gr-qc/9504004
UMDGR-95-114

Thermodynamics of Spacetime: The Einstein Equation of State[1]

Ted Jacobson

Department of Physics, University of Maryland
College Park, MD 20742-4111, USA
jacobson@umdhep.umd.edu

Abstract

The Einstein equation is derived from the proportionality of entropy and horizon area together with the fundamental relation $\delta Q = TdS$ connecting heat, entropy, and temperature. The key idea is to demand that this relation hold for all the local Rindler causal horizons through each spacetime point, with δQ and T interpreted as the energy flux and Unruh temperature seen by an accelerated observer just inside the horizon. This requires that gravitational lensing by matter energy distorts the causal structure of spacetime in just such a way that the Einstein equation holds. Viewed in this way, the Einstein equation is an equation of state. This perspective suggests that it may be no more appropriate to canonically quantize the Einstein equation than it would be to quantize the wave equation for sound in air.

The four laws of black hole mechanics, which are analogous to those of thermodynamics, were originally derived from the classical Einstein equation[1]. With the discovery of the quantum Hawking radiation[2], it became clear that the analogy is in fact an identity. How did classical General Relativity know that horizon area would turn out to be a form of entropy, and that surface gravity is a temperature? In this letter I will answer that question by turning the logic around and deriving the Einstein equation from the proportionality of entropy and horizon area together with the fundamental relation $\delta Q = TdS$ connecting heat Q, entropy S, and temperature T. Viewed in this way, the Einstein equation is an equation of state. It is born in the thermodynamic limit as a relation between thermodynamic variables, and its validity is seen to depend on the existence of local equilibrium conditions. This perspective suggests that it may be no more appropriate to quantize the Einstein equation than it would be to quantize the wave equation for sound in air.

The basic idea can be illustrated by thermodynamics of a simple homogeneous system. If one knows the entropy $S(E, V)$ as a function of energy and volume, one can deduce the equation of state from $\delta Q = TdS$. The first law of thermodynamics yields $\delta Q = dE + pdV$, while differentiation yields the identity $dS = (\partial S/\partial E)\,dE + (\partial S/\partial V)\,dV$. One thus infers that $T^{-1} = \partial S/\partial E$ and that $p = T(\partial S/\partial V)$. The latter equation is the equation of state, and yields useful information if the function S is known. For example, for weakly interacting molecules at low density, a simple counting argument yields $S = \ln(\#\text{ accessible states}) \propto \ln V + f(E)$ for some function $f(E)$. In this case, $\partial S/\partial V \propto V^{-1}$, so one infers that $pV \propto T$, which is the equation of state of an ideal gas.

In thermodynamics, heat is energy that flows between degrees of freedom that are not macroscopically observable. In spacetime dynamics, we shall define heat as energy that flows across a causal horizon. It can be felt via the gravitational field it generates, but its particular form or nature is unobservable from outside the horizon. For the purposes of this definition it is not necessary that the horizon be a black hole event horizon. It can be simply the boundary of the past of any set $\mathcal{O}$ (for "observer"). This sort of horizon is a null hypersurface (not necessarily smooth) and, assuming cosmic censorship, it is composed of generators which are null geodesic segments emanating backwards in time from the set $\mathcal{O}$. We can

[1] Published in Phys.Rev.Lett.**75**,1260(1995)

consider a kind of local gravitational thermodynamics associated with such causal horizons, where the "system" is the degrees of freedom beyond the horizon. The outside world is separated from the system not by a diathermic wall, but by a causality barrier.

That causal horizons should be associated with entropy is suggested by the observation that they hide information[3]. In fact, the overwhelming majority of the information that is hidden resides in correlations between vacuum fluctuations just inside and outside of the horizon[4]. Because of the infinite number of short wavelength field degrees of freedom near the horizon, the associated "entanglement entropy" is divergent in continuum quantum field theory. If, on the other hand, there is a fundamental cutoff length l_c, then the entanglement entropy is finite and proportional to the horizon area in units of l_c^2, as long as the radius of curvature of spacetime is much longer than l_c. (Subleading dependence on curvature and other fields induces subleading terms in the gravitational field equation.) We shall thus assume for most of this letter that the entropy is proportional to horizon area. Note that the area is an extensive quantity for a horizon, as one expects for entropy[5].

As we will see, consistency with thermodynamics requires that l_c must be of order the Planck length (10^{-33} cm). Even at the horizon of a stellar mass black hole, the radius of curvature is 10^{38} times this cutoff scale. Only near the big bang or a black hole singularity, or in the final stages of evaporation of a primordial black hole, would such a vast separation of scales fail to exist. Our analysis relies heavily on this circumstance.

