Found Phys (2018) 48:542–557
https://doi.org/10.1007/s10701-018-0150-7

The Matter-Gravity Entanglement Hypothesis

Bernard S. Kay[1]

Received: 30 June 2017 / Accepted: 1 March 2018 / Published online: 27 March 2018

Abstract I outline some of my work and results (some dating back to 1998, some more recent) on my matter-gravity entanglement hypothesis, according to which the entropy of a closed quantum gravitational system is equal to the system's matter-gravity entanglement entropy. The main arguments presented are: (1) that this hypothesis is capable of resolving what I call the second-law puzzle, i.e. the puzzle as to how the entropy increase of a closed system can be reconciled with the asssumption of unitary time-evolution; (2) that the black hole information loss puzzle may be regarded as a special case of this second law puzzle and that therefore the same resolution applies to it; (3) that the black hole thermal atmosphere puzzle (which I recall) can be resolved by adopting a radically different-from-usual description of quantum black hole equilibrium states, according to which they are total pure states, entangled between matter and gravity in such a way that the partial states of matter and gravity are each approximately thermal equilibrium states (at the Hawking temperature); (4) that the Susskind–Horowitz–Polchinski string-theoretic understanding of black hole entropy as the logarithm of the degeneracy of a long string (which is the weak string coupling limit of a black hole) cannot be quite correct but should be replaced by a modified understanding according to which it is the entanglement entropy between a long string and its stringy atmosphere, when in a total pure equilibrium state in a suitable box, which (in line with (3)) goes over, at strong-coupling, to a black hole in equilibrium with its thermal atmosphere. The modified understanding in (4) is based on a general result, which I also describe, which concerns the likely state of a quantum system when it is weakly coupled to an energy-bath and the total state is a random

✉ Bernard S. Kay
bernard.kay@york.ac.uk

[1] Department of Mathematics, University of York, York YO10 5DD, UK

pure state with a given energy. This result generalizes Goldstein et al.'s 'canonical typicality' result to systems which are not necessarily small.

Keywords Matter-gravity entanglement · Information loss · String theory approach to black hole entropy · Gravitational decoherence · Second law of thermodynamics · Canonical typicality

1 The Second Law Puzzle

Let me begin my talk[1] by recalling one version of the second law of thermodynamics:

The entropy of the universe begins low and increases monotonically.

There are long-established and well-known arguments—see the discussion of 'branch systems' in [1] as also reviewed e.g. in [2])—that other statements of the second law, in terms of what can and cannot happen with heat engines, refrigerators etc. follow from the above statement. As also explained in these references, the above statement leads to an explanation of time asymmetry; i.e. why, for example, it is commonplace to observe wine-glasses fall off tables and smash into pieces, but we never see lots of smashed pieces assemble themselves into wine-glasses and jump onto tables (Fig. 1).

But how do we define the entropy of a closed system? And why *does it* increase?

A standard way of answering this (essentially due to Boltzmann around 1870) might be to consider for example what will happen if one starts with a system of N gas molecules in the left half of a box (see Fig. 2) and removes a partition, allowing the particles to diffuse into the right half of the box.

In a classical discussion, one describes the states of this system with some given energy in terms of a $6N - 1$ dimensional phase space, the points of which are called 'microstates' and (see Fig. 3) one imagines this phase space to be divided up into cells—called 'macrostates'—with the property that we cannot in practice distinguish between any pair of microstates in any single macrostate. One then defines the ('coarse-grained') entropy, S, of a microstate by

$$S = k \log W \tag{1}$$

where k is Boltzmann's constant and W is the volume of the macrostate containing that given microstate.

The standard argument then is that (see Fig. 3) the macrostate corresponding to "all the particles are in the left half of the box" will have a vastly smaller volume in phase space than the large macrostate which corresponds to "the molecules fill the box with roughly uniform density". Hence, as time goes on and the state of the system wanders around the phase space accordingly, it is highly likely that the entropy—as defined by (1) will get bigger and stay bigger.

[1] This article is a written version of a talk given at the 18th UK and European Conference on Foundations of Physics (16–18 July 2016, LSE, London)

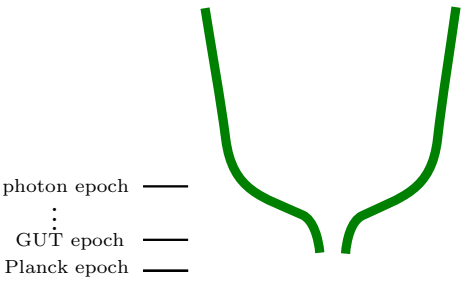

Fig. 1 Schematic diagram of the universe showing how its radius increases with time

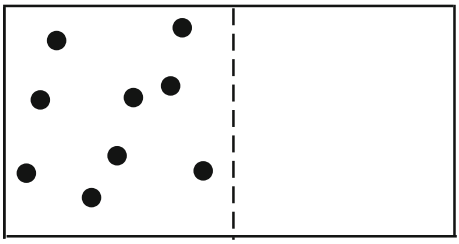

Fig. 2 A box of gas molecules, initially confined to the left half

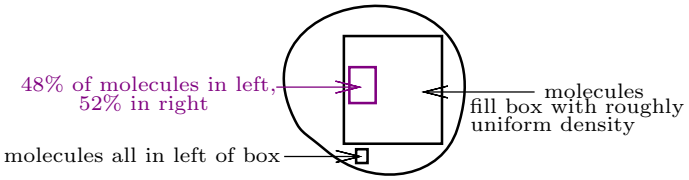

Fig. 3 The phase space for the gas in the box, indicating some possible macrostates

