow operators that preserve this condition, so for example $e^{{\rm i}p}$, which does not preserve $q \geq 0$, should be replaced with $\Pi e^{{\rm i}p}\Pi$, where $\Pi = \Theta(q)$ is the projection operator onto $q \geq 0$.

We now want to allow only operators that commute with

$$\widehat{H} = H_{\rm bulk} + H_{\rm obs},$$

where $H_{\rm bulk}$ is (any) gravitational constraint operator that generates a shift of τ along the worldline. Since

$$[H_{\rm bulk}, \phi(\tau)] = -{\rm i}\dot{\phi}(\tau),$$

we need

$$[q, \phi(\tau)] = {\rm i}\dot{\phi}(\tau),$$

which we can achieve by just setting

$$\tau = p$$

or more generally

$$\tau = p + s$$

for a constant s.

So a typical allowed operator is $\phi(p+s)$, or more precisely

$$\widehat{\phi}_s = \Pi\phi(p+s)\Pi = \Theta(q)\phi(p+s)\Theta(q).$$

In addition to these operators (with ϕ possibly replaced by any local field along the worldline such as the electromagnetic field or the Riemann tensor) there is one more obvious operator that commutes with $\widehat{H}$, namely q itself. So we define an algebra $\mathcal{A}_{\rm obs}$ that is generated by the $\widehat{\phi}_s$ as well as q.

The setup hopefully sounds "background independent," since we described it without picking a background. However, background independence really depends on interpreting the formulas properly. We will not get background independence if we interpret $\widehat{\phi}_s$ and q as Hilbert space operators. To get a Hilbert space on which $\widehat{\phi}_s$ and q act, we have to pick a spacetime in M which the observer is propagating. Then we won't have background independence. The algebras for different M's are inequivalent representations of the same underlying operator product algebra. To get background independence, we have to think of $\mathcal{A}_{\mathrm{obs}}$ as an operator product algebra, rather than an algebra of Hilbert space operators.

In the absence of gravity, we would characterize the objects $\phi(\tau)$ by their universal short distance singularities:

$$\phi(\tau)\phi(\tau') \sim C(\tau - \tau' - \mathrm{i}\epsilon)^{-2\Delta} + \cdots .$$

This characterization does not require any knowledge about the quantum state. After coupling to gravity and including the observer and the constraint, the short distance expansion in powers of $\tau - \tau'$ becomes an expansion in $1/q$. We characterize $\mathcal{A}_{\rm obs}$ purely by the universal short distance or $1/q$ expansion of operator products. With that understanding, $\mathcal{A}_{\rm obs}$ is background-independent.

By a “state” of the observer algebra $\mathcal{A}_{\rm obs}$, we mean a linear function $\mathcal{O} \to \langle \mathcal{O} \rangle$ which

(1) is positive, in the sense that $\langle \mathcal{O}\mathcal{O}^\dagger \rangle \geq 0$ for all $\mathcal{O} \in \mathcal{A}_{\rm obs}$

(2) is consistent with all universal OPE relations.

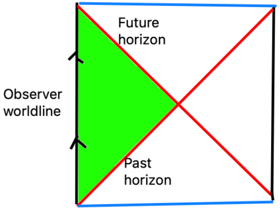
(This is analogous to a definition given by Hollands and Wald in quantum field theory in a fixed curved spacetime background.)

If M is any spacetime in which the observer is found, $\mathcal{H}$ is the Hilbert space that describes the fields in M together with the observer, and $\Psi \in \mathcal{H}$, then

$$\mathcal{O} \to \langle \Psi | \mathcal{O} | \Psi \rangle$$

is a state of the algebra $\mathcal{A}_{\rm obs}$, by that definition. Though these definitions make sense for any M, they are most interesting when, because of black hole or cosmological horizons, the part of the universe that the observer can see does not include a complete Cauchy hypersurface.

There is a very special case that turns out to be important. This is the case that M is an empty de Sitter space, with some positive value of the effective cosmological constant.



The green region is called a static patch, because it is invariant under a particular de Sitter generator H that advances the proper time of the observer.