So far we have argued that energy flux across a causal horizon is a kind of heat flow, and that entropy of the system beyond is proportional to the area of that horizon. It remains to identify the temperature of the system into which the heat is flowing. Recall that the origin of the large entropy is the vacuum fluctuations of quantum fields. According to the Unruh effect[8], those same vacuum fluctuations have a thermal character when seen from the perspective of a uniformly accelerated observer. We shall thus take the temperature of the system to be the Unruh temperature associated with such an observer hovering just inside the horizon. For consistency, the same observer should be used to measure the energy flux that defines the heat flow. Different accelerated observers will obtain different results. In the limit that the accelerated worldline approaches the horizon the acceleration diverges, so the Unruh temperature and energy flux diverge, however their ratio approaches a finite limit. It is in this limit that we analyse the thermodynamics, in order to make the arguments as local as possible.

Up to this point we have been thinking of the system as defined by any causal horizon. However, in general, such a system is not in "equilibrium" because the horizon is expanding, contracting, or shearing. Since we wish to apply equilibrium thermodynamics, the system is further specified as follows. The equivalence principle is invoked to view a small neighborhood of each spacetime point p as a piece of flat spacetime. Through p we consider a small spacelike 2-surface element $\mathcal{P}$ whose past directed null normal congruence to one side (which we call the "inside") has vanishing expansion and shear at p. It is always possible to choose $\mathcal{P}$ through p so that the expansion and shear vanish in a first order neighborhood of p. We call the past horizon of such a P the "local Rindler horizon of $\mathcal{P}$", and we think of it as defining a system—the part of spacetime beyond the Rindler horizon—that is instantaneously stationary (in "local equilibrium") at p. Through any spacetime point there are local Rindler horizons in all null directions.

The fundamental principle at play in our analysis is this: The equilibrium thermodynamic relation $\delta Q = TdS$, as interpreted here in terms of energy flux and area of local Rindler horizons, can only be satisfied if gravitational lensing by matter energy distorts the causal structure of spacetime in just such a way that the Einstein equation holds. We turn now to a demonstration of this claim.

First, to sharpen the the above definitions of temperature and heat, note that in a small neighborhood of any spacelike 2-surface element $\mathcal{P}$ one has an approximately flat region of spacetime with the usual Poincaré symmetries. In particular, there is an approximate Killing field χ^a generating boosts orthogonal to $\mathcal{P}$ and vanishing at $\mathcal{P}$. According to the Unruh effect[8], the Minkowski vacuum state of quantum fields—or any state at very short distances— is a thermal state with respect to the boost hamiltonian at temperature $T = \hbar\kappa/2\pi$, where κ is the acceleration of the Killing orbit on which the norm of χ^a is unity (and we employ units with the speed of light equal to unity.) The heat flow is to be defined by the boost-energy current of the matter, $T_{ab}\chi^a$, where T_{ab} is the matter energy-momentum tensor. Since the temperature and heat flow scale the same way under a constant rescaling of χ^a, this scale ambiguity will not prevent us from inferring the equation of state.

Consider now any local Rindler horizon through a spacetime point p. (See Fig. 1.) Let χ^a be an

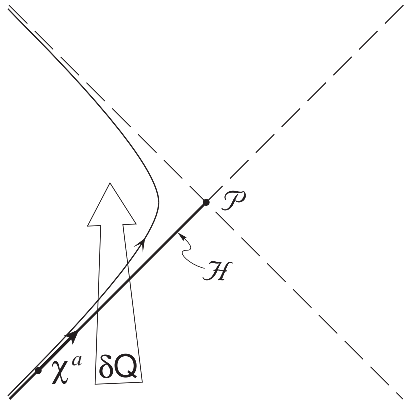

Figure 1: Spacetime diagram showing the heat flux δQ across the local Rindler horizon $\mathcal{H}$ of a 2-surface element $\mathcal{P}$. Each point in the diagram represents a two dimensional spacelike surface. The hyperbola is a uniformly accelerated worldline, and χ^a is the approximate boost Killing vector on $\mathcal{H}$.

