On the origin of thermality

Bernard S. Kay

Department of Mathematics, University of York, York YO10 5DD, UK

It is well-known that a small system weakly coupled to a large energy bath will, when the total system is in a microcanonical ensemble, find itself to be in an (approximately) thermal state (i.e. canonical ensemble) and, recently, it has been shown that, if the total state is, instead, a random pure state with energy in a narrow range, then the small system will still be approximately thermal with a high probability (defined by ‘Haar measure’ on the total Hilbert space). Here we ask what conditions are required for something resembling either/both of these ‘traditional’ and ‘modern’ thermality results to still hold when the system and energy bath are of comparable size. In Part 1, we show that, for given system and energy-bath densities of states, \( \sigma_S(\epsilon) \) and \( \sigma_B(\epsilon) \), thermality does not hold in general, as we illustrate when \( \sigma_S \) and \( \sigma_B \) both increase as powers of energy, but that it does hold in certain approximate senses, in both traditional and modern frameworks, when \( \sigma_S \) and \( \sigma_B \) both grow as \( e^{b\epsilon} \) or as \( e^{q\epsilon^2} \) (for constants \( b \) and \( q \)) and we calculate the system entropy in these cases. In their ‘modern’ version, our results rely on new quantities, which we introduce and call the \( S \) and \( B \) ‘modapprox’ density operators, which are defined for any positively supported, monotonically increasing, \( \sigma_S \) and \( \sigma_B \), and which, we claim, will, with high probability, closely approximate the reduced density operators for the system and energy bath when the total state of system plus energy bath is a random pure state with energy in a narrow range. In Part 2 we clarify the meaning of these modapprox density operators and give arguments for our claim.

The prime examples of non-small thermal systems are quantum black holes. Here and in two companion papers, we argue that current string-theoretic derivations of black hole entropy and thermal properties are incomplete and, on the question of information loss, inconclusive. However, we argue that these deficiencies are remedied with a modified scenario which relies on the modern strand of our methods and results here and is based on our previous matter-gravity entanglement hypothesis.

PACS numbers: 03.65.Yz, 05.30.Ch, 04.70.Dy, 04.60.Cf

I. INTRODUCTION

A. Background

This paper is concerned with the general question: “How do physical systems get to be hot?” By ‘hot’ here, we do not simply mean ‘having lots of energy’. We shall reserve the word ‘energetic’ for that. Rather, we mean the more specialized notion of being in what is known, in (quantum) statistical mechanics, as a Gibbs state, i.e. a state described by a density operator of form

\[
\rho_\beta^{\text{Gibbs}} = Z_\beta^{-1} e^{-\beta H}
\]

where \( H \) is a suitable (usually, of necessity, approximate) Hamiltonian (assumed to have discrete spectrum) for the system and \( \beta \) is related to the system’s temperature, \( T \), by \( \beta = 1/kT \) where \( k \) is Boltzmann’s constant (henceforth set to 1). Here \( Z_\beta \) stands for \( \text{tr}(e^{-\beta H}) \) and is the normalization constant which ensures that \( \rho_\beta^{\text{Gibbs}} \) will have unit trace. (When regarded as a function of \( \beta \) it is, of course, the system’s ‘partition function’.) Such states are also known as ‘canonical states’ or ‘thermal equilibrium states’ or ‘KMS states’. We shall sometimes refer to them simply as ‘thermal’ states. A possible source of confusion here is the fact that it is sometimes found to be convenient to adopt the fiction that a system which is merely energetic is in a Gibbs state at a temperature chosen so as to give it the same mean energy. Additionally, given a system with a density of states \( \sigma(\epsilon) \), it can sometimes be convenient to assign to it a ‘temperature’, \( T(\epsilon) \), at each energy, \( \epsilon \), according to the formula \( 1/T(\epsilon) = d\log \sigma(\epsilon)/d\epsilon \) \[1\]. We wish to underline that we shall not be concerned with such a fiction, nor with such an assignment of an energy-dependent ‘temperature’, here. Rather we are interested in how systems get into states which are actually Gibbs states. In particular, we are interested in black bodies, and, more particularly, black holes (in suitable boxes; here we refer...
to the remarkable developments in ‘Euclidean Quantum Gravity’ and in (Quantum) ‘Black Hole Thermodynamics’ which arose from Hawking’s pioneering work \cite{2} on ‘Black Hole Evaporation’ – see e.g. the papers on quantum black holes in the collections \cite{3, 4}.

Of course, one way for a system to get into a Gibbs state is for it to be weakly coupled to a (much larger) heat bath which is already in a Gibbs state at the desired temperature. There is a considerable literature, which, with varying degrees of mathematical rigour and generality, shows that, as one might expect, a typical such system will, more or less irrespective of its initial state, approximately get into a Gibbs state at the same temperature at late times – see e.g. \cite{5, 6}. However, what we are really interested in when we ask our general question “How do physical systems get to be hot?” is:

“How does any physical system ever get to be hot in the first place?”

Obviously, an explanation of how one system gets to be hot which invokes the existence of another system (the above-mentioned heat bath) which is assumed already to be hot can’t help to answer this version of our question!

Another traditional explanation for the propensity of some systems to be in Gibbs states goes along the following lines (see e.g. \cite{7} and, for a treatment of some of the related mathematical aspects, e.g. \cite{8} as well as the paper \cite{9}, which recalls this traditional explanation as a preliminary to its main purpose – for which see below): One assumes one’s system of interest, say described by a Hamiltonian, $H_S$, on a Hilbert space, $\mathcal{H}_S$, to be weakly coupled to a much larger ‘energy bath’, with Hamiltonian, $H_B$, on a Hilbert space, $\mathcal{H}_B$ – both Hamiltonians being assumed to have a finite number of energy levels in any finite energy interval, with the number of states of the energy bath in an energy interval, $\delta$, being approximately given in terms of a ‘density of states’, $\sigma_B$, as $\sigma_B(\delta) - \sigma_B$ being assumed to have some typical, say, power-law form (see below) – and one assumes the whole system to be in a total microcanonical state. Before we explain what we mean by this, we pause to remark, first, that, in order to avoid ambiguous usages of the word ‘system’, we shall, from now on, adopt the word *totem* (short for ‘total system’) to denote what we referred to above as our ‘whole system’. So we shall talk about a ‘totem’ which consists of a ‘system’, $S$, and an ‘energy bath’, $B$. Our assumption of weak coupling is then the assumption that the totem Hamiltonian will take the form

$$H = H_S \otimes 1 + 1 \otimes H_B + \text{weak coupling term}$$

(2)

on the totem Hilbert space, $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B$, and we shall assume further that the coupling term is so weak that it can be neglected for state and energy-level counting purposes. To say that our totem is in a microcanonical state then means to assume it is described by the density operator

$$\rho_{\text{microc}} = M^{-1} \sum |\epsilon\rangle \langle \epsilon|$$

(3)

on the totem Hilbert space, $\mathcal{H}$, where the sum is over a basis of energy eigenstates for the subspace of $\mathcal{H}$ consisting of energy levels with energies in an interval, $[E, E + \Delta]$, which is small, yet large enough for the total number of totem energy eigenstates in this range to be very large, while the normalization constant, $M$ (which is expected to roughly scale with $\Delta$) is the total number of such basis eigenstates. We further pause to note that we shall assume throughout the present paper, as is usually assumed for ‘ordinary’ physical systems, that both Hamiltonians, $H_S$ and $H_B$, are positive and their densities of states monotonically increasing. We remark though that, as we will discuss further in Section \(\text{VIII}\) were any of these assumptions to be relaxed, then the prospects for systems to become hot become much less constrained and, in particular, there are ways in which a system can be hot while the totem is in a pure state which differ from the ‘modern’ scenarios we discuss below.

Proceeding with the above assumptions, the states, $|\epsilon\rangle$, in the sum in (3) will each take the form $|\epsilon_S\rangle \otimes |\epsilon_B\rangle$ and the sum over totem energy levels will become (see (4) below) a double sum over system energy levels, $\epsilon_S$, and energy-bath energy levels, $\epsilon_B$, which satisfy the condition $\epsilon_S + \epsilon_B \in [E, E + \Delta]$. The resulting state of the system is then represented mathematically, in the usual way, by the reduced density operator, $\rho_S^{\text{microc}}$ on $\mathcal{H}_S$ i.e. by the partial trace of $\rho_{\text{microc}}$ over $\mathcal{H}_B$.

To remind ourselves how thermality of our system can then come about in this traditional explanation, it is instructive first to consider an oversimplified model in which our system Hilbert space, $\mathcal{H}_S$, is two-dimensional with only two energy levels with energies $\epsilon_1^S$ and $\epsilon_2^S$ such that $\epsilon_2^S - \epsilon_1^S \gg \Delta$ and in which the density of states, $\sigma_B$, of the energy bath grows exponentially – we shall write $\sigma_B(\epsilon) = e^{\epsilon \beta}$. (We shall discuss the case where both system and energy bath both have such a density of states in Sections \(\text{III}\) and \(\text{IV}\).

Then we easily see that $\rho_S^{\text{microc}}$ will be approximately

$$\rho_S^{\text{microc}} = n^{-1}(e^{\Delta \epsilon} e^{b(E - \epsilon_1)}|\epsilon_1^S\rangle \langle \epsilon_1^S| + e^{\Delta \epsilon} e^{b(E - \epsilon_2)}|\epsilon_2^S\rangle \langle \epsilon_2^S|)$$

where $n$ denotes the appropriate normalization constant, and this is clearly the same as the Gibbs state

$$\rho_\beta = Z_\beta^{-1}(e^{-\beta \epsilon_1}|\epsilon_1^S\rangle \langle \epsilon_1^S| + e^{-\beta \epsilon_2}|\epsilon_2^S\rangle \langle \epsilon_2^S|)$$
for $\beta = b$ for a suitable, normalizing, $Z_\beta$.

In the full story, where we now assume that also the states of the system are approximately given by a density of states, $\sigma_S$, it is convenient to assume that $E$ is an integral multiple of $\Delta$ and locally to slightly distort the spectra of system and energy bath so that their energy levels are evenly spaced at intervals $\Delta, 2\Delta, \ldots, E$ with each system level having degeneracy

$$n_S(\epsilon) = \sigma_S(\epsilon)\Delta$$

and each energy-bath level having degeneracy

$$n_B(\epsilon) = \sigma_B(\epsilon)\Delta.$$  

If, as we shall further assume, this can be done in such a way as to maintain the same ‘smoothed out’ densities of states, then it will not seriously alter the values of any quantities of interest. Choosing a basis within the degeneracy subspace of $H_S$ with each energy, $\epsilon$, and labelling its elements $|\epsilon, i\rangle$, where, for each $\epsilon$, $i=1,\ldots,n_S(\epsilon)$ while $\epsilon$ ranges from $\Delta$ to $E$ in integer steps of $\Delta$ (and similarly for the energy bath) we then easily have that $\rho_{\text{microc}}$ can be rewritten as

$$\rho_{\text{microc}} = M^{-1} \sum_{\epsilon_S} \sum_{\epsilon_B} \sum_i \sum_j |\epsilon_S, i\rangle \otimes |\epsilon_B, j\rangle \langle \epsilon_S, i| \otimes \langle \epsilon_B, j|$$

where the sum over $i$ goes from 1 to $n_S(\epsilon_S)$, the sum over $j$ goes from 1 to $n_B(\epsilon_B)$ and the sums over $\epsilon_S$ and $\epsilon_B$ are over values which are positive-integer multiples of $\Delta$ and are constrained to have $\epsilon_S + \epsilon_B = E$, while the normalization constant, $M$, defined after (3), is also given by

$$M = \sum_{\epsilon = \Delta}^E n_S(\epsilon)n_B(E-\epsilon)$$

or, roughly equivalently [14], by making the replacement

$$\sum_{\epsilon = \Delta}^E \text{by} \Delta^{-1} \int_0^E d\epsilon$$

by the approximate formula

$$M = \Delta \int_0^E \sigma_S(\epsilon)\sigma_B(E-\epsilon)d\epsilon.$$  

Moreover, $\delta/\Delta$ times the summand in (7) or $\delta\Delta$ times the integrand in (9) is, for suitable (small but not too small) $\delta$ (approximately) the number of energy eigenstates for which the energy of the totem lies in the interval $[E, E+\Delta]$ while the energy of the system lies in the interval $[\epsilon, \epsilon+\delta]$. When our totem is in the microcanonical state (3), (6), this summand divided by $M$ may thus be interpreted as the probability that the system energy lies in this latter interval. We shall denote it by $P_S(\epsilon)$ and call $P_S(\epsilon)$ the system’s energy probability density so we have

$$P_S(\epsilon) = \frac{\Delta}{M} \sigma_S(\epsilon)\sigma_B(E-\epsilon) \simeq \frac{1}{M\Delta} n_S(\epsilon)n_B(E-\epsilon),$$

and we notice, in passing, that

$$P_B(\epsilon) = P_S(E-\epsilon).$$

The reduced density operator, $\rho_S^{\text{microc}}$, of $\rho_{\text{microc}}$ on $H_S$ will clearly be

$$\rho_S^{\text{microc}} = M^{-1} \sum_{\epsilon = \Delta}^E n_B(E-\epsilon) \sum_{i=1}^{n_S(\epsilon)} |\epsilon, i\rangle \langle \epsilon, i|.$$  

(Here and below, to avoid cluttering up our formulae, we drop the ‘s’ suffix on $\epsilon$ – also in $|\epsilon, i\rangle$ – when there can be no ambiguity.)
One can then show, for a wide range of ‘realistic’ energy-bath models that, in the limit as the energy bath gets large while the system remains unchanged, $\rho_{\text{micro}}$ will converge to a thermal state at an inverse temperature $\beta$ given, in terms of the large-size behaviour of the energy bath’s density of states.

In particular, and specializing now to a case (cf. again e.g. [7]) that will interest us further below, if the density of states, $\sigma_B$, has the typical power-law form of ordinary (radiationless) matter:

$$\sigma_B(\epsilon) = A_B \epsilon^{N_B},$$

where $A_B$ is a constant and $N_B$ is an ‘Avogadro-sized’ number which could stand e.g. for ‘$3/2$ times the number of molecules’ in the energy bath or the ‘number of oscillators’ in the energy bath (see again e.g. [7] for the origin of the $3/2$ etc.) etc. then, in the limit as the total energy, $E$, of the totem gets larger while the size of the energy bath gets larger – in the sense that $N_B$ gets larger – while the system remains unaltered and $N_B/E$ converges to a constant, $\beta$, $\rho_{\text{micro}}$ will converge to a thermal state at inverse temperature $\beta$ – i.e. to the $\rho_{S,\beta}^{\text{Gibbs}}$ of Equation (13) below.

In the special case that the system has a density of states also of power-law form (see (18)) we shall provide a proof of this result, in passing, in Section (II) below which is particularly instructive in relation to our present purposes. See the last paragraph in Section (II) So, in this way, one shows that a small system in contact with a large energy bath with a suitable density of states will approximately be in a Gibbs state when the totem is in a microcanonical state.

Above, a Gibbs state (1) of our system will obviously take the form (assuming again the spectrum to be slightly distorted as explained before equation (6))

$$\rho_{S,\beta}^{\text{Gibbs}} = Z_{S,\beta}^{-1} \sum_{\epsilon = \Delta}^{\infty} e^{-\beta \epsilon} \sum_{i=1}^{n_B(\epsilon)} |\epsilon, i\rangle \langle \epsilon, i|$$

where (approximating the obvious sum by an integral as we did when we passed from (7) to (9))

$$Z_{S,\beta} = \int_0^{\infty} \sigma_S(\epsilon) \exp(-\beta \epsilon) d\epsilon.$$
state is constrained to have a definite energy. (Popescu et al. [12] then generalized these developments by allowing for more general constraints, and also made them mathematically rigorous.)