However, this definition of entropy and this argument for its increase depends, unsatisfactorily, on the need to make judgments about what *we* can distinguish. For example, if (see again Fig. 3) after previously ignoring such fine distinctions, we were to take the view that we can distinguish a state where, say, 48% of the particles are in the left half of the box and 52% in the right half from a state with roughly equal proportions[2] then, at times for which the system's microstate lies in the accordingly-defined new macrostate (obviously a subregion of the previously discussed large macrostate) then Eq. (1) would ascribe a different value to the entropy.

Moreover, this unsatisfactory arbitrariness and vagueness in the definition of entropy is even more of a problem if we want to account for the version of the second law with which we began. For *we* are not even present to make any distinctions in the early universe!

Turning to the quantum setting, von Neumann gave us long ago a quantum translation of Boltzmann's equation (1). Given a description of our system in terms of a density operator, ρ acting on the system's Hilbert space $\mathcal{H}$, one defines its von Neumann entropy, $S^{\mathrm{vN}}(\rho)$, by

[2] These numbers were not entirely randomly chosen, the talk being given shortly after the June 2016 Brexit referendum.

$$S^{\mathrm{vN}}(\rho) = -k\mathrm{tr}(\rho \log \rho). \tag{2}$$

But if we were to equate the physical entropy, S^{physical}, with $S^{\mathrm{vN}}(\rho)$ and if ρ satisfies the usual unitary time evolution rule

$$\rho(t) = U(t)\rho(0)U(t)^{-1}$$

then we would conclude that

$$S^{\mathrm{physical}}(\rho(t)) = \text{constant}.$$

in contradiction with the second law. We shall call this the *second law puzzle*. One can overcome this difficulty by defining quantum counterparts to the above classical coarse-graining, but of course one then would have the same unsatisfactory vagueness and subjectivity as we discussed above in the classical case.

More interestingly, one can seek to exploit a feature of quantum mechanics which has no classical counterpart: If we have a pure state, described by a density operator, $\rho = |\Psi\rangle\langle\Psi|$, which is a projector onto a vector, Ψ, in a Hilbert space, $\mathcal{H}_{\mathrm{total}}$, which arises as the tensor product,

$$\mathcal{H}_{\mathrm{total}} = \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$$

of two Hilbert spaces, $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$, then the reduced density operator, ρ_A on $\mathcal{H}_{\mathrm{A}}$, defined as the partial trace, $\mathrm{tr}_{\mathcal{H}_{\mathrm{B}}}(\rho)$, of ρ over $\mathcal{H}_{\mathrm{B}}$, will typically have $S^{\mathrm{vN}}(\rho_{\mathrm{A}}) > 0$.

We remark that

– This partial trace is characterized by the property that, if O is a (self-adjoint) operator on $\mathcal{H}_{\mathrm{A}}$, then

$$\mathrm{tr}(\rho_{\mathrm{A}} O)_{\mathcal{H}_{\mathrm{A}}} = \langle \Psi(O \otimes I)|\Psi\rangle_{\mathcal{H}_{\mathrm{total}}}.$$

– Both reduced density operators have equal von Neumann entropies:

$$S^{\mathrm{vN}}(\rho_A) = S^{\mathrm{vN}}(\rho_{\mathrm{B}}) \tag{3}$$

and this common value is often known as the A–B entanglement entropy of the total state-vector Ψ.

In a variant of the 'environment paradigm for decoherence' or, from another point of view, a variant of a possible approach to quantum statistical mechanics, this formalism is often applied in the case that A is interpreted as standing for some 'system' and B for the system's 'environment' or 'energy bath' and $S^{\mathrm{vN}}(\rho_{\mathrm{A}})$ is then interpreted as the entropy of the system due to entanglement with the environment.

So the environment paradigm gives us an objective notion of entropy. However, there remain problems:

– It only offers a notion of entropy for *open* systems.

– There are lots of ways of decomposing a given $\mathcal{H}$ as $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$. How we choose to decompose it depends on subjective choices and, again, *we* are not around in the early universe to make those choices.

What I'd like to point out is that one can envisage an alternative physical use of this mathematical fact: Suppose there's some decomposition that's physically natural, then maybe we could define the entropy of a total *closed* system by

$$S^{\mathrm{total}} = S^{\mathrm{vN}}(\rho_{\mathrm{A}}) \quad (= S^{\mathrm{vN}}(\rho_{\mathrm{B}})) \quad (= \text{A–B entanglement entropy}) \tag{4}$$

rather than interpreting this mathematical quantity as the entropy of the A-subsystem!

We propose that the identification:

$$\mathrm{A} = \textit{matter}; \quad \mathrm{B} = \textit{gravity},$$

is the right choice. This is our *matter-gravity entanglement hypothesis*. (See [3–5] for early papers, and [6] and the remainder of the present article for recent partial overviews and further references.)