approximate local boost Killing field generating this horizon, with the direction of χ^a chosen to be future pointing to the "inside" past of $\mathcal{P}$. We assume that all the heat flow across the horizon is (boost) energy carried by matter. This heat flux to the past of $\mathcal{P}$ is given by

$$\delta Q = \int_{\mathcal{H}} T_{ab}\chi^a d\Sigma^b. \tag{1}$$

(In keeping with the thermodynamic limit, we assume the quantum fluctuations in T_{ab} are negligible.) The integral is over a pencil of generators of the "inside" past horizon $\mathcal{H}$ of $\mathcal{P}$. If k^a is the tangent vector to the horizon generators for an affine parameter λ that vanishes at $\mathcal{P}$ and is negative to the past of $\mathcal{P}$, then $\chi^a = -\kappa\lambda k^a$ and $d\Sigma^a = k^a d\lambda d\mathcal{A}$, where $d\mathcal{A}$ is the area element on a cross section of the horizon. Thus the heat flux can also be written as

$$\delta Q = -\kappa \int_{\mathcal{H}} \lambda\, T_{ab}k^a k^b\, d\lambda d\mathcal{A}. \tag{2}$$

Assume now that the entropy is proportional to the horizon area, so the entropy variation associated with a piece of the horizon satisfies $dS = \eta\,\delta\mathcal{A}$, where $\delta\mathcal{A}$ is the area variation of a cross section of a pencil of generators of $\mathcal{H}$. The dimensional constant η is undetermined by anything we have said so far (although given a microscopic theory of spacetime structure one may someday be able to compute η in terms of a fundamental length scale.) The area variation is given by

$$\delta\mathcal{A} = \int_{\mathcal{H}} \theta\, d\lambda d\mathcal{A}, \tag{3}$$

where θ is the expansion of the horizon generators.

The content of $\delta Q = TdS$ is essentially to require that the presence of the energy flux is associated with a focussing of the horizon generators. At $\mathcal{P}$ the local Rindler horizon has vanishing expansion, so the focussing to the past of $\mathcal{P}$ must bring an expansion to zero at just the right rate so that the area increase of a portion of the horizon will be proportional to the energy flux across it. This requirement imposes a condition on the curvature of spacetime as follows.

The equation of geodesic deviation applied to the null geodesic congruence generating the horizon yields the Raychaudhuri equation

$$\frac{d\theta}{d\lambda} = -\frac{1}{2}\theta^2 - \sigma^2 - R_{ab}k^a k^b, \tag{4}$$

where $\sigma^2 = \sigma^{ab}\sigma_{ab}$ is the square of the shear and R_{ab} is the Ricci tensor. We have chosen the local Rindler horizon to be instantaneously stationary at $\mathcal{P}$, so that θ and σ vanish at $\mathcal{P}$. Therefore the θ^2 and σ^2 terms are higher order contributions that can be neglected compared with the last term when integrating to find θ near $\mathcal{P}$. This integration yields $\theta = -\lambda\, R_{ab}k^a k^b$ for sufficiently small λ. Substituting this into the equation for $\delta\mathcal{A}$ we find

$$\delta A = -\int_{\mathcal{H}} \lambda\, R_{ab}k^a k^b\, d\lambda d\mathcal{A}. \qquad (5)$$

With the help of (2) and (5) we can now see that $\delta Q = TdS = (\hbar\kappa/2\pi)\eta\,\delta A$ can only be valid if $T_{ab}k^a k^b = (\hbar\eta/2\pi)\, R_{ab}k^a k^b$ for all null k^a, which implies that $(2\pi/\hbar\eta)T_{ab} = R_{ab} + f g_{ab}$ for some function f. Local conservation of energy and momentum implies that T_{ab} is divergence free and therefore, using the contracted Bianchi identity, that $f = -R/2 + \Lambda$ for some constant Λ. We thus deduce that the Einstein equation holds:

$$R_{ab} - \frac{1}{2}Rg_{ab} + \Lambda g_{ab} = \frac{2\pi}{\hbar\eta}T_{ab}. \qquad (6)$$

The constant of proportionality η between the entropy and the area determines Newton's constant as $G = (4\hbar\eta)^{-1}$, which identifies the length $\eta^{-1/2}$ as twice the Planck length $(\hbar G)^{1/2}$. The undetermined cosmological constant Λ remains as enigmatic as ever.

Changing the assumed entropy functional would change the implied gravitational field equations. For instance, if the entropy density is given by a polynomial in the Ricci scalar $\alpha_0 + \alpha_1 R + ...$, then $\delta Q = TdS$ will imply field equations arising from a Lagrangian polynomial in the Ricci scalar[9]. It is an interesting question what "integrability" conditions must an entropy density satisfy in order for $\delta Q = TdS$ to hold for all local Rindler horizons. It seems likely that the requirement is that the entropy density arises from the variation of a generally covariant action just as it does for black hole entropy. Then the implied field equations will be those arising from that same action.