In the absence of gravity, there is a distinguished de Sitter invariant state $\Psi_{\rm dS}$ such that correlation functions in this state are thermal at the de Sitter temperature $T_{\rm dS} = 1/\beta_{\rm dS}$ (Gibbons and Hawking; Figari, Nappi, and Hoegh-Krohn). For example, this means that two point functions $\langle\Psi_{\rm dS}|\phi(\tau)\phi'(\tau')|\Psi_{\rm dS}\rangle$ have two key properties:

(1) Time translation symmetry:

$$\langle\Psi_{\rm dS}|\phi(\tau+s)\phi'(\tau'+s)|\Psi_{\rm dS}\rangle = \langle\Psi_{\rm dS}|\phi(\tau)\phi'(\tau')|\Psi_{\rm dS}\rangle.$$

(2) The KMS condition, which says roughly:

$$\langle\Psi_{\rm dS}|\phi(\tau)\phi'(0)|\Psi_{\rm dS}\rangle = \langle\Psi_{\rm dS}|\phi'(0)\phi(\tau-{\rm i}\beta)|\Psi_{\rm dS}\rangle.$$

(A precise statement involves holomorphy of the correlation function in a strip in the complex plane.)

Including gravity and the observer, we define a special state in which the observer energy has a thermal distribution at the de Sitter temperature

$$\Psi_{\rm max} = \Psi_{\rm dS} e^{-\beta_{\rm dS} q/2} \sqrt{\beta_{\rm dS}},$$

and we replace operators $\phi(\tau)$ by "gravitationally dressed" operators $\widehat{\phi}_s = \Pi \phi(p+s)\Pi$. Then a straightforward computation shows that

(1$'$) We still have time-translation symmetry

$$\langle \Psi_{\rm max} | \widehat{\phi}_s \widehat{\phi}'_{s'} | \Psi_{\rm max} \rangle = \langle \Psi_{\rm max} | \widehat{\phi}'_{s+c} \widehat{\phi}'_{s'+c} | \Psi_{\rm max} \rangle, \quad c \in \mathbb{R}.$$

(2$'$) The KMS condition simplifies:

$$\langle \Psi_{\rm max} | \widehat{\phi}_s \widehat{\phi}'_{s'} | \Psi_{\rm max} \rangle = \langle \Psi_{\rm max} | \widehat{\phi}'_{s'} \widehat{\phi}_s | \Psi_{\rm max} \rangle.$$

Condition $(2')$ tells us that if, for any $\mathsf{a} \in \mathcal{A}_{\rm obs}$, we define

$$\mathrm{Tr}\, \mathsf{a} = \langle \Psi_{\rm max} | \mathsf{a} | \Psi_{\rm max} \rangle,$$

then the function Tr does have the algebraic property of a trace:

$$\mathrm{Tr}\, \mathsf{ab} = \mathrm{Tr}\, \mathsf{ba}, \quad \mathsf{a}, \mathsf{b} \in \mathcal{A}_{\rm obs}.$$

This function has the property that $\mathrm{Tr}\, \mathsf{a}^\dagger \mathsf{a} > 0$ for all $\mathsf{a} \neq 0$, meaning in particular that it is "nondegenerate." Note that if $\Psi_{\rm max}$ is normalized then

$$\mathrm{Tr}\, 1 = 1.$$

Let $\mathcal{H}_{\mathrm{dS}}$ be the Hilbert space that we get by quantizing fields in de Sitter space (in perturbation theory). It is important to understand what $\mathcal{H}_{\mathrm{dS}}$ describes and what it doesn't describe. There are states in $\mathcal{H}_{\mathrm{dS}}$ with any number of graviton excitations, but not a number of them of order $1/\hbar$. An $\hbar$-independent number of graviton excitations produces a back reaction on the geometry that is of order $\hbar$ – or order G. $\mathcal{H}_{\mathrm{dS}}$ doesn't describe states with an $\mathcal{O}(1)$ change in the geometry, only an $\mathcal{O}(G)$ change.