The modern (see Endnote [13]) results, [9 12], of Goldstein, Lebowitz et al. and of Popescu et al. are an advance on the traditional results in that they replace the assumption of a total microcanonical state by the assumption of a total pure state. However, they still share the limitation of the traditional approach of still only being capable of explaining how, at most, only a small subsystem of a given ‘large’ totem can get to be (approximately) thermal. The main purpose of the present paper will be to explore to what extent, and/or under what altered circumstances, this limitation can be overcome. Our main motivation relates to the theory of quantum black holes. Black holes are a puzzle in relation to the above results if one believes, as seems compelling, that the totem consisting of a black hole in equilibrium with its atmosphere in a box at approximately fixed energy is completely (approximately) thermal [10].

B. Quantum black holes

In such black hole equilibrium states we may roughly (albeit not exactly, see Endnote (iii) in [19]) identify the black hole itself with ‘gravity’ and the atmosphere with ‘matter’. In an earlier proposal (see [17] [18] and especially Endnotes (i), (ii), (iii) and (v) in [19] of the author (which predated the work [9 12] in a more general, but non-gravitational context of Goldstein, Lebowitz et al. and of Popescu et al. by around seven years) a radically-different-from-usual hypothesis was put forward as to the nature of quantum black hole equilibrium states according to which the total state is a pure state (in line with what we are calling here the ‘modern’ approach – see Endnote [13] – but in contrast to the usual assumption in work on quantum black holes that it is a Gibbs state at the Hawking temperature) while the reduced state of the gravitational field alone and also the reduced state of the matter fields alone are each thermal (i.e. each Gibbs states) at the appropriate Hawking temperature (see below). (Here we use the word ‘matter’ to include e.g. the electromagnetic field.) Below, we shall sometimes call such a total pure state bithermal. This hypothesis formed, in turn, just a part of our wider hypothesis [17 19] (which we shall sometimes refer to here as our matter-gravity entanglement hypothesis) according to which, quite generally, one should always take into account the quantum gravitational field as well as all matter fields in describing the full dynamics of any physically closed totem, and that, while the state of the totem is always pure and evolves unitarily, the ‘physically relevant’ quantum state is to be identified with the reduced density operator of the matter alone and, concomitantly (see Section I E and, in particular, Endnote [18]), the physical entropy of a closed totem is to be identified with its matter-gravity entanglement entropy. Interpreted according to this wider hypothesis, our hypothesis that quantum black hole equilibrium states are bithermal then implies that, physically, such states are completely thermal. We remark that, given our wider hypothesis, what is required for this complete thermalism is, of course, just thermality of the reduced state of the matter. However, there are strong reasons (particularly the fact [4] that the Euclideanized Schwarzschild metric is periodic in imaginary time with period \(8\pi GM\)) for believing that the mathematical nature of the reduced state of gravity will also be thermal and this is what we have assumed above and will continue to assume in the remainder of this section and in Section [19].

To summarize and also to recall the relevant formulae: While we accept the (conventional) belief that, in black hole equilibria, both matter and gravity are each separately thermal at the Hawking temperature, \(T_H\), we propose (unconventionally in comparison to other work on quantum black holes) that the total state of matter-gravity is pure (rather than itself being a thermal state). The thermality of each of the reduced states (i.e. of matter and of gravity separately) will then arise as the result of entanglement between matter and gravity in the pure totem state. We shall refer to this picture of black hole equilibrium states as our entanglement picture of black hole equilibrium. (We shall assume in Section [19] and in [22] [23] that, in this picture, the overall (i.e. totem) state of black hole equilibrium is not only pure but also close to an energy eigenstate.) We further emphasize that while this proposal is unconventional when compared to other work on quantum black holes, it seems to fit well with modern approaches (such as those of [9 12] towards understanding the origin of thermalism which have recently been proposed in non-gravitational contexts. Here, we recall that the Hawking temperature, \(T_H\), is given [3 4], in the case of a Schwarzschild (i.e. spherical, uncharged) black hole of mass \(M\), by \(T_H = 1/(8\pi GM)\) (in general the surface gravity multiplied by 2\(\pi\)). Here, \(G\) denotes Newton’s constant and we set \(c\) and \(\hbar\) to 1. Moreover, we accept the conventional belief that the physical entropy – again in the spherical, uncharged case – has the Hawking value of \(4\pi GM^2\) (in general, one quarter of the area of the event horizon, divided by \(G\)) and what is new about our proposal is our claim that this entropy-value should ultimately be explainable as the matter-gravity entanglement entropy of a pure state of the overall matter-gravity totem.

Finally, we note that our matter gravity entanglement hypothesis and our entanglement picture of black hole equilibrium also offer a natural resolution to the Information Loss Puzzle [20]. This puzzle arose because, as long as it was believed that black holes were correctly described by mixed states, then, in a dynamical process in which black holes were formed from collapsing stars etc., it appeared that an initial pure state would dynamically evolve
into a mixed state, contradicting unitarity. On the other hand, there is no difficulty in reconciling our matter-gravity
entanglement hypothesis with a unitary quantum mechanical time evolution and, once we identify entropy as matter-
gravity entanglement entropy, this is entirely consistent with increasing entropy (i.e. information loss). We note that
this proposed resolution to the Information Loss Puzzle is, in fact, just a special case of our proposed resolution to
the Second Law Puzzle [17, 18, 22].

C. Our specific question

The specific question we shall endeavour to answer in this paper assumes, as its basic setting, that a totem be given
which consists of a pair of weakly coupled systems, S and B, each with its own Hilbert space, \( \mathcal{H}_S \) and \( \mathcal{H}_B \), and each
with its own density of states, \( \sigma_S \) and \( \sigma_B \).

Our specific question is then:

If the systems, S and B, are of comparable size [24], what modifications need to be made either to the traditional
‘total microcanonical state’ approach or, more relevantly since we believe it to be a step closer to the right answer, to
the more modern ‘total pure state’ approach of Goldstein, Lebowitz et al. and of Popescu et al. and others, as described
above, so as to ensure that when the totem has a total state with energy in an interval \([E, E + \Delta]\), the reduced states
of S and B will each likely be approximately thermal states? (and, in particular, in the ‘total-pure state approach’, the
total state will likely be approximately bithermal).

(What is meant here by ‘comparable size’ has, of course, to be encoded into the functional form of the densities of
the states \( \sigma_S(\epsilon) \) and \( \sigma_B(\epsilon) \). How this is done will be clear from the specific examples we discuss.)

We hope the answers we obtain below may be of interest in their own right and that the formalism we deploy to
answer them may find a variety of other applications. But the immediate application we have in mind is to the theory
of quantum black holes. In Section IX and in our two companion papers, [22, 23], we shall argue that our answers help
to strengthen the case for, and give concrete form to, our matter-gravity entanglement hypothesis and particularly
our entanglement picture of black hole equilibrium discussed in Section I B.

D. Answers

The key to answering our specific question, in the ‘traditional total microcanonical state’ approach is the formula
\[ \rho_{\text{micro}} = M^{-1} \left( \sum_{\epsilon = \Delta} n_S(\epsilon) \sum_{i=1} {\epsilon,i} \langle \epsilon,i | \sum_{\epsilon = E_c + \Delta} E n_B(E-\epsilon) \sum_{i=1} \langle \epsilon,i | \end{align} \]

On the right hand side of this equation, we continue to assume the spectrum to be slightly distorted in the way we
explained before equation (6), \( n_S \) and \( n_B \) to be defined as in (4) and (5) and the sums to be over integral multiples of \( \Delta \), and we also continue to assume, as will be the case in our examples in Part 1, that \( \sigma_S \) and \( \sigma_B \) are monotonically
increasing functions – defining \( E_c \) to be the energy value at which \( \sigma_S(E_c) = \sigma_B(E - E_c) \). When \( \epsilon > E_c \), the \( \epsilon,i \) then
denote the elements of an orthonormal basis of an \( n_B(E-\epsilon) \)-dimensional subspace of the \( n_S(\epsilon) \)-dimensional energy-\( \epsilon \)
subspace of \( \mathcal{H}_S \) which will depend on \( \Psi \). As we shall see, this dependence on \( \Psi \) will not matter for the developments
in Part 1. We will postpone a full explanation of the way in which the subspace depends on \( \Psi \) to Section XII in
Part 2.

It is important to notice that, as is easy to check, the constant, \( M \), by which one needs to divide in order to
normalize (15) has the same value, given by (7) and (9) (and as explained after those equations, equal to the total
number of states of the totem with energy in the interval \([E, E + \Delta]\) as the constant, \( M \), by which one needs to divide
in order to normalize (11)). Moreover, while the states, \( \rho_{\text{micro}} \) and \( \rho_{\text{modern}} \), are clearly (usually) very different,
both states share the same energy probability density, \( P_S(\epsilon) \) (10). (There is of course a similar pair of equations to
(11) and (15) with obvious reversals of the letters ‘S’ and ‘B’ and, in the case of (15), with \( E_c \) replaced by \( E - E_c \).

We now claim that the sense in which (15) is the appropriate replacement for (11) in the modern approach is then
made clear by the following proposition, our argument for the correctness of which is given in (and is the main purpose
of) Part 2:

Proposition. [21] For a given, randomly chosen, pure state, \( \Psi \), on the Hilbert space of our totem, with energy
restricted to be in the range \([E, E + \Delta]\), the reduced density operator, \( \rho_{\text{modern}} \) of the system may, as far as physical
quantities of interest are concerned, with very high probability, be considered to be very close to the $\rho_S^{modapprox}$ of (15) for the appropriate (i.e. to the chosen vector $\Psi$) $n_B(E - \epsilon)$-dimensional subspaces of $\mathcal{H}_S$ spanned by the $|\epsilon, i\rangle$ (see above and Part 2). (And a similar statement of course holds with system, S, replaced by bath, B.)

What makes this proposition particularly useful is the fact that, while the $n_B(E - \epsilon)$-dimensional subspaces (spanned by the $|\epsilon, i\rangle$) of $\mathcal{H}_S$ will depend on the choice of $\Psi$ (in a way which we shall explain in Section XII in Part 2 where we point out, by the way, that they might themselves be said to be 'random subspaces') as is easy to see and as we shall illustrate in Part 1, the values of physical quantities of interest, such as the mean energy and the von Neumann entropy of the system S (see (16) below and Section I E) calculated using $\rho_S^{modapprox}$, do not depend on which $n_B(E - \epsilon)$-dimensional subspaces they are. Therefore we can conclude that, to the extent that the approximation of $\rho_S^{modern}$ by $\rho_S^{modapprox}$ is good (and we shall argue in Part 2 that, in our situations of interest, and when it is used for the purpose of calculating mean energy and entropy, it is very good) the actual values of these quantities must (with a very high probability) be largely independent of the choice of $\Psi$! (Aside from mean energy, in fact we expect the entire energy probability density function, $P_S(\epsilon)$, will most likely be close to that of $\rho_S^{modapprox}$ and hence also, similarly, for higher moments of the energy.)

Above, we recall that, for an arbitrary density operator, $\rho$, the von Neumann entropy is given by the formula

$$S(\rho) = -\text{tr}(\rho \log \rho).$$

We remark that it is easy to see from a comparison between (11) and (15) that the above proposition implies the ‘Canonical Typicality’ result [9] of Goldstein, Lebowitz et al., thus fulfilling our promise in Section I to re-obtain the latter. For, in the relevant limit (see after Equation (12)) $E_{\epsilon}$ in (15) will tend to $E$ and therefore $\rho_S^{modapprox}$ in (15) will tend to $\rho_S^{microc}$ (11), which, in turn, will tend, by the traditional argument we reviewed in Section I, to a Gibbs state (namely the $\rho_{S,\beta}^{Gibbs}$ of (13) for $\beta$ equal to the limiting value of $N_B/E$).

To start now to address our specific question, we first observe that, whatever the densities of states, $\sigma_S$ and $\sigma_B$ (provided only they are monotonically increasing) as long as the total energy, $E$, of our totem is finite, then, of course neither of the density operators, $\rho_S^{microc}$ (11) and $\rho_S^{modapprox}$ (15), can be exactly thermal. To see this easily, it suffices to notice that the energy probability density, $P_S(\epsilon)$ (10), which these states share will obviously be zero for $\epsilon > E$, whereas, when $\sigma_S(\epsilon)$ is sufficiently slowly growing for $\rho_S^{Gibbs}$ (see (13)) to exist, the energy probability density for the Gibbs state (13) will obviously take the form

$$P_{S,\beta}^{Gibbs}(\epsilon) = Z_{S,\beta}^{-1}\sigma_S(\epsilon) \exp(-\beta\epsilon),$$

where $Z_{S,\beta}$ is as in (14), which (for a rising density of states $\sigma_S$) will be non-zero for all $\epsilon$. However, one can ask whether $\rho_S^{microc}$ and/or $\rho_S^{modapprox}$ can be approximately thermal, say at sufficiently low energies.

We shall find that, for physically ordinary densities of states such as (cf. the discussion around (12))

$$\sigma_S(\epsilon) = A_S \epsilon^{N_S}, \quad \sigma_B(\epsilon) = A_B \epsilon^{N_B},$$

then, when system, S, and energy bath, B, are large and of comparable size – i.e. when $N_S$ and $N_B$ are both large, but comparably sized numbers – then neither $\rho_S^{microc}$ nor $\rho_S^{modapprox}$ can even be approximately thermal. In particular, this is the case when both system and energy-bath densities of states are identical (i.e. when $A_S = A_B$ and $N_S = N_B$). Rather, we will show that, when system and energy bath are of comparable size (or identical) in the sense just explained, the energy probability density of both S and B will, instead of having the behaviour one would expect of a thermal state, deviate from the most likely distribution of energies between S and B according to a Gaussian probability distribution with width of the order of $E$ divided by the square root of $N_S$ (equivalently $N_B$).

On the other hand, we shall show that in certain well-defined senses, ‘approximately thermal’ states are obtained for system, S, and energy bath, B, both on the traditional total microcanonical state approach and also on the modern total pure state approach if they both have identical densities of states which either rise exponentially with energy or rise as ‘quadratic exponentials’ – i.e. each as the exponential of a constant times the square of the energy – the notion of ‘approximately thermal’ depending both on the approach (i.e. the traditional total microcanonical state approach or the modern total pure state approach) and also on the behaviour of the densities of states (i.e. on whether they rise as the exponential of energy or of energy squared). See especially the notions of ‘$E$-approximately thermal’ and ‘$E$-approximately semi-thermal’ introduced in Section III for the case of an exponentially rising density of states. (The extent to which these results generalize to non-identical densities of states is briefly discussed for the exponential case in Endnote [29] to Section III.)
E. Results on the origin of entropy

Although it is not indicated in our title, besides our main question concerning the origin of thermality, we shall be greatly concerned throughout the paper, with the origin of entropy. And we are particularly interested in understanding how the very large entropies of black holes come about.

To this end, we will obtain formulae (Equations (54), (55) in Section V and Equations (69) and (70) in Section VI) for the entropy of our system, S, on both traditional and modern approaches, when system and energy bath both have either identical exponential or identical quadratic exponential densities of states. (We will also obtain formulae for the mean energy of S and B.) In the traditional approach, this is simply the mean entropy of the reduced density operator of the system when the totem is in a microcanonical state with given energy, $E$. In the modern approach, we remark, first, that, for every pure totem state, whether or not S and B have identical densities of states, the system entropy is necessarily always equal to the energy-bath entropy and both of these quantities are, in fact, identical $\rho_S$ with the {system} - {energy bath} entanglement entropy. Second, the value of the entropy in the modern case is to be interpreted, in the light of our proposition, as the value that the system entropy (= energy-bath entropy) = {system} - {energy bath} entanglement entropy) of a randomly chosen totem pure state will, with very high probability, be very close to. One of the most significant of our overall conclusions, dependent on our proposition, which we argue for in Part 2, is the fact that there is such a value at all – i.e. the fact that, with our basic general assumptions and for system and energy-bath densities of states of the sorts we discuss, the vast majority of totem states will have a system entropy close to one single value, namely $-\text{tr}(\rho_S^{\text{modapprox}} \log(\rho_S^{\text{modapprox}}))$. In terms of the language of Quantum Information Theory, this may be stated in the following way (below we temporarily suspend our terminological conventions, calling both S and B ‘systems’ and our ‘totem’ the ‘total system’):

Given two comparably-sized large systems, (S and B), which are either uncoupled or weakly coupled, then (for physically reasonable densities of states and even some maybe physically unreasonable ones) if their total state is a random pure state, their degree of entanglement (as measured by their entanglement entropy, $S$) will, with high probability, be close to the single value $-\text{tr}(\rho_S^{\text{modapprox}} \log(\rho_S^{\text{modapprox}}))$.