Our thermodynamic derivation of the Einstein equation of state presumed the existence of local equilibrium conditions in that the relation $\delta Q = TdS$ only applies to variations between nearby states of local thermodynamic equilibrium. For instance, in free expansion of a gas, entropy increase is not associated with any heat flow, and this relation is not valid. Moreover, local temperature and entropy are not even well defined away from equilibrium. In the case of gravity, we chose our systems to be defined by local Rindler horizons, which are instantaneously stationary, in order to have systems in local equilibrium. At a deeper level, we also assumed the usual form of short distance vacuum fluctuations in quantum fields when we motivated the proportionality of entropy and horizon area and the use of the Unruh acceleration temperature. Viewing the usual vacuum as a zero temperature thermal state[11], this also amounts to a sort of local equilibrium assumption. This deeper assumption is probably valid only in some extremely good approximation. We speculate that out of equilibrium vacuum fluctuations would entail an ill-defined spacetime metric.

Given local equilibrium conditions, we have in the Einstein equation a system of local partial differential equations that is time reversal invariant and whose solutions include propagating waves. One might think of these as analogous to sound in a gas propagating as an adiabatic compression wave. Such a wave is a travelling disturbance of local density, which propagates via a myriad of incoherent collisions. Since the sound field is only a statistically defined observable on the fundamental phase space of the multiparticle system, it should not be canonically quantized as if it were a fundamental field, even though there is no question that the individual molecules are quantum mechanical. By analogy, the viewpoint developed here suggests that it may not be correct to canonically quantize the Einstein equations, even if they describe a phenomenon that is ultimately quantum mechanical.

For sufficiently high sound frequency or intensity one knows that the local equilibrium condition breaks down, entropy increases, and sound no longer propagates in a time reversal invariant manner. Similarly, one might expect that sufficiently high frequency or large amplitude disturbances of the gravitational field would no longer be described by the Einstein equation, not because some quantum operator nature of the metric would become relevant, but because the local equilibrium condition would fail. It is my hope that, by following this line of inquiry, we shall eventually reach an understanding of the nature of "non-equilibrium spacetime".

I am grateful to S. Corley, J.C. Dell, R.C. Myers, J.Z. Simon, and L. Smolin for helpful comments on the presentation in earlier drafts of this letter. This work was supported in part by NSF grant PHY94-13253.

References

[1] J.M. Bardeen, B. Carter and S.W. Hawking, *Comm. Math. Phys.* **31** 161 (1973).

[2] S.W. Hawking, *Comm. Math. Phys.* **43** 199 (1975).

[3] J.D. Bekenstein, *Phys. Rev. D* **7**, 2333 (1973).

[4] L. Bombelli, R.K. Koul, J. Lee, and R.D. Sorkin, *Phys. Rev. D* **34**, 373 (1986).

[5] Another argument that might be advanced in support of the proportionality of entropy and area comes from the holographic hypothesis[6, 7], ie, the idea that the state of the part of the universe inside a spatial region can be fully specified on the boundary of that region. However, currently the primary support for this hypothesis comes from black hole thermodynamics itself. Since we are trying to account for the occurrence of thermodynamic-like laws for classical black holes it would therefore be circular to invoke this argument.

[6] G. 't Hooft, "Dimensional Reduction in Quantum Gravity", Utrecht preprint THU-93/26, gr-qc/9310026 .

[7] L. Susskind, "The World as a Hologram", Stanford preprint SU-ITP-94-33, hep-th/9409089, to appear in *J. Math. Phys.*, November 1995.

[8] W.G. Unruh, *Phys. Rev. D* **14**, 870 (1976).

[9] If the entropy density on the horizon takes the form $e^\rho d\mathcal{A}$, with ρ a function of the Ricci scalar, one finds that $\delta Q = TdS$ implies $(2\pi/\hbar\eta)T_{ab} = R_{ab} - \nabla_a\nabla_b\rho + fg_{ab}$ for some function f[10]. Imposing the local energy conservation equation $\nabla^a T_{ab} = 0$ then allows one to solve uniquely for f up to a constant.

[10] T. Jacobson, G. Kang, R.C. Myers, "Increase of Black Hole Entropy in Higher Curvature Gravity", McGill preprint 94-95, Maryland preprint UMDGR-95-047, gr-qc/9503020, to be published in *Phys. Rev. D*.

[11] D.W. Sciama, P. Candelas, and D. Deutsch, *Adv. in Phys.*, **30**, 327 (1981).

FIGURE 1: Spacetime diagram showing the heat flux δQ across the local Rindler horizon $\mathcal{H}$ of a 2-surface element $\mathcal{P}$. Each point in the diagram represents a two dimensional spacelike surface. The hyperbola is a uniformly accelerated worldline, and χ^a is the approximate boost Killing vector on $\mathcal{H}$.