Let Ψ be any state in $\mathcal{H}_{\rm dS}$ and consider the function $\mathsf{a} \to \langle\Psi|\mathsf{a}|\Psi\rangle$, $\mathsf{a} \in \mathcal{A}_{\rm obs}$. Roughly speaking, because $\mathcal{A}_{\rm obs}$ has the nondegenerate trace Tr, we can hope that there is a "density matrix" $\rho \in \mathcal{A}_{\rm obs}$ such that

$$\langle\Psi|\mathsf{a}|\Psi\rangle = \mathrm{Tr}\,\mathsf{a}\rho, \quad \mathsf{a} \in \mathcal{A}_{\rm obs}.$$

Actually, such a ρ does exist but not as an element of the universal OPE algebra $\mathcal{A}_{\rm obs}$ that we've defined so far. We really have to replace $\mathcal{A}_{\rm obs}$ by its completion $\mathcal{A}_{\rm obs,dS}$ among operators on $\mathcal{H}_{\rm dS}$. Then ρ exists as an operator in (or more precisely, in general affiliated to) $\mathcal{A}_{\rm obs,dS}$. In this setup, a density matrix is defined as a positive operator ρ in (or affiliated to) $\mathcal{A}_{\rm obs,dS}$ such that

$$\mathrm{Tr}\,\rho = 1.$$

There is such a ρ for each $\Psi \in \mathcal{H}_{\rm dS}$.

The definition of the trace makes it clear that the density matrix of the state $\Psi_{\rm max}$ is $\sigma_{\rm max}=1$, since to satisfy

$$\langle\Psi_{\rm max}|\mathsf{a}|\Psi_{\rm max}\rangle=\mathrm{Tr}\,\mathsf{a}\sigma_{\rm max}\equiv\langle\Psi_{\rm max}|\mathsf{a}\sigma_{\rm max}|\Psi_{\rm max}\rangle,$$

we set

$$\sigma_{\rm max}=1.$$

This means that $\Psi_{\rm max}$ is "maximally mixed," similar to a maximally mixed state in ordinary quantum mechanics whose density matrix is a multiple of the identity.

Once we know that every state has a density matrix, we can define entropies as well. The von Neumann entropy of a density matrix ρ is as usual

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

In ordinary quantum mechanics, a maximally mixed state has a density matrix that is a multiple of the identity, and it has the maximum possible von Neumann entropy. The analog here is Ψ_{max}, with density matrix $\sigma_{\mathrm{max}} = 1$. It is clear that

$$S(\sigma_{\mathrm{max}}) = -\mathrm{Tr}\,1 \log 1 = 0,$$

and by imitating an argument that in ordinary quantum mechanics proves that a maximally mixed state has maximum possible entropy, one can prove that every other density matrix $\rho \neq 1$ has strictly smaller entropy:

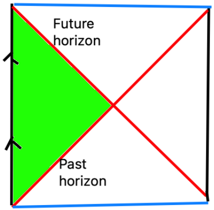
$$S(\rho) < 0.$$

Thus, the system consisting of an observer in a static patch in de Sitter space has a state of maximum entropy

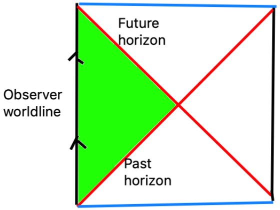
$$\Psi_{\mathrm{max}} = \Psi_{\mathrm{dS}} e^{-\beta_{\mathrm{dS}} q/2} \sqrt{\beta_{\mathrm{dS}}},$$

consisting of empty de Sitter space with a thermal distribution of the observer energy. Why did this happen?

The original argument that empty de Sitter space has maximum entropy is due to Bousso (2000), who argued that this must be true, based on the Second Law of Thermodynamics, because the static patch is empty in the far future:



The original argument that empty de Sitter space has maximum entropy is due to Bousso (2000), who argued that this must be true, based on the Second Law of Thermodynamics, because the static patch is empty in the far future:



In the present context, we've defined the static patch by the presence of the observer, so by definition the observer doesn't leave the static patch even in the far future. But we can expect in the far future that the observer will be in thermal equilibrium with the bulk quantum fields, and that is what we see in the state $\Psi_{\max}$. So the maximum entropy state that we found is the one suggested by Bousso's argument.