(Similarly, we expect that the mean value of the energy of the system, $S$, will, with high probability, be close to the single value $-\text{tr}(\rho_S^{\text{modapprox}} H_S)$. Indeed we expect the full energy probability density function, $P_S(\epsilon)$ [and hence also other moments of the energy], of S to be, with high probability, close to that of $\rho_S^{\text{modapprox}}$ [and similarly with $S$ replaced by B].)

Our results are that, for a totem with total energy $E$, for identical exponentially rising densities of states, $\sigma_S(\epsilon) = \sigma_B(\epsilon) = ce^{b\epsilon}$, on the traditional approach, the entropy, $S_S^{\text{microc}}$, will be $bE/2$ (up to a logarithmic correction) while, on the modern approach, the entropy (i.e. the single value as discussed in the previous paragraph) $S_S^{\text{modapprox}}$, will be $bE/4$ (up to a logarithmic correction). For identical quadratic exponential densities of states, $\sigma_S(\epsilon) = \sigma_B(\epsilon) = Ke^{q\epsilon^2}$, we find that $S_S^{\text{microc}} = qE^2/2$ (up to a correction of order 1 in $E$), while $S_S^{\text{modapprox}}$ will be tiny (i.e. a term of order 1 in $E$). (In both traditional and modern cases and with both equal exponential and equal quadratic exponential densities of states the mean energy of both system and energy bath will, of course be $E/2$ – in the modern case, ‘mean energy’ here meaning the value, $-\text{tr}(\rho_S^{\text{modapprox}} H_S)$, that the mean energy of a random pure totem state will most likely be very close to.)

F. Outline of the rest of the paper

We shall give full details of the results outlined in Section II in Part 1, the main sections of which comprise Section II which discusses the case where the density of states of both system and energy bath goes as a power of the energy, Sections III and IV which discuss the exponential case, and Section V which discusses the quadratic exponential case. Section VI develops the mathematical formalism to enable efficient computation of the expected energy and entropy of system, S, and energy bath, B, for the states $\rho_S^{\text{microc}}$ and $\rho_S^{\text{modapprox}}$ and this formalism is applied in Sections V and VI to obtain formulae for these quantities in the cases of exponential and quadratic exponential densities of states.

Two further sections, VII and VIII, discuss some further related matters and can be skipped on a first reading. Section VII discusses the special features of the entropy, in both modern and microcanonical cases, when the densities of states of system and energy bath are such that the energy probability density (10) is sharply peaked (as is, for example, the case for our power law densities of states) and derives some general formulae which enable us, e.g. to calculate the entropy for the states considered in Section II. In passing, we clarify the relation with some traditional work on the microcanonical ensemble (where peaks are normally presupposed) and dispel some myths. We also discuss the connection between the sum of the entropies of the partial states of system and energy bath with the totem entropy log($M$). In Section VIII we point out that if some of our basic assumptions are relaxed, then the prospects for systems
to become hot become much less constrained and, in particular, there are ways in which a system can be hot while
the totem is in a pure state which differ from the ‘modern’ scenarios we discuss below. In particular, we discuss the
notion of ‘purification’ (closely related to ‘thermofield dynamics’).

The entropy formulae we obtain in Sections III, V and VI (as outlined at the end of Section I E) will play an
important role in Section IX and in two companion papers [22, 23] where we discuss the application of the ideas and
formulae of these sections to the theory of quantum black holes. In Section IX A we point out an intriguing resemblance
between our entropy and temperature formulae for quadratic exponential densities of states in the microcanonical
strand of Section VI with Hawking’s energy and temperature formulae for (Schwarzschild) quantum black holes and
point out an apparent lack of success for the modern strand of Section VI in modelling black holes. However, we
argue that it is difficult to conclude anything decisive from these observations since (at least in a description in terms
of a quantized Einsteinian metric) black holes presumably do not satisfy the basic assumptions underlying our results
here – in particular our assumption (see Equation (2)) of weak coupling.

What seems more promising is a connection between the formulae and results for entropy and temperature which
we obtain in Sections III and V for exponentially growing densities of states and scenarios in which quantum black
holes are viewed as strong string-coupling limits of certain states of weakly coupled strings. In Section IX B and in
our two companion papers, [22] and [23], we recall some of the existing work [38–41] in this direction, and point out
that, despite its great computational success, what is computed in this work is the degeneracy of certain black hole
states; the fact that the resulting degeneracy formulae happen to agree with the previously known values of black hole
entropy does not seem to have been explained hitherto. We then go on to propose a modification of the existing string
theory scenario, and in particular of the work of Susskind [38] and Horowitz and Polchinski [39, 41] based on the
modern strand of the present paper and on our matter-gravity entanglement hypothesis and our entanglement picture
of black hole equilibrium (see Section I B). We argue that this modified scenario, which is based on an understanding
of black hole equilibrium states as strong string-coupling limits of equilibria involving a long string coupled to a
stringy atmosphere, does offer an explanation of black hole entropy and thereby also a satisfactory resolution to the
Information Loss Puzzle. The companion paper [22] gives a brief announcement of the main results of the present
paper with a focus on the main results and formalism, as well as discussing further our matter-gravity entanglement
hypothesis and outlining the application of that, with the results of Sections III and V, to this string scenario. The
further companion paper [23] develops the string scenario further.

Part 2, which comprises Sections X, XI, XII and XIII, clarifies the meaning of Equation (15) and presents our
arguments in favour of our proposition in Section I D. A fuller description of the contents of Part 2 is given towards
the end of Section X.

Part 1: Results for power law, exponential and quadratic exponential (equal) densities
of states

II. POWER-LAW DENSITIES OF STATES

If S and B have densities of states as in (18) then, by (9) and the remarks in the subsequent paragraph, we have
that \( M \), i.e. the total number of totem states with energy in \([E, E + \Delta]\), is given by

\[
M = A_S A_B \Delta \int_{0}^{E} \epsilon^{N_S} \left(E - \epsilon\right)^{N_B} d \epsilon
\]  

(19)

which can be rewritten

\[
M = A_S A_B \Delta E^{N_S + N_B + 1} B(N_S + 1, N_B + 1)
\]  

(20)

where \( B(x, y) \) is the usual beta function (see e.g. [20]) – related to the gamma and factorial functions by

\[
B(x + 1, y + 1) = \frac{\Gamma(x + 1)\Gamma(y + 1)}{\Gamma(x + y + 2)} = \frac{x! y!}{(x + y)! (x + y + 1)}.
\]  

(21)

(For fractional arguments, we take \( x! \) to mean \( \Gamma(x + 1) \).) On the other hand, the number of such totem states with
system energy in an interval \([\epsilon, \epsilon + \delta]\) will, for suitable \( \delta \), be well-approximated by

\[
P_S(\epsilon) \delta = A_S A_B \Delta M^{-1} \epsilon^{N_S} \left(E - \epsilon\right)^{N_B}
\]
Thus, combining (20), (22) and (21) we have that
\[ P_S(\epsilon) = \frac{N_S + N_B + 1}{E} b(N_S; N_S + N_B, \frac{\epsilon}{E}) \] (23)
where (see e.g. [26])
\[ b(k; n, p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \] (24)
is, when \( n \) and \( k \) are integers, the binomial distribution function which has the famous interpretation as the probability that \( n \) ‘Bernoulli’ trials, each with probability \( p \) for success and \( q = 1 - p \) for failure, result in \( k \) successes and \( n - k \) failures. In order to take advantage of the insight afforded by this connection with probability theory we shall (with negligible error when \( N_S \) and \( N_B \) are large) assume from now on that \( N_S \) and \( N_B \), if not already integers, are replaced by their nearest integers.

First we notice that we may use the well-known connection between the binomial and the Poisson distribution to give an alternative derivation of the fact that, in the limit as \( E \) and \( N_B \) grow while \( N_S \) remains constant and the ratio \( N_B/E \) converges to \( \beta \), S’s energy probability density, \( P_S(\epsilon) \) (23), converges to the Gibbs energy probability density \( P_{S,\beta}^{\text{Gibbs}}(\epsilon) \) (see (17) and (14)) with inverse temperature \( \beta = N_B/E \) for \( \sigma_S(\epsilon) \) as in (18) – the latter Gibbs energy probability density being given explicitly by
\[ P_{S,\beta}^{\text{Gibbs}}(\epsilon) = \frac{\beta^{N_S+1} e^{N_S e^{-\beta \epsilon}}}{N_S!} \] (25)
as one sees from (17) and (18) after easily checking from (14) and (18) that
\[ Z_{S,\beta} = A_S^{-1} \frac{N_S!}{\beta^{N_S+1}}. \]

This convergence result is of course a special case (i.e. the case where \( S \), as well as \( B \), has a power-law density of states) of an easy corollary both of the traditional thermality result (on the total microcanonical state approach) and (bearing in mind the equality of the energy probability density for both (11) and (15)) of the ‘Canonical Typicality’ result of Goldstein Lebowitz et al. (on the ‘modern’ total pure state approach) which, as we discussed in Section I, both hold in the same limit; we shall see shortly that the alternative proof which we next give for this corollary easily implies an alternative proof to the traditional thermality result itself and thus also, by a remark we made in Section ID to an alternative argument for ‘Canonical Typicality’ when the system and energy-bath densities of states both have power-law form.

As Feller puts it in [26], “If \( n \) is large and \( p \) is small, whereas the product \( \lambda = np \) is of moderate magnitude” then the binomial distribution goes over to the Poisson distribution, i.e.
\[ b(k; n, p) \propto \frac{\lambda^k}{k!} e^{-\lambda}. \] (26)
In particular (cf. e.g. [27]) for fixed \( k \), the right hand side of (26) is the limit of \( b(n; k, p) \) as \( n \to \infty \) while \( p \to 0 \) in such a way that \( np \to \lambda \). From this, and (23), we easily conclude that the limit, as \( E \to \infty \) while \( N_B/E \to \beta \) with \( N_S \) fixed, of \( P_S(\epsilon) \) is equal to \( P_{S,\beta}^{\text{Gibbs}}(\epsilon) \) (25). So, to summarize, in the appropriate limit of a large energy bath, the energy probability density of \( S \) goes over to the energy probability density of the appropriate Gibbs state:
\[ P_S(\epsilon) \to P_{S,\beta}^{\text{Gibbs}}(\epsilon) = \frac{\beta^{N_S+1} e^{N_S e^{-\beta \epsilon}}}{N_S!}. \] (27)
We remark that, by inspecting (17) and (14) this is easily seen to be equivalent to the statement that, in the same limit,
\[ M^{-1} n_B(E - \epsilon) \to Z_{S,\beta} e^{-\beta \epsilon}, \]
and, by inspecting (11) and (13), one easily sees that this implies that in the same limit
\[ P_S^{\text{micro}} \to P_{S,\beta}^{\text{Gibbs}}. \]
thus providing the alternative proof, which we promised in Section 4 of the traditional result on the thermality of a small system in contact with an energy bath in the traditional limit of a large energy bath, in the case where both the energy bath, B, and system, S, have a power-law density of states (and by our remark in Section 10 thus also providing an alternative proof of ‘Canonical Typicality’ for such S and B).

We will now demonstrate, however, that, when S and B are of comparable size, then, if they both have power-law densities of states as in (18), both the total microcanonical state approach and the ‘modern’ approach (i.e. with a total pure state) predict that the reduced density operators of each of S and B will be quite different from thermal!

We shall show this by showing that the energy probability density of each of S and B (which we again recall from the paragraph after (15) is the same in each approach) will have a quite different form from the thermal form of $P_{\text{Gibbs}}(\epsilon)$.

First we notice that, when $k$ is a fixed fraction, $pn$, of $n$ (in such a way that $0 < p < 1$ and also $pn$ is an integer) then, if $p$ and $n$ are regarded as fixed, the binomial distribution function (24) $b(pn; n, p')$ is maximized when $p' = p$ and we easily obtain the approximation (28) (now writing $p = p' = p + x$)

$$b(pn; n, p + x) \simeq \frac{1}{\sqrt{2\pi np(1-p)}} \exp \left( -\frac{nx^2}{2p(1-p)} \right).$$

(28) is obtained by expressing the left hand side in terms of factorials and powers according to (24). We then adopt Stirling’s approximation, $N! \simeq \sqrt{2\pi N}N^{N+\frac{1}{2}}e^{-N}$ for each of the factorials and, introducing $q = 1 - p$, write the term $(p + x)^n q^n (1 - p - x)^n (1 - x/q)^n$ and approximate the latter by $\exp(-nx^2/2pq)$. Clearly, as long as $n$ is extremely large and $p$ is not extremely close to zero or 1, then this will be an excellent approximation.

Combining (23) with the definition of $p$ before (28) we see that, if we identify $n$ with $N_S + N_B$, then

$$P_S(\epsilon) = \frac{n + 1}{E} b\left(pn; n, \frac{\epsilon}{E}\right)$$

where

$$p = \frac{N_S}{N_S + N_B}$$

and that, provided $n$ is extremely large and S and B are of ‘comparable size’, which of course, in view of (29), corresponds exactly to $p$ not being extremely close to zero or 1, then, by (28), to a high degree of accuracy, we will have the approximation

$$P_S(\epsilon) \simeq \frac{1}{E} \sqrt{\frac{\gamma}{\pi}} \exp \left( -\frac{\gamma(\epsilon - \epsilon_0)^2}{E^2} \right)$$

(30)

i.e. a Gaussian with a peak located (See Section VII A for an alternative perspective on Equation (31)) at

$$\epsilon_0 = pE \quad (p \text{ as in (29)}) = \frac{N_S}{N_S + N_B}E$$

(31)

and

$$\gamma = \frac{n}{2p(1-p)} = \frac{(N_S + N_B)^3}{2N_SN_B}$$

(32)

and there will of course be an obvious counterpart formula for the energy probability density, $P_B$, of B, similar to the above formula but with $p$ replaced by $1 - p$. (This of course changes the value of $\epsilon_0$ but not of $\gamma$.) So the energy of S will be in a Gaussian band around a most likely energy of $\epsilon_0$, the energy of B will be in a Gaussian band around a most likely energy of $E - \epsilon_0$, each having the same width which will be $E$ divided by a number (i.e. $\sqrt{2\gamma}$) which is of the order of the square root of either of the (comparable!) numbers $N_S$, $N_B$. Moreover it is easy to see that, in both the traditional microcanonical and the modern total pure state approaches, the two energy probability densities will be perfectly anticorrelated – i.e. when S has energy in a small interval around $\epsilon$, then B with have energy in a similar small interval around $E - \epsilon$.

Above, by ‘width’ we mean the standard deviation, $s$, from the mean of the energy probability density, i.e.

$$s = (\epsilon^2 - \bar{\epsilon}^2)^{1/2}$$
FIG. 1: Plot of the energy probability density \( P_S(\epsilon) \) in the case S and B have the same density of states \( \sigma(\epsilon) = A\epsilon^N \) for the ‘unusually’ small value \( N = 10 \).

where

\[
\bar{\epsilon} = \int_0^E \epsilon^n P_S(\epsilon) d\epsilon
\]

\( (= \text{tr}(\rho_S^{\text{microc}} H_S^n) = \text{tr}(\rho_S^{\text{modapprox}} H_S^n)) \) – cf. Section IV.