In support of this, we note that the decomposition has to be meaningful throughout the entire history of the universe: E.g. we could not identify A with *photons* and B with *nuclei + electrons* because these notions are not even meaningful until the photon epoch. We content ourselves, though, with going back to just after the Planck epoch; we assume that a low-energy quantum gravity theory holds there and throughout the entire subsequent history of the universe and that this is a conventional (unitary) quantum theory with $\mathcal{H} = \mathcal{H}_{matter} \otimes \mathcal{H}_{gravity}$. We will also assume that the initial degree of matter-gravity entanglement is low. (We leave it for a future theory of the pre-Planck era to explain that.)

These assumptions then appear to be capable of offering an explanation of the second law in the form stated at the outset since one can argue that an initial state with a low degree of matter-gravity entanglement will, because of matter-gravity interaction, get more entangled, plausibly monotonically, as time increases. At least the question of whether the second law holds becomes a question which, in principle, can be answered mathematically once we specify the (low-energy) quantum gravity Hamiltonian (i.e. the generator of the unitary time-evolution) and the initial state. What we have called the second law puzzle would then be resolved because once we define entropy as matter-gravity entanglement entropy (rather than as the von Neumann entropy of the total state) there is no conflict between its increase and a unitary time-evolution.

2 The Information Loss Puzzle (Hawking 1976)

The celebrated result of Hawking [7] is that a black hole formed by the dynamical collapse of a star will emit thermal radiation at the Hawking temperature, given, in the case of a spherically symmetric electrically neutral black hole (Fig. 4) by

$$kT_{\mathrm{Hawking}} = \frac{1}{8\pi GM} \tag{5}$$

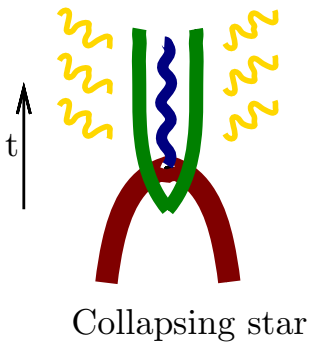

Fig. 4 A schematic picture of the spacetime of a star which collapses to a black hole and then Hawking-evaporates. The thick brown lines represent the boundary of the surface of a collapsing star, the green lines the horizon, the blue wiggly line the future spacetime singularity. The thin yellow wiggles indicate the Hawking radiation predicted in [7] (Color figure online)

where M is the black hole mass (and we take $c = \hbar = 1$).

As Hawking explained in that work, one expects that such a radiating black hole will lose mass, increasing further its temperature, and eventually evaporate.

During this whole process of collapse to a black hole and subsequent evaporation, one expects the entropy of the total system to increase monotonically.[3]

The version of the *information loss puzzle* [8] that I shall adopt here is the puzzle as to how this entropy increase can be reconciled with an assumption of unitary time evolution.

Stated in this way, I think it is clear that the information loss puzzle is nothing but a special case of our Second Law Puzzle; we recall here that this is the puzzle that, if one equates S^{physical} with $S^{\text{vN}}(\rho_{\text{total}})$, then S^{physical} must be constant.

I suggested in [3,4] that the resolution to the information loss puzzle is simply the special case of the above proposed resolution to the second law puzzle. Namely, S^{physical} is *not* $S^{\text{vN}}(\rho_{\text{total}})$. Rather S^{physical} is the total state's matter-gravity entanglement entropy. As I already said in the more general context in Sect. 1, this is not a unitary invariant and—it is reasonable to assume—would increase, thus offering to resolve the puzzle. That it also offers this resolution to the information loss puzzle lends, is, in my view, further evidence that our matter-gravity entanglement hypothesis is on the right track.

3 The Thermal Atmosphere Puzzle

A black hole in a box in equilibrium with its thermal atmosphere (see Fig. 5) is traditionally taken to be in a total Gibbs state (in particular a total mixed state) at the Hawking temperature.

[3] Without wishing to imply that they are necessarily exactly additive, we note that while the entropy of the black hole (given by (6)) will decrease because the horizon area will decrease, one expects that this will be more than compensated by the increased entropy of the sphere of emitted Hawking radiation which is growing in size at the speed of light and within which, moreover, the later radiation will be hotter than that emitted earlier.

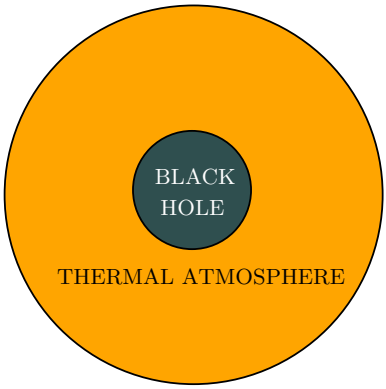

Fig. 5 A schematic picture of a black hole in equilibrium with its thermal atmosphere in a box

Everyone agrees that the entropy of this system has (at least up to small corrections) the value

$$S^{\text{Hawking}} = 4\pi kGM^2 = kA/4G. \tag{6}$$

where A is the surface area of the event horizon ($= 16\pi G^2M^2$). The thermal atmosphere puzzle [9,10] is that one can give seemingly convincing arguments for each of the following three, at first sight seemingly mutually contradictory, statements about the nature and origin of this entropy:

- It is the entropy of the gravitational field (so mostly 'residing' in the black hole).
- It is the entropy of the thermal atmosphere (so apart from the graviton component, consisting mainly of matter).
- It is the sum of the above two entropies.