It is possible to show (CLPW 2022) that entropy defined as I have explained agrees *up to an additive constant independent of the state* with the usual definition of the generalized entropy

$$S_{\mathrm{gen}} = \frac{A}{4G} + S_{\mathrm{out}},$$

for a suitable class of semi-classical states $\Psi \in \mathcal{H}_{\mathrm{dS}}$. (The argument involves comparing to a description of the generalized entropy by Wall (2011).) The constant discrepancy is simply that in this approach the maximum entropy is defined to be 0, rather than $S_{\mathrm{dS}} = A_{\mathrm{dS}}/4G$ as usual. So S_{dS} has been subtracted from all entropies. A limitation: in this comparison of the entropy defined algebraically to the generalized entropy, we are only seeing contributions of $\mathcal{O}(1)$, not $\mathcal{O}(1/G)$, from the $A/4G$ term. It would be much nicer to eliminate this restriction.

I will conclude by explaining how I think this might work, but this will be speculative. First of all, recall the definition of relative entropy. The relative entropy between two density matrices ρ, σ is

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

If we take σ to be the density matrix of maximum entropy $\sigma = \sigma_{\max} = 1$, then we see

$$S(\rho) = -S(\rho|\sigma_{\max}).$$

We can reinterpret empty de Sitter space as the Hartle-Hawking no boundary state $\Psi_{\rm HH}$, restricted to the de Sitter spacetime. (Here we actually want a version of the Hartle-Hawking state for spacetimes with an observer.) I want to assume that $\Psi_{\rm HH}$ makes sense for arbitrary spacetimes in which the observer might be found.

Under this assumption, let M be any spacetime in which the observer may be, with corresponding Hilbert space $\mathcal{H}^M$. Let ρ be the density matrix for an arbitrary state $\Psi \in \mathcal{H}^M$, and let σ_{HH} be the density matrix for Ψ_{HH}. Then a possible definition of entropy is

$$S(\rho) = -S(\rho|\sigma_{\mathrm{HH}}).$$

As will be clear, this definition involves only one arbitrary additive constant, independent of the spacetime.

In one simple situation, one can check that this definition of entropy is sensible. Consider a theory that has many de Sitter vacua M^α with different values of the cosmological constant and therefore different horizon areas A^α, and different Hilbert spaces $\mathcal{H}^\alpha$. Each $\mathcal{H}^\alpha$ has a maximum entropy state $\Psi^\alpha_{\rm max}$ with density matrix $\sigma^\alpha_{\rm max} = 1|_{\mathcal{H}^\alpha}$. The Hartle-Hawking state, in the approximation of considering only these spacetimes, is

$$\Psi_{\rm HH} = \frac{1}{\sqrt{Z}} \sum_\alpha e^{A^\alpha/8G} \Psi^\alpha_{\rm max}.$$

The density matrix $\sigma_{\rm HH}$, projected to $\mathcal{H}^\alpha$, is

$$\sigma_{\rm HH}|_{\mathcal{H}^\alpha} = \frac{1}{Z} e^{A^\alpha/4G} \sigma^\alpha_{\rm max} = \frac{1}{Z} e^{A^\alpha/4G} 1|_{\mathcal{H}^\alpha}.$$

Using the relative entropy formula, the entropy of the maximum entropy state of M^α is

$$S = -S(\sigma^\alpha_{\rm max}|\sigma_{\rm HH}) = \frac{A^\alpha}{4G} - \log Z,$$

which is the standard result except for the α-independent constant $-\log Z$.

If we consider $\mathcal{O}(1)$ perturbations around the empty state of M^α, we get the expected

$$S = \frac{\widehat{A}^\alpha}{4G} + S_{\rm out} - \log Z.$$

Here $\widehat{A}^\alpha$ is the corrected horizon area, with back reaction taken into account.