(In the special case where \( N_S = N_B = N \) say, one sees that \( \gamma \) becomes \( 4N \) so the width, \( s \), will be \( E/2\sqrt{2N} \).) As we anticipated, this is a qualitatively very different behaviour from the energy probability density of thermal states and we conclude therefore, as promised, that, in both the traditional total microcanonical and the modern total pure state approaches, the reduced density operators of S and B must, when, S and B are of comparable size, be quite different in character from thermal density operators. To illustrate this point, we include a figure (Figure 1) for the energy probability density, \( P_S(\epsilon) \), in the case S and B have the same density of states \( \sigma(\epsilon) = A\epsilon^N \) for the (unrealistically small) value \( N = 10 \) and a comparison figure, Figure 2, showing the the energy probability density, \( P_{S,\beta}^\text{Gibbs}(\epsilon) \) for a thermal state at the inverse temperature, \( \beta = 2(N + 1)/E \), chosen so that the mean energy takes the same value, \( E/2 \) – again in the case \( N = 10 \).

For the sake of a quantitative result, we note that, for general \( N \), the width, \( s \), of the energy probability density, \( P_{S,\beta}^\text{Gibbs}(\epsilon) \), of this comparison thermal state is (as is easily calculated) \( E/2\sqrt{N + 1} \) – i.e. (to a very good approximation for large \( N \)) a factor of \( \sqrt{2} \) wider than the width of \( P_S(\epsilon) \) while the height is (again by an easy calculation) a factor of \( \sqrt{2} \) smaller.

We shall postpone to Section VII a calculation of the (microcanonical and modern) entropies of S and B for general \( N_S \) and \( N_B \). Suffice it to remark that, like the width, \( s \), the microcanonical entropy of S, differs, in general, from its value in the comparison thermal state at inverse temperature \( \beta = 2(N + 1)/E \), albeit the difference is just a ‘small’ constant (it is smaller by \( \approx \log 2/2 \) independent of \( \epsilon \).

Finally, we remark that, in this power-law density-of-states case, it is clear from the developments in this section that the ‘canonical’ (i.e. thermal) behaviour of \( \rho_S^{\text{microc}} \) (or indeed of \( \rho_S^{\text{modapprox}} \)) in the case that the system, S, is very much smaller than the energy bath, B, may be reconciled with the above-discussed Gaussian behaviour, when S and B are of comparable size, in that the relationship between the two may be regarded as an instance of the well-
FIG. 2: Plot of the energy probability density, $P_{S, \beta}^{Gibbs}(\epsilon)$ for the thermal state at inverse temperature, $\beta$, on our system, $S$, 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 $\beta$ for which the mean energy is $E/2$). To be contrasted with the $P_S(\epsilon)$ of Figure 1.

known relationship (see e.g. [26] or [27]) between the Poisson and Gaussian distributions in probability theory. (This obviously easily follows from the way we derived, above, both the canonical behaviour and the Gaussian behaviour as limits of the binomial distribution.)

III. EXPONENTIALLY RISING DENSITIES OF STATES

We now turn to discuss the quite different behaviour of the reduced density operators $\rho^{\text{micro}}_S$ and $\rho^{\text{modapprox}}_S$ when the densities of states of $S$ and $B$ increase exponentially. We shall confine our interest here to the case where both densities of states, $\sigma_S$ and $\sigma_B$, behave as $c e^{b \epsilon}$ with the same constants $c$ and $b$ in each expression:

$$\sigma_S(\epsilon) = c e^{b \epsilon}, \quad \sigma_B(\epsilon) = c e^{b \epsilon}. \quad (34)$$

We remark, however, that, as may quite easily be checked, allowing different values of $c$ (say $c_S$ in the first formula and $c_B$ in the second) will not essentially change our conclusions [29].

The normalization constant $M$ is now easily seen – either by using (9) or, on recalling (4), by using (7) – to be given by

$$M = c^2 e^{b E} E \Delta. \quad (35)$$

We note that this will be large provided neither $c$ nor $\Delta$ are ‘too small’ and provided also

$$bE \gg 1 \quad (36)$$

which will hold in cases of interest.

The formula, (11) for $\rho^{\text{micro}}_S$ is then easily seen to coincide with the formula, (13), for a thermal density operator $\rho^{\text{Gibbs}}_{S, \beta}$, for the density of states $\sigma(\epsilon)$ as in (34) at inverse temperature $\beta = b$, provided the latter formula is modified
so that the sum over $\epsilon$ is truncated at the upper energy, $E$ and the partition function, $Z_{S,\beta}$, is replaced by $cE$. Of course, the un-truncated formula \( \rho_{\text{micro}}^S \) will only make mathematical sense for $\beta > b$. Nevertheless, the reduced density operator $\rho_{\text{micro}}^S$ (and similarly also $\rho_{\text{micro}}^B$) clearly deserves to be called an approximately thermal state at inverse temperature $b$. (This will generalize from equal systems to comparably sized systems if, by this, we mean systems with densities of states with unequal $c_S$ and $c_B$ as discussed in Endnote \[29\]). We shall refer to the relevant notion of being approximately thermal here as being $E$-approximately thermal.

Turning from the traditional total microcanonical state approach to the modern total pure state approach, we see, on substituting (34) into (15) and noting that $E_c$ will obviously become $E/2$, that the $\epsilon$-summand in (15) still agrees with the $\epsilon$-summand in (13) at inverse temperature $\beta = b$ up to energy $E/2$ and, moreover, as always (cf. after Equation (15)) the system energy probability density of $\rho_{\text{micro}}^S$ is equal to that of $\rho_{\text{micro}}^S$ and thus it agrees with the energy probability density of a Gibbs state, for the same density of states, up to energy $E$. We shall refer to the relevant notion of being approximately thermal here (i.e. agreement of the summand in the formula (15) with the summand in the formula (13) up to $\epsilon = E/2$ – with a suitable change in the value of $Z_{S,\beta}$ – and agreement of the energy probability density up to $E$) as being $E$-approximately semi-thermal.

We note here that, with the densities of states as in (34), the energy probability density $P_S(\epsilon)$, which we recall by \[10\] is given in general by

$$P_S(\epsilon) = \frac{\Delta M}{M} \sigma_S(\epsilon) \sigma_B(E - \epsilon),$$

will, with $M$ as in (36) and $\sigma_S$ and $\sigma_B$ as in (34), reduce to

$$P_S(\epsilon) = \frac{1}{E}.$$ \hspace{1cm} (37)

See Figure 3. Of course (cf. the paragraph after Equation (32)) the energies of S and B will, again, be perfectly anticorrelated.

Similar results to the the above results for S will obviously hold for B. We thus conclude, in fulfillment of our promise (cf. the start of Section I C) that, with the appropriate meaning in each case for the expression “approximately thermal”, as above, when the densities of states of S and B take the exponential form of (34) then – in contrast to
the situation for power-law densities of states discussed in Section IV – on both the traditional total microcanonical state approach and also on the modern total pure state approach, the reduced density operators of both S and B will be approximately thermal (at inverse temperature \( \beta = b \)) in an appropriate sense.

IV. GENERAL FORMULAE FOR MEAN ENERGY AND FOR ENTROPY

It is interesting (especially in preparation for the discussion in Section IX and in our companion papers \[22, 23\] about the connection with quantum black hole physics both of the results in Sections III and V concerning densities of states which grow exponentially with energy and, in Section VI, concerning those which grow as quadratic exponentials in the energy) to calculate the mean energy, \( \bar{\epsilon} \), and also the von Neumann entropy, \( S \), for each of the density operators, \( \rho_S^{\text{micro}} ; \rho_S^{\text{modapprox}} \) (and also for \( \rho_B^{\text{micro}} ; \rho_B^{\text{modapprox}} \)). The former density operator will just give the usual mean energy and von Neumann entropy (defined as in (16)) of our system, S, when the totem is in the microcanonical ensemble. The latter density operator will, according to our proposition in Section I D, give an energy value and an entropy which grow exponentially with energy and, in Section VI, concerning those which grow as quadratic exponentials in energy.

By (11), (15), and (7), we have, in general, that, with obvious notation, \( \bar{\epsilon}_S^{\text{micro}} := \text{tr}(\rho_S^{\text{micro}} H_S) = M^{-1} \sum_{\epsilon = \Delta}^E \epsilon n_S(\epsilon) n_B(\epsilon) \).

Similarly

\[ \bar{\epsilon}_B^{\text{micro}} := \text{tr}(\rho_B^{\text{micro}} H_B) = M^{-1} \sum_{\epsilon = \Delta}^E \epsilon n_B(\epsilon) n_S(\epsilon) \]

and one easily sees that necessarily, \( \bar{\epsilon}_B^{\text{micro}} = E - \bar{\epsilon}_S^{\text{micro}} \). Moreover, we have

\[ \bar{\epsilon}_S^{\text{modapprox}} = \text{tr}(\rho_S^{\text{modapprox}} H_S) \]

\[ = M^{-1} \left( \sum_{\epsilon = E_c}^E \epsilon n_B(\epsilon) n_S(\epsilon) + \sum_{\epsilon = E_c + \Delta}^E \epsilon n_S(\epsilon) n_B(\epsilon) \right) \]

(38)

which is easily seen to be equal to \( \bar{\epsilon}_S^{\text{micro}} \). Similarly, \( \bar{\epsilon}_B^{\text{modapprox}} = \bar{\epsilon}_B^{\text{micro}} \).

On the other hand, by (11) and (16), we have

\[ S_S^{\text{micro}} := -\text{tr}(\rho_S^{\text{micro}} \log \rho_S^{\text{micro}}) = -M^{-1} \sum_{\epsilon = \Delta}^E n_S(\epsilon) n_B(\epsilon) \log(M^{-1} n_B(\epsilon)) \]

(40)

and by (15) and (16)

\[ S_S^{\text{modapprox}} := -\text{tr}(\rho_S^{\text{modapprox}} \log \rho_S^{\text{modapprox}}) \]

\[ = -M^{-1} \left( \sum_{\epsilon = \Delta}^E n_S(\epsilon) n_B(\epsilon) \log(M^{-1} n_B(\epsilon)) + \sum_{\epsilon = E_c + \Delta}^E n_S(\epsilon) n_B(\epsilon) \log(M^{-1} n_S(\epsilon)) \right) \]

(41)

and similarly with S replaced by B. We remark that it is not difficult to see from (15) and the counterpart equation for \( \rho_B^{\text{modapprox}} \) that \( \rho_S^{\text{modapprox}} \) will necessarily equal \( S_B^{\text{modapprox}} \). This is of course consistent with the fact that, by the general result recalled in Endnote \[48\], for any pure totem state, \( \Psi \), we necessarily have that the von Neumann entropies of the resulting reduced density operators \( \rho_S^{\text{modern}} \) and \( \rho_B^{\text{modern}} \) will necessarily be equal. After all, as we claim in our Proposition in Section I D and argue in Part 2, for random \( \Psi \), \( \rho_S^{\text{modapprox}} \) most probably gives a very good approximation of \( \rho_S^{\text{modern}} \) and \( \rho_B^{\text{modapprox}} \) of \( \rho_B^{\text{modern}} \).
By referring to the second equality in (10), it is easy to see that the formulae for \(S_{\text{micro}}\) and \(S_{\text{modapprox}}\) in (40) and (41) may be rearranged to give the following useful alternative expressions:

\[
S_{\text{micro}}^{\epsilon} = \Delta \sum_{\epsilon = \Delta}^{E} P_{\epsilon}(\epsilon) \log \left( \frac{n_{S}(\epsilon)}{P_{\epsilon}(\epsilon) \Delta} \right),
\]

\[
S_{\text{modapprox}}^{\epsilon} = \Delta \left( \sum_{\epsilon = \Delta}^{E} P_{\epsilon}(\epsilon) \log \left( \frac{n_{S}(\epsilon)}{P_{\epsilon}(\epsilon) \Delta} \right) + \sum_{\epsilon = E_{c} + \Delta}^{E} P_{\epsilon}(\epsilon) \log \left( \frac{n_{B}(E - \epsilon)}{P_{\epsilon}(\epsilon) \Delta} \right) \right).
\]

Referring to (4) and making the replacement (8) (and with the proviso made in the cautionary remark in [14]) we see that (42) and (43) have, as their continuum versions:

\[
S_{\text{micro}}^{\epsilon} = \int_{0}^{E} P_{\epsilon}(\epsilon) \log \left( \frac{\sigma_{S}(\epsilon)}{P_{\epsilon}(\epsilon)} \right) \, d\epsilon,
\]

\[
S_{\text{modapprox}}^{\epsilon} = \int_{0}^{E_{c}} P_{\epsilon}(\epsilon) \log \left( \frac{\sigma_{S}(\epsilon)}{P_{\epsilon}(\epsilon)} \right) \, d\epsilon + \int_{E_{c}}^{E} P_{\epsilon}(\epsilon) \log \left( \frac{\sigma_{B}(E - \epsilon)}{P_{\epsilon}(\epsilon)} \right) \, d\epsilon.
\]

We notice, in passing, that the absence of the quantity \(\Delta\) (or of any quantity that scales with \(\Delta\)) in the formulae \(S_{\text{micro}}^{\epsilon}\) and \(S_{\text{modapprox}}^{\epsilon}\) shows us the interesting fact that (for \(\Delta\) in an appropriate not-too-large and not-too-small range, and to what, in typical applications will be an extremely good approximation) neither of these entropies depends on \(\Delta\)!

Finally, further useful insight concerning the form of Equations (42) and (43) can be had by noticing that they can alternatively be derived as corollaries of the following easily proved Lemma, which we will also need to refer to in Section XIII in Part 2.

**Lemma:** Given density operators \(\rho_{1}, \rho_{2}, \ldots\) on Hilbert spaces \(H_{1}, H_{2}, \ldots\) respectively, with von Neumann entropies \(S(\rho_{1}), S(\rho_{2}), \ldots\) and given positive real numbers \(\lambda_{1}, \lambda_{2}, \ldots\) with \(\sum_{i} \lambda_{i} = 1\). Then the density operator \(\rho = \lambda_{1} \rho_{1} \oplus \lambda_{2} \rho_{2} \oplus \ldots\) on the direct sum Hilbert space \(H = H_{1} \oplus H_{2} \ldots\) will have an entropy, \(S\), given by

\[
S = \sum_{i} \lambda_{i} S(\rho_{i}) - \sum_{i} \lambda_{i} \log \lambda_{i}.
\]
V. FORMULAE FOR MEAN ENERGY AND ENTROPY FOR EXPONENTIALLY RISING DENSITIES OF STATES

Specialising to \( n_S \) and \( n_B \) given by 
\[
\sigma_S(\epsilon) \Delta \quad \text{and} \quad \sigma_B(\epsilon) \Delta
\]
with \( \sigma_S \) and \( \sigma_B \) as in (34), we have that
\[
\bar{\epsilon}_S^{\text{micro}} = \bar{\epsilon}_S^{\text{modapprox}} = E/2,
\]
and similarly with \( S \) replaced by \( B \). These results for the mean energies of \( S \) and \( B \) are of course anyway obvious since the assumption of very weak coupling (see (2) and the subsequent paragraph) implies that the mean energies of \( S \) and \( B \) will add to \( E \), while (34) is symmetric under the replacement of \( S \) by \( B \). However, we remark that (53) turns out to remain exactly unchanged even when the densities of states are generalized so as to have different pre-factors \( c_S \) and \( c_B \) (see Endnote [29]). (Returning to the case of equal densities of states) we caution that the mean energies are just that, averages; they are not in any sense ‘most likely’ energies. In fact, as we saw in Section III, the energy probability density (see (37) and Figure 3) is flat!

It is also straightforward to calculate, using the formulae of Section IV, that the entropies take the values
\[
S_S^{\text{micro}} = bE/2 + \log(cE),
\]
(54)
\[
S_S^{\text{modapprox}} = bE/4 + \log(cE),
\]
(55)
and similarly with \( S \) replaced by \( B \). (Again, see Endnote [29] for the generalization to different prefactors, \( c_S \) and \( c_B \), in the first and second formulae of (34)).