Our proposed resolution of the puzzle begins by postulating that it is not actually the case that the total state is a Gibbs state; rather, we propose, *the total state is pure*, but entangled between gravity ($\simeq$ the black hole) and matter ($\simeq$ its atmosphere) in such a way that each are approximately Gibbs states (at the Hawking temperature).

We further suggest, in line with our matter-gravity entanglement hypothesis, that S^{Hawking} is really this state's *matter-gravity entanglement entropy*. This offers to resolve the puzzle in the following way: The first entropy can be regarded, according to the environment paradigm, as the entropy of the open system consisting of the gravitational field due to its matter environment; the second the entropy of the open system consisting of the matter due to its gravity environment. But, by (3), these are actually equal and so, in this environment-paradigm sense, both statements are therefore true, without contradiction. On the other hand, there is no reason why the third statement should be true in any sense and in fact, on our hypothesis it is clearly not true—the total entropy being, by (4) not the sum of the first two, but rather, equal to each of them.

The fact that it seems capable of providing this resolution to the thermal atmosphere puzzle provides further support for the validity of our matter-gravity entanglement hypothesis.

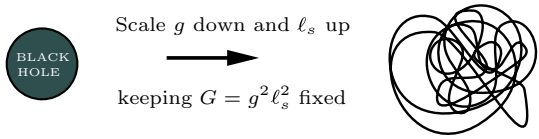

Fig. 6 The weak string-coupling limit of a black hole is a long string

4 The Weak String-Coupling Limit of Black-Hole Equilibrium States and Black Hole Entropy

Some of the most interesting work towards computing (in certain cases) or, at least, gaining a better understanding of, black hole entropy has been within string theory. Here I shall briefly recall the basic idea due to Susskind [11] and one particular line of development by Horowitz and Polchinski [12,13] which leads to an explanation of how the entropy of spherically symmetric black holes scales with M^2 (the square of the black-hole mass), albeit the argument is semi-qualitative and does not tell us the constant term (so does not explain the factor of $1/4$ in (6)).

First I will outline the Susskind–Horowitz–Polchinski (SHP) argument. Then I will criticize it. Then I will propose a modification of the SHP argument which is free from the criticisms I raise and is consistent with the understanding of black-hole equilibrium states on the matter-gravity entanglement hypothesis that I outlined in Sect. 3.

The SHP argument [12,13] is in two steps[4]: First (see Fig. 6) one argues that, as one scales the string coupling-constant, g, down and the string length, ℓ_s up, keeping Newton's constant $G = g^2\ell_s^2$ fixed, a black hole goes over to a long string. This will have density of states (i.e. number of states per unit energy, where we use ϵ to denote energy) $\sigma_{\text{longstring}}(\epsilon)$ approximately of the form of a constant times $e^{\ell_s\epsilon}$.

Secondly, one equates the entropy, $S_{\text{blackhole}}$, with "$k\log(\sigma_{\text{longstring}}(\epsilon))$" $= k\ell_s\epsilon$ at $\epsilon =$ constant times M when $\ell_s =$ constant times GM whereupon $S_{\text{blackhole}} =$ constant times kGM^2.

Our criticism of this is that it is not correct to equate an entropy with the logarithm of a density of states. (Nor indeed, in other string theory work, with the logarithm of a degeneracy—see [6,15].) Indeed it only ever makes sense in physics to take the logarithm of a dimensionless quantity but a density of states has of course the dimensions of inverse energy!

Our proposed modification of the SHP scenario [14,15] is to consider, in place of the limit

$$\textit{black hole} \rightarrow \textit{long string},$$

the limit

black hole in equilibrium with thermal atmosphere in a box $\rightarrow$
long string in equilibrium with atmosphere of small strings in a suitably rescaled box.

[4] We adopt similar simplifications to those adopted in [12,13]. Thus the spacetime dimension is taken to be 4 and the power-law prefactors in the densities of states are ignored. See however [14] for the importance of those prefactors in my proposed modification of the SHP argument.

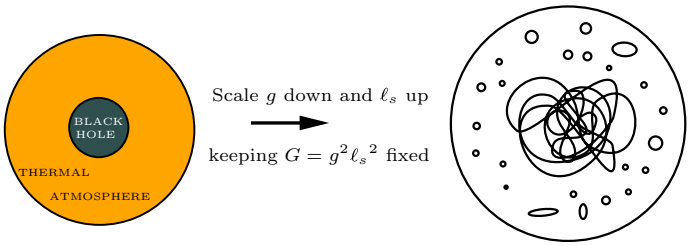

Fig. 7 The weak string-coupling limit of a black hole in equilibrium with its atmosphere in a suitable box is a long string in equilibrium with its stringy atmosphere in another box

The key fact [12,13] about a string equilibrium state of this latter type is that (in a certain approximation where we ignore certain power-law prefactors—see Footnote 4) the long string and its stringy atmosphere will have densities of states of the exponential form:

$$\sigma_{\text{longstring}}(\epsilon) \sim c e^{\ell_s \epsilon}, \quad \sigma_{\text{stringyatmosphere}}(\epsilon) \sim c' e^{\ell_s \epsilon} \tag{7}$$

where the constants c and c' may be different, but, importantly the exponents are the same.