In particular, inserting the formulae (34) and (37) for \( \sigma_S \) and \( P_S \) in (42) and (43) we obtain
\[
S_S^{\text{micro}} = \sum_{\epsilon=\Delta}^{E} \frac{\Delta}{E} (b \epsilon + \log(c \Delta) - \log(\Delta/E))
\]
\[
= \frac{\Delta}{E} \left( b\Delta (E/\Delta)(E/\Delta + 1) / 2 + (E/\Delta) \log(c E) \right)
\]
\[
\simeq \frac{bE}{2} + \log(cE)
\]
(56)
while (assuming \( E/\Delta \) is even)
\[
S_S^{\text{modapprox}} = \sum_{\epsilon=\Delta}^{E/2} \frac{\Delta}{E} (b \epsilon + \log(c \Delta) - \log(\Delta/E)) + \sum_{E/2+\Delta}^{E} \frac{\Delta}{E} (b(E - \epsilon) + \log(c \Delta) - \log(\Delta/E))
\]
\[
= 2 \sum_{\epsilon=\Delta}^{E/2} \frac{\Delta}{E} (b \epsilon + \log(c \Delta) - \log(\Delta/E))
\]
\[
= 2 \frac{\Delta}{E} \left( b\Delta (E/2\Delta)(E/2\Delta + 1) / 2 + (E/\Delta) \log(c E) \right)
\]
\[
\simeq \frac{bE}{4} + \log(cE)
\]
(57)
which are the formulae (54) and (55). In the calculations above, we need to recall that the sums in (11) and (15), and hence also in the direct sums in (47) and (48) and in the above equations, are over \( \epsilon \) values which are positive-integer multiples of \( \Delta \).

We remark that the leading behaviour of \( S_S^{\text{micro}} \) (i.e. the term, \( bE/2 \), which remains when we ignore the logarithmic terms in (54)) arises, in (say) the continuum version, (44), of our general formula for \( S_S^{\text{micro}} \) by replacing the logarithm in this formula by its ‘main part’, by which we mean the exponent, \( b\epsilon \), in the formula, (34) \( \sigma_S(\epsilon) = ce^{b\epsilon} \). Similarly, the leading behaviour of \( S_S^{\text{modapprox}} \) (i.e. the term \( bE/4 \) in (55)) arises by setting \( E_c = E/2 \) in (45) and noticing that the main parts of the two logarithms in this formula are (in order) \( b\epsilon \) and \( b(E/2 - \epsilon) \).
VI. DENSITIES OF STATES WHICH GROW AS QUADRATIC EXPONENTIALS

Next we discuss the behaviour of the reduced density operators, $\rho_{S}^{\text{micro}}$ and $\rho_{B}^{\text{modapprox}}$, when the densities of states of S and B each increase as the exponential of a constant times the square of the energy. We shall confine our interest to the case where both densities of states, $\sigma_S$ and $\sigma_B$, behave as $K e^{q\epsilon^2}$ with the same constants, $K$ and $q$:

$$\sigma_S(\epsilon) = K e^{q\epsilon^2}, \quad \sigma_B(\epsilon) = K e^{q\epsilon^2}$$

and shall just discuss the cases of $\rho_{S}^{\text{micro}}$ and $\rho_{B}^{\text{modapprox}}$ – those of $\rho_{S}^{\text{micro}}$ and $\rho_{B}^{\text{modapprox}}$ obviously being similar.

We shall assume that $qE^2 \gg 1$. (59)

The energy probability density, $P_S(\epsilon)$ (10), now takes the form

$$P_S(\epsilon) = \frac{\Delta}{M} K^2 e^{q\epsilon^2} e^{-2q\epsilon(E-\epsilon)}.$$ (60)

and we sketch its graph in Figure 4.

We note that it is symmetric about $E/2$ and also, in view of (59), $P_S(\epsilon)$ is very close to zero except when $\epsilon$ is close to 0 or to $E$, where it is well approximated by the exponentially decaying function $\frac{\Delta}{M} K^2 e^{q\epsilon^2} e^{-2q\epsilon E}$ (near $\epsilon = 0$), and by the exponentially rising function $\frac{\Delta}{M} K^2 e^{q\epsilon^2} e^{2q\epsilon(E-\epsilon)}$ (near $\epsilon = E$). Approximating the integral from 0 to $E$ of $P_S(\epsilon)$ by the sum of the (equal) integrals (from 0 to $\infty$ and from $-\infty$ to $E$) of these exponential approximations, and demanding that the result must equal 1 we see that $P_S(\epsilon)$ will be well approximated on its domain $[0, E]$ by

$$P_S(\epsilon) \simeq qE(e^{-2q\epsilon E} + e^{2q\epsilon(E-E)})$$

from which we infer that the normalization constant, $M$, will be well approximated by

$M \simeq K^2 qE^2 \Delta$. (62)
FIG. 5: Comparison figure for the (non-normalizable!) "energy probability density" (64) of the “thermal state”, $P^{\text{Gibbs}}_{S,\beta}(\epsilon)$ at inverse temperature $\beta = 2qE$, again for the value $qE^2 = 10$

We pause here to record that (cf. (53)) either by calculation or by the above-noted symmetry of $P_S(\epsilon)$, we will obviously have that the mean energy of both S and E will be $E/2$:

$$\bar{\epsilon}_\text{micro} = \bar{\epsilon}_\text{modapprox} = E/2,$$  \hfill (63)

However, (cf. our remark after Equation (53)) even more emphatically than in the case of exponentially rising densities of states, these are not 'most likely energies'. In fact, the energy probability density, $P_S(\epsilon)$, tells us that the energy (say of S) will be highly likely, and with equal likelikhoods, either to be close to 0 or to be close to $E$ – and highly unlikely to be close to $E/2$ (and similarly for B). In addition of course (cf. after Equation (32) and Equation (37)) the energy of S and the energy of B will be perfectly anticorrelated. So when the energy of S is near 0, the energy of B will be near $E$, and when the energy of S is near $E$, the energy of B will be near 0.

In analogy to what we did in Sections III V we next wish to compare the formulae (60) and (61) for $P_S(\epsilon)$ with $P^{\text{Gibbs}}_{S,\beta}(\epsilon) = Ce^{-\beta\epsilon}Ke^{\beta\epsilon^2}$,  \hfill (64)

where $C$ is a suitable constant, which, but for the fact that it is not integrable on the interval $(0, \infty)$ (for any value of $\beta$) would deserve to be called 'the energy probability density of a thermal state at inverse temperature $\beta$' (cf. 17) for the same density of states (58). If $\beta E \gg 1$, then, on the interval $[0, \beta/q]$, $P^{\text{Gibbs}}_{S,\beta}(\epsilon)$ will be very close to zero except when $\epsilon$ is close to 0 or to $\beta/q$, where it will be well approximated by the exponential decay $Ce^{-\beta\epsilon}$ (near 0), and by the exponentially rising function $Ce^{\beta(\epsilon-\beta/q)}$ (near $\epsilon = \beta/q$). Put otherwise, on the interval $[0, \beta/q]$, $P^{\text{Gibbs}}_{S,\beta}(\epsilon)$ will take the approximate form

$$P^{\text{Gibbs}}_{S,\beta}(\epsilon) \simeq CK(e^{-\beta\epsilon} + e^{\beta(\epsilon-\beta/q)}),$$  \hfill (65)

Beyond $\epsilon = \beta/q$, $P^{\text{Gibbs}}_{S,\beta}(\epsilon)$ will, of course, grow rapidly. If we now choose to make the identification

$$\beta = 2qE, \hfill (66)$$
respectively, will stand in relation to $\rho$ approximately thermal. Quadratic exponentials lead to reduced density operators on $S$ and $B$ which, in the sense we have explained above, III, V, for densities of states which grow exponentially with energy, also densities of states which grow, (58), as for power-law densities of states and with many similarities (but also a few differences) to what we found in Sections III and comparison of (65) or a glance at Figures 4 and 5, immediately shows that these resemble one-another closely – each having an equally-rapidly exponentially decaying peak located near $\epsilon = 0$ and each having an equally rapidly exponentially rising, equal-sized second peak – the only discrepancy being that, in the case of $P_S(\epsilon)$, the second peak is located near $\epsilon = E$ while, in the case of $P_{S,\beta}^{\text{Gibbs}}(\epsilon)$, it is located near $\epsilon = 2E$. Thus, at low energies – and indeed at all energies, $\epsilon$, up to a little below $E$ – the energy probability density, $P_S(\epsilon)$, is closely approximated by the thermal energy probability density for $\beta = 2qE$, while there is a qualitative resemblance between $P_S(\epsilon)$ on its full interval $[0, E]$ and $P_{S,\beta}^{\text{Gibbs}}(\epsilon)$ on the interval $[0, 2E]$ (with the above-mentioned quantitative discrepancy that the second peak in $P_S(\epsilon)$ occurs near $E$ while the second peak in $P_{S,\beta}^{\text{Gibbs}}(\epsilon)$ occurs near $2E$).

Moreover, one may easily check that, except for discrepancies corresponding to the above discrepancy for the energy probability densities, the reduced density operators, $\rho_S^{\text{microc}}$ (defined as in (11)) and $\rho_S^{\text{modapprox}}$ (defined as in (45)), respectively, will stand in relation to $\rho_{S,\beta}^{\text{Gibbs}}$ (defined as in (13)) for $\beta = 2qE$, in a similar way to the relationships which we termed ‘$E$-approximately thermal’ and ‘$E$-approximately semi-thermal’ in Section III.

In conclusion, except for the discrepancy pointed out above, we may say that, in contrast again to the situation than we found, in Section V, for densities of states which grow as quadratic exponentials, there is an even more dramatic difference with many similarities (but also a few differences) to what we found in Sections III and comparison of (65) or a glance at Figures 4 and 5, immediately shows that these resemble one-another closely – each having an equally-rapidly exponentially decaying peak located near $\epsilon = 0$ and each having an equally rapidly exponentially rising, equal-sized second peak – the only discrepancy being that, in the case of $P_S(\epsilon)$, the second peak is located near $\epsilon = E$ while, in the case of $P_{S,\beta}^{\text{Gibbs}}(\epsilon)$, it is located near $\epsilon = 2E$. Thus, at low energies – and indeed at all energies, $\epsilon$, up to a little below $E$ – the energy probability density, $P_S(\epsilon)$, is closely approximated by the thermal energy probability density for $\beta = 2qE$, while there is a qualitative resemblance between $P_S(\epsilon)$ on its full interval $[0, E]$ and $P_{S,\beta}^{\text{Gibbs}}(\epsilon)$ on the interval $[0, 2E]$ (with the above-mentioned quantitative discrepancy that the second peak in $P_S(\epsilon)$ occurs near $E$ while the second peak in $P_{S,\beta}^{\text{Gibbs}}(\epsilon)$ occurs near $2E$).

Moreover, one may easily check that, except for discrepancies corresponding to the above discrepancy for the energy probability densities, the reduced density operators, $\rho_S^{\text{microc}}$ (defined as in (11)) and $\rho_S^{\text{modapprox}}$ (defined as in (45)), respectively, will stand in relation to $\rho_{S,\beta}^{\text{Gibbs}}$ (defined as in (13)) for $\beta = 2qE$, in a similar way to the relationships which we termed ‘$E$-approximately thermal’ and ‘$E$-approximately semi-thermal’ in Section III.

In conclusion, except for the discrepancy pointed out above, we may say that, in contrast again to the situation for the first approximate equality, we have replaced the logarithm in (44) by its ‘main part’ – i.e. by the exponent, $K\epsilon^2$, in the first equation in (58), and, for the second approximate equality, we have used the fact that the energy probability density, $P_S(\epsilon)$ (see (61) and Figure 4) consists of two sharp peaks, each of area 1/2, one located at $\epsilon = 0$ and one at $\epsilon = E$.

Proceeding similarly for $S_{S,\beta}^{\text{modapprox}}$, we similarly expect the leading term to be given by approximating (45) by

$$S_{S,\beta}^{\text{modapprox}} \simeq \int_0^{E/2} P_S(\epsilon) q\epsilon^2 d\epsilon + \int_{E/2}^E P_S(\epsilon) q(E - \epsilon)^2 d\epsilon \simeq 0.$$  

It is straightforward to check that the error in both (69) and (70) is only of order 1 in $E$; one needs only to be careful to realize that this is one situation in which (cf. Endnote 14) it is important to work with the discrete sum versions, (40) and (41) or alternatively (42) and (43), of our entropy formulae; if one were to work unthinkingly with (44) and (45), one might (wrongly) conclude there is a (for some values of $K$, $q$ and $E$, negative!) correction to both equations of form $\log(K/qE) + O(1)$ – the problem being caused by the steeply rising behaviour of $P_S(\epsilon)$ near $\epsilon = 0$ and $\epsilon = E$.

Thus, for our densities of states which grow as quadratic exponentials, there is an even more dramatic difference between the value of $S_{S,\beta}^{\text{microc}}$ and the value of $S_{S,\beta}^{\text{modapprox}}$ than we found, in Section V, for densities of states which grow exponentially with energy (where they differed by a factor of 2).

VI. MORE ABOUT ENTROPY

Note: The reader may wish to skip this, and the next, section on a first reading and go directly to Section IX.
A. Special facts about the entropy when the energy probability density is sharply peaked

In the case of our power-law density of states example, an alternative way of arguing that the location of the peak of the energy probability density, \( P_S(\epsilon) \), of the system, \( S \), is given by the formula \( (31) \) is to assume foreknowledge of the existence of a (single) peak in the energy probability density, \( P_S(\epsilon) \), in the interior of the energy-interval \( [0, E] \) and then to note that, by \( (10) \), this must occur at an energy, \( \epsilon \) for which

\[
\frac{d \log \sigma_S(\epsilon)}{d \epsilon} + \frac{d \log \sigma_B(E - \epsilon)}{d \epsilon} = 0
\]

which easily implies \( (31) \).

In a popular approach (cf. also \[43\]) to such problems involving the microcanonical ensemble of a pair of weakly coupled systems with such a (say unique, interior) peak, such a calculation often appears in the following guise:

One writes \( \epsilon = \epsilon_1 \) and \( E - \epsilon = \epsilon_2 \). One calls \( \log \sigma_S(\epsilon) \) “\( S_1(\epsilon_1) \)”, and one calls \( \log \sigma_B(E - \epsilon) \) “\( S_2(\epsilon_2) \)”. Then one writes the equations

\[
\epsilon_1 + \epsilon_2 = E,
\]

\[
\frac{\partial S_1}{\partial \epsilon_1} - \frac{\partial S_2}{\epsilon_2} = 0,
\]

(equivalent to \( (71) \) and

\[
\frac{\partial^2 S_1}{\partial \epsilon_1^2} + \frac{\partial^2 S_2}{\partial \epsilon_2^2} < 0
\]

(expressing the fact that it is a peak and not a trough).

It is often then assumed, or, at least, tacitly implied, that \( S(\epsilon_1) \) and \( S(\epsilon_2) \) are the “entropies” of Systems 1 and 2 (our systems S and ‘energy bath’ B) and that \( \partial S_1/\partial \epsilon_1 \) and \( \partial S_2/\partial \epsilon_2 \) are the “temperatures” of Systems 1 and 2. Finally, the equation \( (72) \) is interpreted as telling us that Systems 1 and 2 are in “stable equilibrium”.

Concerning this popular approach, we would remark and emphasize:

(a) \( S(\epsilon_1) \) and \( S(\epsilon_2) \) (our \( \log \sigma_S(\epsilon) \) and \( \log \sigma_B(E - \epsilon) \)) are not entropies (they are logarithms of densities of states). To make sense of the logarithms one would, at least, need to multiply \( \sigma_S(\epsilon) \) and \( \sigma_B(E - \epsilon) \) by “constants” with the dimensions of energy, first, to make the overall arguments of the logarithms dimensionless. This may not matter if one wishes to talk meaningfully about the logarithms themselves (evaluated at the peak values of \( \epsilon \) and \( E - \epsilon \) as ‘entropies’. One could, of course, insert, in each logarithm, an arbitrary constant with the dimensions of energy, and try to argue that it doesn’t make much difference, in practice, what is the precise value of this constant provided it is of a “reasonable” order of magnitude. But, even if it were the case that all that was at stake was such a “constant”, one would expect, in a fundamental understanding of the origin of entropy, its value to be determined in terms of the physical parameters of the problem. In fact, as we shall see below, what actually needs to be inserted is not a constant, but rather (for given system and energy-bath densities of states) a quantity (which we call \( Q \) below) with the dimensions of energy which (like the peak values of \( \epsilon \) and \( E - \epsilon \) themselves) depends on the totem energy, \( E \).