I have demonstrated (see Sect. 5 for a discussion of the proof) that:

Theorem 1 *For any pair of weakly coupled systems (to be called here 'system' and 'bath') with densities of states as in* (7) *a randomly chosen* pure *equilibrium state with total energy E will, with very high probability, have a system-bath entanglement entropy approximately equal to $k\ell_s E/4$. It will also be such that the reduced states of system and bath separately each have energy $E/2$ and are each approximately thermal at temperature $T = 1/k\ell_s$*

Applying this theorem and reading 'long string' for 'system' and 'stringy atmosphere' for 'bath' (or vice versa) and equating the black hole mass, M, with a constant times E and the entanglement entropy of this theorem with the matter-gravity entanglement entropy of the black hole equilibrium state at $\ell_s =$ constant times GM (as in the unmodified argument) the latter entropy will thus be a constant times kGM^2. Thus we achieve a corrected string explanation of this formula for the black hole entropy which is not subject to the criticism we made of the original SHP approach. Moreover making the same substitution, $\ell_s =$ constant times GM, the temperature formula for the reduced states of the long string and of its stringy atmosphere goes over to the temperature formula $T =$ a constant times $1/kGM$, which agrees with the Hawking temperature formula (5) (up to a constant).[5]

That ends my discussion of my matter-gravity entanglement hypothesis and of how it offers a resolution to the three puzzles: the second law puzzle, the black hole information loss puzzle, and the thermal atmosphere puzzle and, finally, in this section,

[5] Intriguingly, as pointed out in [15], if one equates M with $E/2$ and equates ℓ_s with $8\pi GM$, then one gets the right value both for the Hawking temperature and the Hawking entropy. However, as explained in [14,15] this numerical coincidence should be interpreted with caution.

of how it enables a modification of the SHP string approach to black hole entropy which is free from the criticism[6] which I made of the original SHP approach.

In the remainder of the talk I would like to supply some of the details about how I proved the above theorem.

5 Explanations of Thermality: Traditional and Modern

Theorem 1 in fact relies on a general theorem—which is stated below as Theorem 2—which I obtained [16] in a general setting where one has a total system (in [16] I abbreviate this with the the term 'totem' and I shall follow that terminology here) consisting of a (quantum) system weakly coupled to an energy bath.

Such a totem will have a Hamiltonian of form

$$H = H_{\text{system}} + H_{\text{bath}} + H_{\text{interaction}}$$

on

$$\mathcal{H}_{\text{system}} \otimes \mathcal{H}_{\text{bath}}$$

where $H_{\text{interaction}}$ is assumed to be sufficiently weak that it can be ignored for the purposes of counting energy levels; $\mathcal{H}_{\text{system}}$ and $\mathcal{H}_{\text{bath}}$ each have positively supported, locally finite, discrete spectrum with monotonically increasing densities of states,

$$\sigma_{\text{system}}(\epsilon) \quad \text{and} \quad \sigma_{\text{bath}}(\epsilon).$$

Theorem 2 may be considered to generalize a result of Goldstein, Lebowitz, Tumulka and Zanghi (GLTZ) [17] (see also [18]) which explains why it is that a small system in contact with a large energy bath will typically be in an (approximate) thermal equilibrium state. So I will first briefly recall that result.

5.1 Thermality in the Case the System is Small

The GLTZ explanation is itself a modern replacement for the earlier traditional explanation of the thermality of a small system in contact with a heat bath, so let me recall that first (Fig. 8).

[6] To provide further perspective on that criticism, let us recall that the attempt to provide a microscopic explanation of thermodynamical behaviour in terms of a classical statistical mechanics has often been criticized because it requires the introduction of an ad hoc quantity with the dimensions of action in order to provide a unit of volume in phase space. It has been said that this shortcoming of classical statistical mechanics is overcome in quantum statistical mechanics where a suitable power of the quantity $\hbar$ effectively provides the right volume element. One might re-express the main thesis of this section by saying that, in a similar way, the need to introduce an ad hoc dimensionful quantity as in the SHP approach to black hole entropy and the resolution of that difficulty along the lines explained in the main text indicates that, to have a satisfactory microscopic explanation of thermodynamical behaviour, a quantum statistical mechanics is *also* insufficient and what is needed, instead, is a quantum-gravitational statistical mechanics based on our matter-gravity entanglement hypothesis.

The traditional explanation is based on a mathematical theorem which tells us that if the totem is in a microcanonical ensemble with energy in a narrow band around some total energy E, then the small system will be approximately in a thermal equilibrium state with temperature, T_{system} given by the formula (note that the dimensionful argument of the logarithm is innocuous here because the logarithm is differentiated):

$$\frac{1}{kT_{\text{system}}} = \frac{d}{d\epsilon} \log \sigma_{\text{bath}}(\epsilon)|_{\epsilon=E}. \tag{8}$$

The modern explanation (Fig. 9) [17] is based on a mathematical theorem (proven in [17]) that if the totem state is a pure state, randomly chosen from the set of all pure states with totem energy in a narrow band around E (where the random choice is with respect to a natural measure on the set of all these pure states) then the small system will very probably be very close to the same thermal equilibrium state with a temperature given by the same formula (8).

The advantage of the "modern" over the "traditional" point of view is that it bases a theory of how systems get themselves into (approximate) Gibbs states on the same foundational assumption that we usually make for the foundations of quantum mechanics—namely that the total state of a full closed system is a pure (vector) state.

5.2 What Happens When System and Energy Bath are of Comparable Size?

One might think that one could apply the GLTZ result directly to the case our totem is the string equilibrium state illustrated in Fig. 7, identifying, say, the long string with our 'system' and the stringy atmosphere with our 'energy bath'. However, neither of these can be regarded as small with respect to the other. Here we should clarify that 'small' in this context would mean having much more widely spaced energy levels, i.e. having a much lower density of states. Instead both densities of states are (ignoring the power-law prefactors I mentioned earlier) of the exponentially increasing form (7).

It turns out in general, that when the system and the energy bath are of comparable size, then—on both the traditional assumption of a totem microcanonical ensemble and the modern assumption of a random total pure state with energy in a small band—it is no longer necessarily the case that either system or energy bath will probably be in a thermal equilibrium state. However, I have shown [16] with regard to the modern approach:

Theorem 2 *There is a special density operator* (*see the Appendix for details*)

$$\rho_{\text{system}}^{\text{modapprox}} \quad \textit{on} \quad \mathcal{H}_{\text{system}} \tag{9}$$

such that, given a random vector, $\Psi \in \mathcal{H}_{\text{system}} \otimes \mathcal{H}_{\text{bath}}$*, with energy in a narrow band around* E*, then the partial trace of* $|\Psi\rangle\langle\Psi|$ *over* $\mathcal{H}_{\text{bath}}$ *is very probably very close to* $\rho_{\text{system}}^{\text{modapprox}}$.

(And similarly with *system* $\leftrightarrow$ *energy bath*).

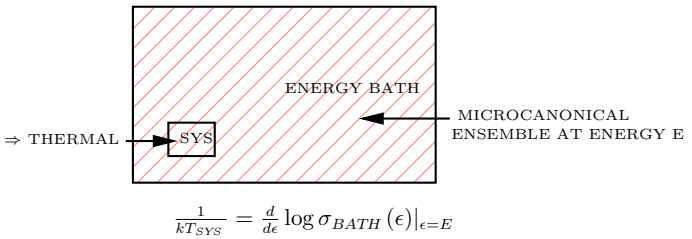

Fig. 8 The traditional explanation of the thermality of a small system

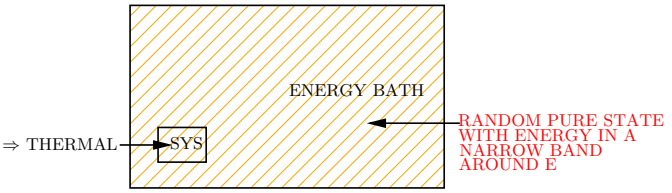

Fig. 9 The modern explanation of the thermality of a small system

But it is important to realize that when system and energy bath are *of comparable size*, $\rho_{\text{system}}^{\text{modapprox}}$ is *not always thermal*. (And neither, by the way, is the reduced state of the system thermal when the total state is in a traditional microcanonical ensemble.)

E.g. if $\sigma_{\text{system}}(\epsilon)$ and $\sigma_{\text{bath}}(\epsilon)$ take, respectively, the power law forms $\sigma_{\text{system}}(\epsilon) = A_S \epsilon^{N_S}$, $\sigma_{\text{bath}}(\epsilon) = A_S \epsilon^{N_S}$ (the typical behaviour of ordinary matter when N_A and N_B are comparable in size to Avogadro's number) then the system 'energy probability density', $P_{\text{system}}(\epsilon)$ [16] will be a Gaussian (in fact the same Gaussian on both traditional and modern assumptions) rather than the Gibbsian distribution characteristic of a thermal state. See Figs. 10 and 11.

5.3 The Special Nature of Exponential Densities of States

However, it is shown in [16], regarding the modern approach[7]

Theorem 3 *When system and energy-bath densities of states both take the exponential form of Eq.* (7)*:*

– $\rho_{\text{system}}^{\text{modapprox}}$ *and* $\rho_{\text{bath}}^{\text{modapprox}}$ *are (close to*[8]*) thermal at temperature* $T = 1/k\ell_s$. *(And each have mean energy* $E/2$.)

[7] A similar result to Theorem 3 holds for the traditional (microcanonical) approach, except that (now neglecting logarithmic terms) in place of $k\ell_s E/4$ one finds [16] that the system and the energy bath have entropy $k\ell_s E/2$. The difference between these two results is interesting since it demonstrates that, in general, the traditional and modern approaches do not give the same results. (It is also interesting since the "right value for the Hawking entropy" mentioned in Footnote 5 depends on the denominator being 4—rather than 2).

[8] See [16] for the sense in which these states are close to thermal.