(b) It is true that one can think of \( \partial S_1/\partial \epsilon_1 \) and \( \partial S_2/\partial \epsilon_2 \) as ‘energy-dependent temperatures’ in the sense (cf. Section I A and Endnote II) that, if System 1 had energy \( \epsilon_1 \) and were uncoupled to System 2, but, rather, coupled to another and much smaller system, then that smaller system would likely get itself into a thermal state at the temperature \( \partial S_1/\partial \epsilon_1 \) evaluated at \( \epsilon_1 \) (and similarly for System 2). However, in the ‘equilibrium’ in question, where System 1 and System 2 are coupled to one-another and neither can be regarded as ‘small’, neither System 1 nor System 2 is in a thermal state (as we have shown in Section I I for our power-law case)

(c) Finally, this ‘popular’ point of view is only of value in cases where the energy probability density (say of System 1) has a peak. Whereas, we emphasize, as explained in this paper, one still predicts definite energy probability density functions when System 1 and System 2 have densities of states (such as our equal exponential and equal quadratic exponential cases discussed in Sections III and VI) which do not lead to an interior peak. (In the equal exponential case, we find an energy probability density which is flat, and, in the quadratic exponential case, it is concave with peaks at the extremities of the range \([0, E]\) which are not ‘maxima’ in the sense of the above equations for \( S_1 \) and \( S_2 \).) In such latter cases, the significant quantity of interest is not the location of a peak (there may even be no peak) but rather the full energy probability density function itself.
We turn next to the value of the entropy of the system in cases where, for given system and energy-bath densities of states, and given total energy, \( E \), the energy probability density, \( P_S(\epsilon) \), has a single peak, say at \( \epsilon = \epsilon_0 \). We are, of course, interested in both of the entropies, \( S_S^{\text{microc}} \) and \( S_S^{\text{modapprox}} \). As we anticipated in Point (a) above, when the totem is in the microcanonical ensemble with energy in our interval \( [E, E + \Delta] \), \( S_S^{\text{microc}} \) will be log(\( Q \sigma_S(\epsilon_0) \)) for a suitable quantity, \( Q \), with the dimensions of energy determined by the parameters of the problem, although we emphasize again that \( Q \) is not a “constant”; rather (for given system and energy-bath densities of states) it is a certain function of totem energy, \( E \), which has the dimensions of energy and which is determined by the detailed shape of the peak in \( P_S(\epsilon) \). In fact, by the general formula (44) (one can see that the issues mentioned in Endnote [14] will not be relevant for a sharp peak which is well inside the interior of the interval \([0, E]\)) we will have

\[
S_S^{\text{microc}} = \int_0^E P_S(\epsilon) \log(L \sigma_S(\epsilon))d\epsilon - \int_0^E P_S(\epsilon) \log(LP_S(\epsilon))d\epsilon
\]

where we have temporarily introduced an arbitrary non-zero constant, \( L \), with the dimensions of energy, which will of course cancel out in the final result.

Since \( P_S(\epsilon) \) is, by assumption, sharply peaked at \( \epsilon = \epsilon_0 \), and assuming \( \sigma(\epsilon) \) is relatively slowly varying (as will be true in typical examples such as the power-law density of states case treated below) the value of the first integral will be very well approximated by log(\( L \sigma_S(\epsilon_0) \)). The second integral will obviously take the form log(\( L/Q \)) where \( Q \) is a quantity with the dimensions of energy which can in principle be computed in terms of our system and energy-bath densities of states and the value of \( E \). So we will have

\[
S_S^{\text{microc}} = \log(Q \sigma_S(\epsilon_0)). \tag{73}
\]

On the other hand, if we consider the totem to be in a pure state, randomly chosen amongst all states with energy in the range \([E, E + \Delta]\), then, by (45), we expect the entropy to most probably be close to \( S_S^{\text{modapprox}} \) given by

\[
S_S^{\text{modapprox}} = \int_0^{E_c} P_S(\epsilon) \log(L \sigma_S(\epsilon))d\epsilon + \int_{E_c}^E P_S(\epsilon) \log(L \sigma_B(E - \epsilon))d\epsilon - \int_0^E P_S(\epsilon) \log(LP_S(\epsilon))d\epsilon
\]

which, by a similar argument, and in view of the definition of \( E_c \) (see after Equation (15)), will be very well approximated by

\[
S_S^{\text{modapprox}} = \min(\log(Q \sigma_S(\epsilon_0)), \log(Q \sigma_B(E - \epsilon_0))
\]

for the same value of \( Q \).

We next illustrate the computation of \( Q \) in the case where both of system, \( S \), and energy bath, \( B \), have the power law densities of states [18] which we discussed in Section II. We have

\[
\log(L/Q) = \int_0^E P_S(\epsilon) \log(LP_S(\epsilon))d\epsilon
\]

where \( P_S(\epsilon) \) is given by (30) with \( \gamma \) as in (32). As long as \( N_S \) and \( N_B \) are comparable in size, the location, \( \epsilon_0 \), of the peak, \( \epsilon_0 \), will be far from the extremities of the interval \([0, E]\) and we may clearly replace the limits of the above integral by \(-\infty \) and \( \infty \) with very little error.

So

\[
\log(L/Q) \simeq \frac{1}{E} \sqrt{\frac{\pi}{\gamma}} \int_{-\infty}^{\infty} \log \left( \frac{L}{E} \sqrt{\frac{\pi}{\gamma}} \exp \left( -\gamma \frac{(\epsilon - \epsilon_0)^2}{E^2} \right) \right) \exp \left( -\gamma \frac{(\epsilon - \epsilon_0)^2}{E^2} \right) d\epsilon
\]

\[
= -\log \left( \frac{L}{E} \sqrt{\frac{\gamma}{\pi}} \right) + \frac{\gamma}{E^3} \sqrt{\frac{\gamma}{\pi}} \int_{-\infty}^{\infty} (\epsilon - \epsilon_0)^2 \exp \left( -\gamma \frac{(\epsilon - \epsilon_0)^2}{E^2} \right) d\epsilon.
\]

The second term above may easily be calculated by writing it as

\[-(\gamma^{3/2}/E^{1/2})\partial/\partial \gamma \int_{-\infty}^{\infty} \exp \left( -\gamma \frac{(\epsilon - \epsilon_0)^2}{E^2} \right) d\epsilon. \]

This is equal to \((\gamma^{3/2}/E^{1/2})\partial/\partial \gamma (\gamma^{1/2})^{-1/2} = 1/2\). So

\[
\log(L/Q) \simeq -\log \left( \frac{L}{E} \sqrt{\frac{\gamma}{\pi}} \right)
\]
or, equivalently,

$$Q = \sqrt{\frac{e\pi}{\gamma}} E.$$  \hspace{1cm} (74)

For the purposes of the comparison we make in Section \[III\], we also calculate $Q$ for a thermal state at some given inverse temperature, $\beta$, of a system, $S$, with a power law density of states $\sigma(\epsilon) = A \epsilon^N$. By \[27\] we will now have

$$\log(L/Q) = \frac{\beta^{N+1}}{N!} \int_0^\infty \log \left( \frac{\beta^{N+1}}{N!} \epsilon^{\beta N} e^{-\beta \epsilon} \right) e^{\beta \epsilon} d\epsilon$$

$$= -\log \left( \frac{\beta}{N!} \right) - \frac{\beta^{N+1}}{N!} \int_0^\infty (N \log(\beta \epsilon) - \beta \epsilon) e^{\beta \epsilon} d\epsilon.$$ We may do the integral here by noticing that $\int_0^\infty x^n \log x e^{-bx} dx = d/d\alpha \int_0^\infty x^\alpha e^{-bx} dx|_{\alpha=n} = d/d\alpha (b^{-(\alpha+1)} \Gamma(\alpha + 1))|_{\alpha=n} = -(\log b) b^{-(n+1)} n! + b^{-(n+1)} \Gamma(n+1) \psi(n+1)$ where $\Gamma$ denotes the gamma function and $\psi(n+1)$ the digamma function (see e.g. \[30\]) of $n+1$ which (see again \[30\]) is equal to $\sum_{k=1}^{\infty} 1/k - \gamma$ where $\gamma$ is Euler’s constant ($= 0.5772 \ldots$). Using this, we conclude that $\log(L/Q) = -\log(\beta) + \log N! - N \psi(N + 1) + N + 1$

which, using Stirling’s approximation (which tells us that $\log n! = (n+1/2) \log n - n + 1/2 \log(2\pi) + 1/12n + O(1/n^2)$) and the asymptotic expansion of $\psi(n+1) = (\log n + 1/(2n) + O(1/n^2))$ is equal to

$$-\log(\beta) + \frac{1}{2} \log N + \frac{1}{2} \log(2\pi) + \frac{1}{2} + O(1/N)$$ \hspace{1cm} (75)

from which we conclude that

$$Q \simeq \sqrt{\frac{2e\pi N}{\beta}}.$$ \hspace{1cm} (76)

We note that if we identify $N$ here with $N_S$ and if $N_B \gg N_S$, then, by \[32\], $\gamma \simeq N_B^2/2N_S$. If, additionally, we take $\beta$ in \[76\] to be $d\log \sigma_B(\epsilon)/d\epsilon|_{\epsilon=E}$ which (with $\sigma_B(\epsilon) = A \epsilon^{N_B}$) equals $N_B/E$, then the $Q$ of \[74\] and \[76\] both take the same value $\sqrt{2e\pi N_S E/N_B}$. This agreement is to be expected since, as we discussed in Sections \[I\] and \[II\], in this situation, where $S$ is small and in the case of a total microcanonical state, the reduced density operator of $S$ will be close to that of a thermal state at inverse temperature $d\log \sigma_B(\epsilon)/d\epsilon|_{\epsilon=E}$. So the agreement of the two $Q$ in this regime serves as a check on the correctness of our two formulae \[74\] and \[76\].

However, in Section \[III\] we were interested in comparing the properties of the (as we show there) non-thermal reduced state of $S$ when $N_S$ and $N_B$ are of comparable size and the totem is in a microcanonical state with the properties of a thermal state of $S$ with the same expected energy. Treating, for simplicity, the case where $N_S = N_B = N$, say, the relevant inverse temperature, $\beta$, is $(2(N+1))/E = (2N/E)$ for large $N$) and $\gamma \simeq 4N$. We then find that the ‘thermal’ $Q$ (Equation \[76\]) becomes $\sqrt{e\pi/2NE}$ whereas the ‘microcanonical’ $Q$ (Equation \[74\]) becomes $(1/2)\sqrt{e\pi/N}$ in $E$. Thus the entropy of the comparison thermal state of $S$ is bigger than the entropy, $S_{\text{micro}}$, of the reduced state of $S$ by $\log 2/2$. While this is a small difference it is conceptually significant that it is not zero.

### B. On the entropy of the totem and more about $\Delta$

Our general framework involves a system, $S$, and an energy bath, $B$, comprising a totem. A natural question is: What is the relationship between the entropy of $S$, the entropy of $B$, and the entropy of the totem? The answer to this question depends, first of all, on whether we are contemplating the, traditional, microcanonical, scenario, or the modern scenario in which the state of the totem is pure – albeit chosen at random amongst the set of states in our energy range. In the latter, modern, scenario, there is only one entropy: As we mentioned in Sections \[LE\] and \[LV\], the (von Neumann) entropy of $S$ is equal to the (von Neumann) entropy of $B$ and both are equal to the {system}–{energy bath} entanglement entropy of the totem; the von Neumann entropy of the totem is of course zero.

In the microcanonical scenario, one might, naively, expect the entropy of the totem to be the sum of the entropy of $S$ and the entropy of $B$. But, as we shall see, this is not true. One way to see that it cannot be true is to notice (see again below) that the entropy of the totem (which will obviously have the value $\log M$, $M$ as in \[3\] and \[7\])
we shall call this $S_{\text{totem}}^\text{microc}$ – depends on the width, $\Delta$, of our energy band $[E, E + \Delta]$ whereas, as we saw in Section IV (for a suitable range of $\Delta$) the entropies of $S$ and $B$ do not. What is of course always true (and applies to both modern and microcanonical scenarios) is the property of subadditivity [31] which guarantees that the the entropy of the totem, must be less than or equal to the sum of the entropies of $S$ and $B$.

Actually, it turns out, in all three cases we have studied here (i.e. with system and energy bath densities of states of power-law form and [equal] exponential or quadratic exponential form) that the sum of the entropies of $S$ and $B$ is close to the entropy of the totem. We have not attempted to formulate any general precise statement of what we mean by this, nor have we attempted to offer any general explanation as to why this should be the case but content ourselves simply with making the content of this statement manifest for each of our three density-of-states models:

For our (equal) exponential densities of states [34], we notice that, by [35]

$$S_{\text{totem}}^\text{microc} = \log M = bE + \log(c^2E\Delta)$$

whereas, by [54]

$$S_S^\text{microc} + S_B^\text{microc} = bE + \log(c^2E^2).$$

For our (equal) quadratic exponential densities of states [58], by [62]

$$S_{\text{totem}}^\text{microc} = \log M = qE^2 + \log\left(\frac{K^2\Delta}{2E}\right)$$

whereas (see [69] and the paragraph after [70])

$$S_S^\text{microc} + S_B^\text{microc} = qE^2 + O(1).$$

For our power-law densities of states [18], we first obtain a good approximate formula for $M$ by noting that the value of the integral in [19] is equal to the ratio of the maximum value of its integrand, $\epsilon_0^{N_S}(E - \epsilon_0)^{N_B}$ ($\epsilon_0$ as in [31]) to its maximum value when normalized, which, by our Gaussian approximation, [30], is well-approximated by $(1/E)\sqrt{\gamma/E}$, $\gamma$ as in [32]. Thus we have (to a very good approximation)

$$S_{\text{totem}}^\text{microc} = \log M = \log\left(\frac{A_S A_B E \Delta \sqrt{\frac{\pi}{\gamma}}}{\gamma_0^{N_S}(E - \epsilon_0)^{N_B}}\right)$$

whereas, by [73] for $S$ and its obvious counterpart for $B$ and [74],

$$S_S^\text{microc} + S_B^\text{microc} = \log\left(\frac{A_S A_B E^2 \epsilon \pi}{\gamma_0^{N_S}(E - \epsilon_0)^{N_B}}\right).$$

We see that subadditivity in the exponential case entails $\Delta < E$. In the quadratic exponential case, (and neglecting the $O(1)$ term) it entails $\Delta < 2E/K^2$, and in the power-law case, it entails $\Delta < \epsilon\sqrt[4]{\pi/\gamma}E$. When $N_S = N_B = N$, say, $\gamma = 4N$ and this latter inequality amounts to $\Delta < (\epsilon\sqrt[4]{\pi}/2)(E/\sqrt{N})$. The first of these inequalities ($\Delta < E$) is obviously consistent with almost any sort of smallness assumption on $\Delta$. The other two inequalities indicate a need to be more precise than we were, in our rather sketchy remarks in Section I and in our subsequent derivations, about what is the appropriate range of $\Delta$ for any given pair of densities of states, $\sigma_S$ and $\sigma_B$, in order for our arguments and approximations to be valid. We shall, however, not pursue this issue further in the present paper except to deduce that the above inequalities must be necessary conditions on the value of $\Delta$.