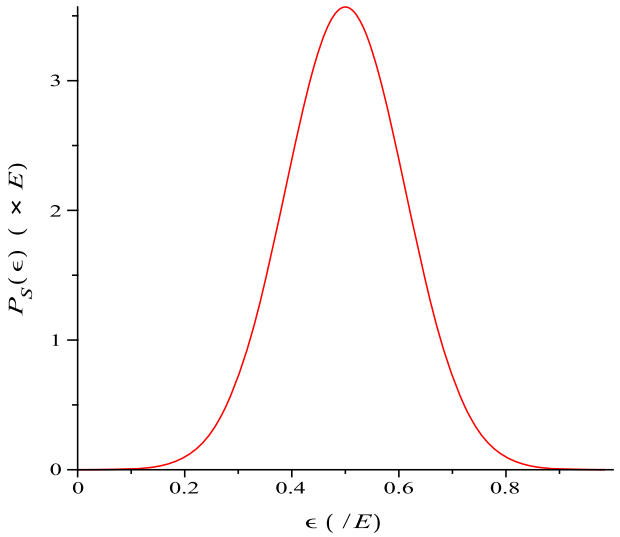

Fig. 10 Plot of the energy probability density, $P_{\text{system}}(\epsilon)$, when *system* and *energy bath* have the same power-law density of states $\sigma(\epsilon) = A\epsilon^N$ for the ('unusually' small) value $N = 10$

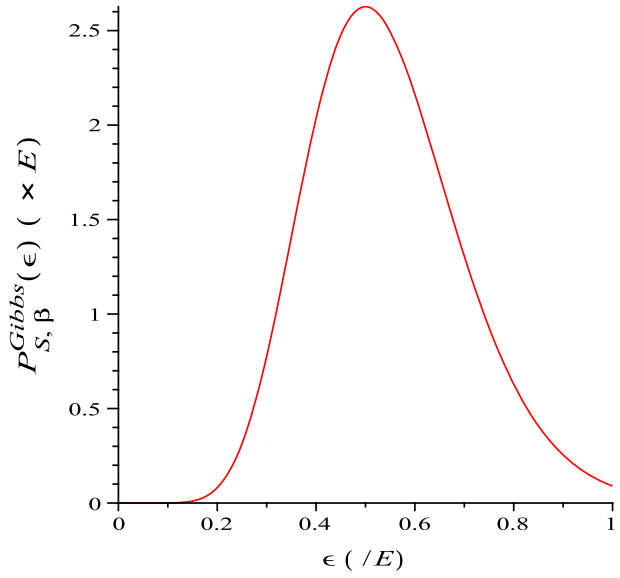

Fig. 11 Plot of the energy probability density, $P^{\text{Gibbs}}_{\text{system},\beta}(\epsilon)$ for the thermal state at inverse temperature, β, on our *system* with density of states $\sigma(\epsilon) = A\epsilon^N$, for the same ('unusually' small) value $N = 10$ and for $\beta = 22/E$ (i.e. the value of β for which the mean energy is $E/2$)

– *Also, the system-energy bath entanglement entropy, S,* $(= S^{\mathrm{vN}}(\rho_{\mathrm{system}}^{\mathrm{modapprox}}) = S^{\mathrm{vN}}(\rho_{\mathrm{bath}}^{\mathrm{modapprox}}))$ *is approximately* $k\ell_s E/4$.[9]

Theorem 1 of Sect. 4 clearly follows immediately from Theorems 2 and 3.

Appendix: Details on $\rho_{\mathbf{system}}^{\mathbf{modapprox}}$

In this appendix we give the detailed formula for the special density operator (9).

Define the (M-dimensional) Hilbert space, $\mathcal{H}_M$ (M assumed large) consisting of elements with total energy in a narrow band $[E, E+\Delta]$ to be the closed span of eigenstates of the total Hamiltonian with energies, $\epsilon \in [E, E+\Delta]$.

For convenience, replace the system of interest by a system with equally spaced energy levels with spacing equal to Δ—each energy level, ϵ, having degeneracy, $n(\epsilon) = \sigma(\epsilon)\Delta$ (so that the new system will have the same density of states, $\sigma(\epsilon)$ as the original system).
(Note that then $M = \sum_{\epsilon=\Delta}^{E} n_{\mathrm{system}}(\epsilon) n_{\mathrm{bath}}(E-\epsilon)$.)

We note first that the traditional microcanonical density operator, $\sum_{\text{basis for } \mathcal{H}_M} |\psi_i\rangle\langle\psi_i|$ is then easily seen to have reduced density operator on $\mathcal{H}_{\mathrm{system}}$ equal to

$$\rho_{\mathrm{system}}^{\mathrm{microc}} = M^{-1} \sum_{\epsilon=\Delta}^{E} n_{\mathrm{bath}}(E-\epsilon) \sum_{i=1}^{n_{\mathrm{system}}(\epsilon)} |\epsilon, i\rangle\langle\epsilon, i|$$

where $|\epsilon, i\rangle$ denotes a basis for the $n_{\mathrm{system}}(\epsilon)$-dimensional degeneracy subspace of $\mathcal{H}_{\mathrm{system}}$ with energy ϵ (assumed to be a multiple of Δ) and the sum over ϵ is over multiples of Δ.