VIII. MORE ABOUT THERMALLITY: PURIFICATION

Throughout the preceding sections we have assumed (see the paragraph after Equation [2]) that both our system, $S$, and our energy bath, $B$, have densities of states which are positively supported (i.e. the Hamiltonians $H_S$ and $H_B$ are positive) and monotonically increasing and we have been concerned exclusively with totem states which are (close to) stationary states for a totem Hamiltonian, [2], which weakly couples $S$ and $B$. In this short section, we point out that the prospects for the thermality of either $S$ or $B$ become much less constrained if we relax some of these assumptions. In particular, and in the spirit of the ‘modern’ approach, given any system density of states, $\sigma_S(\epsilon)$, whatsoever (provided only it grows sufficiently slowly for the desired thermal state to exist) one can always find an energy bath density of states and a pure totem state such that the reduced state of the system is an exactly thermal state at any given temperature.
In fact, there is a well-known procedure, in the spirit of the modern approach, known as ‘purification’ (see e.g.\textsuperscript{[32]} and references therein; see also the papers on ‘thermofield dynamics’\textsuperscript{[34]} and\textsuperscript{[35]} which are based on the same idea – we note that the term ‘purification’ seems to be due to Powers and Størmer\textsuperscript{[33]} by which a system, $S$, with any density of states whatsoever (but we shall assume it to be positively supported and monotonically increasing) and in any non-pure state one wishes to prescribe (but we are interested in a thermal state at some inverse temperature $\beta$) may be provided with a notional energy bath, $B$, such that there is a choice of pure state on the resulting totem for which the reduced density operator on $S$ is equal to the prescribed state.

The essential idea of purification is based on the fact that any density operator, $\rho$, on a Hilbert space, $\mathcal{H}$, takes the form

$$\rho = \sum_i \rho_i |\psi_i\rangle \langle \psi_i|,$$

the $\rho_i$ being positive numbers which sum to 1 and $\psi_1, \psi_2, \ldots$ being an orthonormal basis for $\mathcal{H}$, and on the observation that this can be viewed as arising as the partial trace over the second copy of $\mathcal{H}$, in the tensor product $\mathcal{H} \otimes \mathcal{H}$, of the pure density operator, $|\Psi\rangle \langle \Psi|$, where

$$\Psi = \sum_i \rho_i^{1/2} \psi_i \otimes \psi_i.$$

The easiest way to see this is to notice that, for any linear operator, $A$, on $\mathcal{H}$,

$$\langle \Psi|(A \otimes I)\Psi|\Psi\rangle_{\mathcal{H} \otimes \mathcal{H}} = \text{tr}(\rho A)_{\mathcal{H}}.$$

If we now specialize to the thermal situation where the $\psi_i$ are the energy eigenstates of a Hamiltonian, $H$, on $\mathcal{H}$ with energy eigenvalues say $e_i$ and $\rho_i = Z^{-1} e^{-\beta e_i}$, and identify $\mathcal{H}$ both with the system Hilbert space, $\mathcal{H}_S$, and also with the Hilbert space, $\mathcal{H}_B$, of our notional energy bath, then, if we also take the energy bath Hamiltonian to equal $H$ (and therefore the energy levels of the energy bath to be the same as the energy levels of the system) then $\Psi$ provides us with a pure totem state with the property that the reduced state of the system (and also the reduced state of the energy bath) is exactly thermal. With the notation of Section I A, we would write $\mathcal{H}_B = \mathcal{H}_S$ and take $\Psi$ on $\mathcal{H}_S \otimes \mathcal{H}_B$ to be given by

$$\Psi = Z^{-1/2} \sum_{\epsilon = \Delta}^{\infty} e^{-\beta \epsilon/2} \sum_{i=1}^{nS(\epsilon)} |\epsilon, i\rangle \otimes |\epsilon, i\rangle.$$  \hspace{1cm} (77)

Then the partial trace of $|\Psi\rangle \langle \Psi|$ over $\mathcal{H}_S$ will equal the $\rho_{S,\beta}^{\text{Gibbs}}$ of Equation (13).

This achievement of exact thermality contrasts with the situation discussed in Section I D (see in and after the paragraph containing Equation (17), where (on the ‘modern approach’) the totem state is assumed to be randomly chosen from amongst totem states in a narrow range of energies for a weakly coupled total Hamiltonian. As we saw there and in the rest of Part 1, with that assumption, and for systems of comparable size, thermality can never be achieved exactly and can only be approximately achieved for certain special densities of states – such as, in particular, the exponential and quadratic exponential cases discussed in Sections III and VII. However, in the purification mechanism described here, the state, $\Psi$, of the totem is – say if we regard the totem Hamiltonian to be given by Equation (2) with $H_B = H_S$ (and, say, with no coupling term at all) – clearly not even close to an energy eigenstate; in other words, the totem is in a highly non-stationary state. We remark that this purification mechanism does not seem to play much of a role in everyday physics as a mechanism by which a system can get to be hot, although, interestingly, essentially this mechanism has been made use of in the laboratory\textsuperscript{[36]} to produce thermal states of photons. (See also the remark about the Unruh effect in the next paragraph and the remarks about quantum black holes in Section IV)

We further remark that there is an alternative reinterpretation of Equation (77) in which one ascribes to the ‘energy bath’, $B$, the Hamiltonian $H_B = -H_S$ (and substitutes these Hamiltonians into the totem Hamiltonain formula (2)). With this interpretation, the state of the totem is an eigenstate of totem energy (with totem energy eigenvalue zero!) and so a stationary state. But now the density of states of the energy bath is negatively supported! This latter interpretation can be said to be realized in the Unruh effect (see e.g.\textsuperscript{[37]} and reference therein) whereby the vacuum state of a relativistic quantum field theory in Minkowski space, restricted to a right-Rindler wedge, is a thermal state with respect to Lorentz boosts; the left-Rindler wedge plays the role of our energy bath, $B$, and this can be thought of as a copy of the right-wedge quantum system but with a Hamiltonian which is the negative of the right-wedge Rindler Hamiltonian.
IX. IMPLICATIONS FOR QUANTUM BLACK HOLES

A. The (problematic) connection between our results on quadratic exponential densities of states in Section VI and black hole thermodynamics

There is a striking, at least superficial, resemblance between Equations (66) and (69) in Section VI – i.e. \( \beta = 2qE \) and \( S_{B}^{\text{micro}} = qE^2/2 \) (up to an O(1) correction) – for the inverse temperature and the traditional microcanonical entropy of a weakly coupled system, \( S \), and energy bath, \( B \), with equal quadratic exponential densities of states (58) and the equations (see Equation (B)) \( \beta = 8\pi GM \) and \( S = 4\pi GM^2 \) for the Hawking inverse temperature and entropy of a Schwarzschild black hole. In fact, if we identify, say (see below), the mean energy of \( B \) – i.e. (see (63)) half the total mean energy, \( E/2 \) – with the black hole mass, \( M \), and identify \( q \) with \( 2\pi \), the entropy, \( S_{B}^{\text{micro}} \) matches the Hawking entropy and the (inverse) temperatures match too.

This might seem to suggest that, if one identifies the system, \( S \), with ‘matter’ and the energy bath, \( B \), with ‘gravity’, then the traditional microcanonical strand of Section VI may provide a good model for a black hole in contact with its thermal atmosphere in a box and a good explanation for the microscopic origin of the entropy of this system. (Of course, we must bear in mind that, in this model, the state is only thermal in the approximate sense explained in Section VI). And on the other hand, our result, (70), that, with the densities of states (58), the ‘modern’ totem pure state entropy, \( S_s \) vanishes (up to a term of order 1 in \( E \)) might seem to be at odds with our matter-gravity entanglement hypothesis described in [17–19] and in Section IX – which entails that the total matter-gravity state of a black hole is a pure state. However, we need to realize that the results of Section VI assume that the dynamics of the totem is governed by a totem Hamiltonian of the schematic form (2) with both \( S \) and \( B \) (now to be interpreted as ‘matter’ and ‘gravity’) Hamiltonians positive and weakly coupled. Yet, notoriously, it seems unlikely to be possible to have a quantum theory of gravity within the scope of these basic assumptions (albeit these assumptions do seem to apply to the weak-string coupling limit if the fundamental degrees of freedom are taken to be those of a string rather than of the gravitational field itself – see Section IX). Already classical general relativity is nonlinear and, unlike the situation for (2), energy (mass) is not additive. So the fact that the mean energies of system and energy bath (modelling matter and gravity) are equal in our model seems strange. (It is also unclear whether we should identify the entropy of the black hole with \( S_{B}^{\text{micro}} \) as we did above, or with \( S_{B}^{\text{micro}} + S_{B}^{\text{modapprox}} \), which is twice as big, or with \( \log M \) – see also Section VII B). Furthermore (see the remarks after Equation (63)) in this model, the mean energy, \( E/2 \), of each of matter and of gravity is anyway just the mean of an energy probability density (the \( P_{E}^{\text{micro}}(\epsilon) \) of (68) and of Figure 1) which is peaked at the extremes, \( \epsilon = 0 \) and \( \epsilon = E \), while of course (cf. after Equation (63)) the probability densities of matter and gravity are perfectly anticorrelated. So the model predicts large statistical fluctuations, with (to the extent that it makes sense to talk about the energy of subsystems in general relativity) probability distributions for matter and gravity being such that, approximately, with probability one half, gravity has all the energy and matter none, and with probability one half, matter has all the energy and gravity none. The latter case (where there is presumably no black hole) is then a particular problem because (see the next paragraph) the quadratic exponential form of the density of states presupposed in the model for matter becomes invalid when a black hole is not present.

In fact, turning to more specifically quantum aspects, aside from all the usual problems of quantum gravity (non-renormalizability etc.) it would be incorrect to assume that one can ascribe a single density of states to each of gravity and matter throughout changes of state which include the formation of black holes. Rather it would seem that one has to assign, in some sense, a ‘state-dependent density of states’ to matter: in the absence of black holes, the densities of states of common forms of matter (including photons) grow much more slowly than quadratic exponentials, while, plausibly, when a black hole is present, they do grow as quadratic exponentials (with subleading corrections). (Evidence for this latter statement is provided by the ‘brick wall’ approach which suggests that the energy probability densities of matter and gravity are perfectly anticorrelated. So the model predicts large statistical fluctuations, with (to the extent that it makes sense to talk about the energy of subsystems in general relativity) probability distributions for matter and gravity being such that, approximately, with probability one half, gravity has all the energy and matter none, and with probability one half, matter has all the energy and gravity none. The latter case (where there is presumably no black hole) is then a particular problem because (see the next paragraph) the quadratic exponential form of the density of states presupposed in the model for matter becomes invalid when a black hole is not present.

In fact, turning to more specifically quantum aspects, aside from all the usual problems of quantum gravity (non-renormalizability etc.) it would be incorrect to assume that one can ascribe a single density of states to each of gravity and matter throughout changes of state which include the formation of black holes. Rather it would seem that one has to assign, in some sense, a ‘state-dependent density of states’ to matter: in the absence of black holes, the densities of states of common forms of matter (including photons) grow much more slowly than quadratic exponentials, while, plausibly, when a black hole is present, they do grow as quadratic exponentials (with subleading corrections). (Evidence for this latter statement is provided by the ‘brick wall’ approach which suggests that the energy probability densities of matter and gravity are perfectly anticorrelated. So the model predicts large statistical fluctuations, with (to the extent that it makes sense to talk about the energy of subsystems in general relativity) probability distributions for matter and gravity being such that, approximately, with probability one half, gravity has all the energy and matter none, and with probability one half, matter has all the energy and gravity none. The latter case (where there is presumably no black hole) is then a particular problem because (see the next paragraph) the quadratic exponential form of the density of states presupposed in the model for matter becomes invalid when a black hole is not present.

In fact, turning to more specifically quantum aspects, aside from all the usual problems of quantum gravity (non-renormalizability etc.) it would be incorrect to assume that one can ascribe a single density of states to each of gravity and matter throughout changes of state which include the formation of black holes. Rather it would seem that one has to assign, in some sense, a ‘state-dependent density of states’ to matter: in the absence of black holes, the densities of states of common forms of matter (including photons) grow much more slowly than quadratic exponentials, while, plausibly, when a black hole is present, they do grow as quadratic exponentials (with subleading corrections). (Evidence for this latter statement is provided by the ‘brick wall’ approach which suggests that the energy probability densities of matter and gravity are perfectly anticorrelated. So the model predicts large statistical fluctuations, with (to the extent that it makes sense to talk about the energy of subsystems in general relativity) probability distributions for matter and gravity being such that, approximately, with probability one half, gravity has all the energy and matter none, and with probability one half, matter has all the energy and gravity none. The latter case (where there is presumably no black hole) is then a particular problem because (see the next paragraph) the quadratic exponential form of the density of states presupposed in the model for matter becomes invalid when a black hole is not present.

In fact, turning to more specifically quantum aspects, aside from all the usual problems of quantum gravity (non-renormalizability etc.) it would be incorrect to assume that one can ascribe a single density of states to each of gravity and matter throughout changes of state which include the formation of black holes. Rather it would seem that one has to assign, in some sense, a ‘state-dependent density of states’ to matter: in the absence of black holes, the densities of states of common forms of matter (including photons) grow much more slowly than quadratic exponentials, while, plausibly, when a black hole is present, they do grow as quadratic exponentials (with subleading corrections). (Evidence for this latter statement is provided by the ‘brick wall’ approach which suggests that the energy probability densities of matter and gravity are perfectly anticorrelated. So the model predicts large statistical fluctuations, with (to the extent that it makes sense to talk about the energy of subsystems in general relativity) probability distributions for matter and gravity being such that, approximately, with probability one half, gravity has all the energy and matter none, and with probability one half, matter has all the energy and gravity none. The latter case (where there is presumably no black hole) is then a particular problem because (see the next paragraph) the quadratic exponential form of the density of states presupposed in the model for matter becomes invalid when a black hole is not present.

A tentative possible ‘modern’ model with the correct Hawking value for the entropy: ‘Matter’ is modelled as a
positively supported. Finally, there is the same strange feature that we raised above for our microcanonical model, energy eigenstate, at the expense of having densities of states which are not monotonically increasing and/or not lines of the alternative reinterpretation of Equation (77) in Section VIII so that the totem state is modelled as an to force a strong-coupling situation into a weak-coupling mould, or maybe the model should be modified along the totem is far from being an energy eigenstate. It could possibly be that this is the best one can do when one attempts how seriously it should be taken. There is also an apparent flaw in this tentative model in that the pure state of the

\[ qE \]

predict an inverse temperature of 2

\[ \frac{E}{2} \]

the microcanonical state (6)

\[ P \]

each have the same energy probability density (the

\[ \rho \]

and

\[ \epsilon \]

anticorrelated but, instead, perfectly correlated: One sees immediately from (78) that when gravity has energy near

\[ \epsilon = E \]

this model differs from the microcanonical model in that the two energy probability densities are now no longer antico-related but, instead, perfectly correlated: One sees immediately from (78) that when gravity has energy near 0, so will matter, and when gravity has energy near E, so will matter. So at least one of the problems we mentioned above (the one we referred to as a “particular problem”) for the microcanonical model seems to be alleviated in the above proposed ‘modern’ model. Another problem which is alleviated with this tentative modern model is that it is clear, in this modern model that the entropy should be identified with \( S_B^{\text{mod}} \) whereas, as we remarked above, in the microcanonical model it was not clear whether it should be identified with \( S_B^{\text{micr}} \) or with (approximately – see Section VII B) twice this value; in the modern model, there is only one entropy – i.e. the S-B entanglement entropy

\[ S_B^{\text{modern}} = S_S^{\text{modern}} \approx S_B^{\text{modapprox}} = S_S^{\text{modapprox}} \]

One would also wish to be able to relate our discussion, in Section VII A of cases where the probability density is sharply peaked, to the results, [44], of Hawking on his microcanonical approach to quantum black holes. These latter results of Hawking do seem to form a physically compelling and coherent picture and one would like to understand whether and, if so, how, they can be reconciled with the modern strand of results in Section VII A even though they seem, superficially, to be more easy to understand in terms of the microcanonical strand of work there and, seemingly, to be at odds with the modern strand. We hope to address this question elsewhere. Suffice it to to say here that, again, the difficulties and problems mentioned above are of at least equal relevance also to this issue and thus it seems difficult, also for this issue, to reach a fully convincing conclusion either way [44].