The *modern* replacement for this result is that a random pure density operator, $|\Psi\rangle\langle\Psi|$, on $\mathcal{H}_M$ will have a reduced density operator on $\mathcal{H}_{\mathrm{system}}$ which (as is argued in [16]) is very probably very close to $\rho_{\mathrm{system}}^{\mathrm{modapprox}}$ where

$$\rho_{\mathrm{system}}^{\mathrm{modapprox}} = M^{-1} \text{ times}$$
$$\sum_{\epsilon=\Delta}^{E_c} n_{\mathrm{bath}}(E-\epsilon) \sum_{i=1}^{n_{\mathrm{system}}(\epsilon)} |\epsilon, i\rangle\langle\epsilon, i| + \sum_{\epsilon=E_c+\Delta}^{E} n_{\mathrm{system}}(\epsilon) \sum_{i=1}^{n_{\mathrm{bath}}(E-\epsilon)} |\widetilde{\epsilon, i}\rangle\langle\widetilde{\epsilon, i}|$$

where E_c is the energy at which $\sigma_{\mathrm{system}}(\epsilon) = \sigma_{\mathrm{bath}}(E-\epsilon)$ and the $|\widetilde{\epsilon, i}\rangle$ span an orthonormal basis of an $n_{\mathrm{Bath}}(E-\epsilon)$-dimensional subspace of the $n_{\mathrm{System}}(\epsilon)$-

[9] The exact result [16, Endnote 29] is $k\ell_s E/4 + k\log(c_{\mathrm{S}}c_{\mathrm{B}}E^2)/2 - k(\log(c_{\mathrm{S}}/c_{\mathrm{B}}))^2/4E$.

dimensional) energy-ϵ subspace of $\mathcal{H}_{\text{System}}$ which depends on Ψ in a way explained in detail in [16].

Afterword

To end, let me mention some related aspects of the matter-gravity entanglement hypothesis that we have not had time to discuss. One is an extension of the theory beyond closed systems to include open systems. For this, we refer to [5, Endnote (xii)] or [6]. Another concerns the relevance of the matter-gravity entanglement hypothesis to the measurement problem in quantum mechanics and a possible resolution to the Schrödinger Cat puzzle. For this, see [3–5]. Finally, the papers [19,20] (see also [6] for a brief outline of this work) include a discussion of a possible mechanism whereby, when one passes from a quantum field theory in curved spacetime description to a description in which the backreaction of the stress-energy tensor on the metric is taken into account, the horizon of an enclosed (say Kruskal) black hole becomes unstable with the consequence that entanglement between the right and left Kruskal wedges in a quantum theory in curved spacetime context transmutes into entanglement between matter and gravity—in support of the solution to the thermal atmosphere puzzle presented in Sect. 3.

References

1. Reichenbach, H.: The Direction of Time. University of California Press, Berkeley (1971)
2. Davies, P.C.W.: The Physics of Time Asymmetry. University of California Press, Berkeley (1977)
3. Kay, B.S.: Entropy defined, entropy increase and decoherence understood, and some black-hole puzzles solved. arXiv:hep-th/9802172 (1998)
4. Kay, B.S.: Decoherence of macroscopic closed systems within Newtonian quantum gravity. Class. Quant. Grav. **15**, L89–L98 (1998). arXiv:hep-th/9810077
5. Kay, B.S., Abyaneh, V.: Expectation values, experimental predictions, events and entropy in quantum gravitationally decohered quantum mechanics. arXiv:0710.0992 (2007)
6. Kay, B.S.: Entropy and quantum gravity. Entropy **17**, 8174 (2015). arXiv:1504.00882
7. Hawking, S.W.: Particle creation by black holes. Commun. Math. Phys. **43**, 199–220 (1975)
8. Hawking, S.W.: Breakdown of predictability in gravitational collapse. Phys. Rev. D **14**, 2460–2473 (1976)
9. Page, D.: Hawking radiation and black hole thermodynamics. New J. Phys. **7**, 203 (2005). arXiv:hep-th/0409024
10. Wald, R.M.: The thermodynamics of black holes. Living Rev. Relat. **4**, 1 (2001)
11. Susskind, L.: Some speculations about black hole entropy in string theory. arXiv:hep-th/9309145 (1993)
12. Horowitz, G., Polchinski, J.: A correspondence principle for black holes and strings. Phys. Rev. D **55**, 6189–6197 (1997)
13. Horowitz, G.: Quantum states of black holes. In: Wald, R.M. (ed.) Black Holes and Relativistic Stars. University of Chicago Press, Chicago (1998). arXiv:gr-qc/9704072
14. Kay, B.S.: More about the stringy limit of black hole equilibria. arXiv:1209.5110 (2012)
15. Kay, B.S.: Modern foundations for thermodynamics and the stringy limit of black hole equilibria. arXiv:1209.5110 (2012)
16. Kay, B.S.: On the origin of thermality. arXiv:1209.5215 (2012)
17. Goldstein, S., Lebowitz, J.L., Tumulka, R., Zanghi, N.: Canonical typicality. Phys. Rev. Lett. **96**, 050403 (2006). arXiv:cond-mat/0511091
18. Popescu, S., Short, A.J., Winter, A.: The foundations of statistical mechanics from entanglement: individual states vs. averages. Nat. Phys. **2**, 754 (2006). arXiv:quant-ph/0511225

19. Kay, B.S.: Instability of enclosed horizons. Gener. Relat. Gravit. **47**, 31 (2015). arXiv:1310.7395
20. Kay, B.S., Lupo, U.: Non-existence of isometry-invariant Hadamard states for a Kruskal black hole in a box and for massless fields on 1+1 Minkowski spacetime with a uniformly accelerating mirror. Class. Quantum Grav. **33**, 215001 (2016) arXiv:1502.06582