B. Towards a better understanding of black hole entropy in terms of string entropy

Where we have been able to make a, we think, persuasive, case for the relevance of the ‘modern’ strand of ideas in the present paper – used in combination with the (see Section I B) matter-gravity entanglement hypothesis of [17,19] – to the understanding of black hole entropy is with a model which relates the work in Sections IV and VII here, concerning densities of states which grow exponentially with energy, to an understanding of black hole entropy based on the idea that black holes are strong string-coupling limits of states of weakly coupled string theory. This application of our work to quantum black holes seems to be more well-founded because, unlike in the situations discussed above, we would expect the general assumptions we made at the outset here (positive Hamiltonians, weak coupling etc.) to be applicable to the weak-coupling regime of string theory.

In 1993, Susskind [38] proposed, and in 1997, Horowitz and Polchinski [40, 41], gave further evidence, of a semi-qualitative nature, for, the hypothesis that a (say 4-dimensional, Schwarzschild) black hole can be interpreted as the strong string-coupling limit of a certain state of string theory at weak coupling consisting of a (single) long string. These authors argued that one obtains, with this interpretation, an explanation of black hole entropy in terms of

\[ \frac{1}{\sqrt{\mathcal{M}}} \sum_{\epsilon \Delta}^{E} \sqrt{n_B(E - \epsilon)} \sum_{i=0}^{n_B(\epsilon)} |\epsilon, i\rangle_S \langle \epsilon, i|_B \]
known formulae, based on the ‘counting of states’, in string theory at low string coupling. Moreover, in related work on extremal, and near extremal, black holes (an important early paper was Strominger and Vafa [39]), full quantitative agreement was found between the results of such a string theory approach to black hole entropy (and other related quantities) and the previously established Hawking formulae. It was then claimed that this work, not only gave a microscopic explanation for black hole entropy but also, in view of the fact that string theory is a standard quantum mechanical theory with a unitary time evolution, that it resolved the Information Loss Puzzle (see Section [45]. (In fact similar remarks apply to those we make in Paragraph (a) of Section VII A.)

Actually, the Horowitz-Polchinski work is couched in terms, not of the degeneracy of a particular energy level of string theory, but rather of the (averaged out) density of states of a long string. The problem is then compounded by the fact that a density of states is not a dimensionless quantity, so it is not physically meaningful to take its logarithm [45]. (In fact similar remarks apply to those we make in Paragraph (a) of Section VII A.)

Focusing on this Horowitz-Polchinski work, we shall next propose a modified version of their scenario, based on the modern strand of Sections II and V of the present paper, which seems to overcome the above difficulties and to offer the promise of a fully satisfactory explanation of black hole entropy in terms of string theory, consistent with unitarity and consistent with a resolution to the Information Loss Puzzle – namely with the resolution to the Information Loss Puzzle we proposed in [17–19] based on our matter-gravity entanglement hypothesis (see above and Section I B). This will be further discussed in [22] and further developed in [23], whose content we indicate very briefly at the end of this subsection.

We begin by briefly recalling the basic argument of Susskind, Horowitz and Polchinski [38] [10] as expounded in [41]. Their basic hypothesis is that, as one scales the string length scale, $\ell$, up and the string coupling constant, $g$, down from their physical values, keeping Newton’s gravitational constant, $G = g^2 \ell^2$, fixed, a (4-dimensional) Schwarzschild black hole of mass $M$ will turn into a long string with roughly the same energy, $\epsilon = M$. The density of states of such a long string, in the limit of weak coupling, is known, very roughly (i.e. omitting an approximately inverse-power prefactor – see below) to take the exponential form

$$s_{\text{long string}}(\epsilon) = C_{\text{bs}} \epsilon^{\ell \epsilon} \tag{79}$$

($C_{\text{bs}}$ a constant with the dimensions of inverse energy of the same order of magnitude as $\ell$).

The gist of the argument is that the ‘logarithm’ of this is approximately given by

$$S_{\text{long string}} = \ell \epsilon \tag{80}$$

and they refer to this quantity as (approximately) the ‘entropy of the long string at energy $\epsilon$’. They then argue that this should be equated with the entropy of a (Schwarzschild) black hole provided that one does the equating (i.e. during the process of scaling $\ell$ described above) when, to within an order of magnitude or so,

$$\ell = GM \tag{81}$$

which is roughly the ‘size’ of the black-hole. (Cf. the fact that the Schwarzschild radius is $2GM$.)

Combining (80) and (81) (and replacing $\epsilon$ by $M$) they thus claim to predict that the entropy of the black hole will be within an order of magnitude or so of a constant times $GM^2$ which agrees, up to an undetermined value for the constant, with the Hawking value, $4\pi GM^2$, for the entropy of a black hole.

In our view, what one is actually entitled to say, instead, is that, the number of energy eigenstates of a black hole in a band of width $\Delta$ around energy $\epsilon$ will, by (79) be $\ell \epsilon + \log(C_{\text{bs}} \Delta)$, which, for a ‘reasonable-sized’ $\Delta$ will...
be approximately the same as \( \ell \epsilon \). The argument in the previous two paragraphs then tells us that the number of energy eigenstates of a black hole in a band of width \( \Delta \) around energy (i.e. mass) \( M \) will be within an undetermined constant, say \( C \), of the order of 1, times \( G M^2 \) (and thus, by the way, that the density of states of a Schwarzschild black hole should behave roughly as a constant times \( \exp(C G \ell^2) \)). However, in our view, it remains a challenge to explain why the logarithm of this formula for the density of states of a black hole should equal (to within an unknown constant, \( C \) of the order of 1) the Hawking formula for black hole entropy.

In our attempt to meet this challenge, we first posit that the scenario in which a black hole goes over to a single long string should be replaced by a scenario in which an equilibrium state (i.e. energy eigenstate) consisting of a black hole in contact with its matter atmosphere in a suitable box (on our view described by a pure total state – see our ‘entanglement picture of black hole equilibrium’ in Section 13 with approximate total energy, \( E \), goes over (again, as one scales the string length scale, \( \ell \), up and the string coupling constant, \( g \), down from their physical values, keeping Newton’s gravitational constant, \( G = g^2 \ell^2 \) fixed) to a (pure) equilibrium state, with a similar total energy, consisting of a single long string in contact with an atmosphere of small strings in a suitably rescaled box.

We now assume that the density of states of the long string takes (to the same rough approximation as above) the form of (79) and that the density of states of the stringy atmosphere, \( \sigma_{\text{string atmosphere}}(\epsilon) \), takes (again, to a rough approximation) the similar, exponential, form

\[
\sigma_{\text{string atmosphere}}(\epsilon) = C_{\text{sa}} \epsilon^{\ell \epsilon}.
\]

If we now regard (most of) the stringy atmosphere as corresponding to ‘matter’ and as playing the role of our ‘system’, \( S \), and the long string as corresponding to (most of) ‘gravity’ and as playing the role of our energy-bath, ‘B’, then it is plausible that these may be described by Hilbert spaces and Hamiltonians which, since we are at weak string coupling, should fall within the scope of the present paper, and in particular the dynamics should be described by a totem Hamiltonian of form (2). In view of the exponential growth of the densities of states, (79) and (82), we may therefore apply the formalism of Sections 11 and 15 (modified as explained in Endnote 29 to take into account the different prefactors, \( C_{\text{sa}} \) and \( C_b \)). In particular, the modern strand of these sections tells us that a typical pure equilibrium state of our (stringy atmosphere)-(long string) totem with energy around \( E \) will, with a high probability, have an entropy very close to that given by Equation (55) with \( b = \ell \) (with the modification to the logarithmic term given in Endnote 29). I.e. ignoring the logarithmic term, by

\[
S = \ell E/4,
\]

while the (expected) energy of the long string, \( \bar{\epsilon}_{\text{long string}} \) will (see again Endnote 29) be given by

\[
\bar{\epsilon}_{\text{long string}} = E/2
\]

(and, of course the mean energy of the stringy atmosphere will also be \( E/2 \) in this model).

In parallel to the philosophy of 38 40 41, we now assume that, when we scale \( g \) back up and \( \ell \) down, keeping \( G = g^2 \ell^2 \) constant and keeping \( \Psi \) the ‘same’, we can equate \( \epsilon_{\text{long string}} \) with the mass, \( M \), of the black hole when \( \ell = X G M \), say, where \( X \) is an adjustable parameter of the order of 1. We thereby obtain the prediction \( S = X G M^2/2 \), as the value for the entanglement entropy of black hole and thermal atmosphere in the ‘same’ (i.e. after rescaling) state \( \Psi \). But it is also plausible (as indicated above – see the relevant Endnote in 23 for further discussion) that this is approximately the same as the entanglement entropy between gravity and matter which, according to the matter-gravity entanglement hypothesis of 17 19 and Section 13, is the physical entropy of the black hole. We thus predict that the physical entropy of our black hole is (approximately) \( X G M^2/2 \). This agrees with the Hawking entropy of \( 4\pi M^2 \) if we take \( X = 8\pi \).

Furthermore, we showed, in Section 11 that both \( S \) and \( B \) will be ‘\( E \)-approximately semi-thermal’, in the sense explained there, at inverse temperature \( \beta = \ell \). Equating this with the inverse black hole temperature when \( \ell = X G M \) predicts a black hole inverse temperature of \( X G M \) which, intriguingly, agrees with the inverse Hawking temperature for the same value of \( X \) (i.e. \( 8\pi \)). We remark that we would not have correctly predicted both Hawking temperature and Hawking entropy for a single value of \( X \) had we followed the traditional microcanonical, instead of the modern strand, of Section 11 nor if we had adopted the approach of 38 40 41 and defined the inverse temperature by \( \beta = d(\log(\sigma_{\text{long string}}(\epsilon)))/d\epsilon \). (In each case, the necessary values of \( X \) for fitting the Hawking entropy and the Hawking temperature would have differed by a factor of 2.) However we caution that it is not clear whether this nice feature of our modern model with exponential densities of states survives when (see next paragraph) we improve the model to include the appropriate approximately inverse-power prefactors. We discuss this further in 23. Nevertheless, our main point is that a ‘modern’ model for black hole entropy, based on our matter-gravity entanglement hypothesis seems able to predict a temperature of the order of the Hawking temperature and an entropy of the order of the Hawking entropy.
The main deficiency in the above scenario is the adoption of the equations (79) and (82) for the approximate forms of the long-string and stringy-atmosphere densities of states. These formulae omit important (dimension-dependent) approximately inverse-power prefactors, and when one takes these into proper account, it turns out (with some, seemingly reasonable, assumptions) that the account of the origin of black hole entropy above and in [22] needs significant changes and is even, in certain respects, misleading, although one arrives at similar final conclusions. The prefactors are also needed to explain why an equilibrium weakly coupled string state in a box consists of a single long string surrounded by an atmosphere of short strings as we posited above. Also, the statistical spread in energy of the string (and hence the predicted statistical spread in energy of the black hole) around the mean energy $E/2$ will be altered with the correct prefactors. All these matters will be discussed in our second companion paper [23].

Part 2: Full explanation of the formula (15) and arguments for the validity of the proposition in Section ID

X. THE WORK OF LUBKIN AND PAGE AND OTHER PRELIMINARIES AND OUTLINE OF THE REMAINDER OF PART 2

In Section I we mentioned the work of Lubkin, [10], where it is shown that a randomly chosen pure density operator, $\rho^{mn} = |\Psi\rangle\langle\Psi|$, on the tensor-product Hilbert space, $\mathcal{H}_m \otimes \mathcal{H}_n$, of a pair of quantum systems – $\mathcal{H}_m$ being $m$-dimensional and $\mathcal{H}_n$ being $n$-dimensional – will, for fixed $m$ and $n \gg m$, have, with high probability, a reduced density operator, $\rho_m^{mn}$, on $\mathcal{H}_m$, which is close to the maximally mixed density operator – with components, in any Hilbert space basis, diag$(1/m, \ldots, 1/m)$. We first need to recall some more details about this work as well as some further related developments which will be relevant throughout Part 2.

Lubkin justified the above statement and made it precise by obtaining a result which is (easily seen to be) equivalent to the following exact formula for the mean value (i.e. over Haar measure on the set of unit vectors $\Psi$) $\langle \text{tr}((\rho_m^{mn})^2) \rangle$, of $(\rho_m^{mn})^2$: In our notation

$$\langle \text{tr}((\rho_m^{mn})^2) \rangle = \frac{m + n}{mn + 1}.$$  \hfill (83)

We shall re-derive this result of Lubkin with a somewhat different method in the next section (Section XI) since our full explanation of the formula in Equation (15) and our arguments for the validity of our proposition of Section ID will be closely based on it. Lubkin then gave a simple general argument which amounts to the statement that, for any density operator, $\rho_m$, on an $m$-dimensional Hilbert space, whenever $m/\text{tr}(\rho_m^2) - 1 \ll 1$, then the mean value, $\langle S(\rho_m) \rangle$, of the von Neumann entropy, $S(\rho_m)$, of $\rho_m$ will be well-approximated by

$$\langle S(\rho_m) \rangle \approx \log m - \frac{1}{2} \left( m/\text{tr}(\rho_m^2) - 1 \right).$$  \hfill (84)

Applying this result to $\rho_m^{mn}$, (83) implies that whenever $m \ll n$,

$$\langle S(\rho_m^{mn}) \rangle \approx \log m - \frac{m^2 - 1}{2(mn + 1)}$$  \hfill (85)

which may also be regarded as an alternative quantitative expression of the qualitative property that, when $m \ll n$, most $\rho_m^{mn}$ must be close to maximally mixed.

In (84) and (85) above, the von Neumann entropy is defined in the usual way as in Equation (16). Around 15 years later, Page [46] showed that the formula

$$\langle S(\rho_m^{mn}) \rangle \approx \log m - \frac{m}{2n}$$  \hfill (86)

is a good approximation (with error term of order $1/mn$) whenever $1 \ll m \leq n$, and noted that this agrees with (82) on their common domain of validity [47]. We note here, in passing that, combining the two estimates (83) and (86), we can clearly write, simply, that whenever $m \leq n$, $\langle S(\rho_m^{mn}) \rangle = \log m - m/2n + O(1/mn)$ (since, when $m$ and $n$ are both of order 1, the entropy can anyway only be of order 1) and hence, combining this result with $m$ and $n$
interchanged with the equality of $S(\rho_{mn}^{\alpha\beta})$ and $S(\rho_{mn}^{\alpha\beta})$ \[88\] (by which we mean the reduced density operator of $\rho_{mn}^{\alpha\beta}$ on $\mathcal{H}_n$ – see Section XI).

$$\langle S(\rho_{mn}^{\alpha\beta}) \rangle = \log(\min(m,n)) - \frac{\min(m,n)}{2\max(m,n)} + O\left(\frac{1}{mn}\right). \quad (87)$$

In the remainder of Part 2, we first introduce, in Section XI, a useful coordinatization for unit vectors in an $N$-dimensional Hilbert space in terms of which the ‘Haar’ measure of Section IV takes a particularly convenient form. In order to prepare the ground for our subsequent generalization (see below) we then use this coordinatization to obtain an alternative derivation of Lubkin’s result \[83\]. We also discuss in more detail the qualitative consequence of Lubkin’s result concerning the ‘almost maximal mixedness’ of the density operator $\rho_{mn}^{\alpha\beta}$ when $m \ll n$ (as previously pointed out by Lubkin, as mentioned in Section I) and we also point out a related important second qualitative consequence concerning the nature of $\rho_{mn}^{\alpha\beta}$ in the ‘opposite’ situation when $m \gg n$. Then, in Section XII we use a generalization of our alternative derivation of Equation \[83\] as well as a suitable generalization of our argument for its two qualitative consequences to give the full statement of Equation \[15\] including an explanation of how the $w_B(E - \epsilon)$-dimensional subspace of the $(\eta_S(\epsilon))$-dimensional energy-$\epsilon$ subspace of $\mathcal{H}_S$ spanned by the $|\epsilon, i\rangle$ depend on $\Psi$ and also to give our argument for the validity of our proposition (stated in full in Section I.D) that $\rho_{mn}^{\alpha\beta}$ is well approximated by the $\rho_{S_{\text{modern}}}$ (see the discussion after \[15\]) is well approximated by the $\rho_{S_{\text{modern}}}$ of Equation \[15\]. We end, in Section XIII, with two calculations which provide confirmatory evidence of the goodness of our approximation in situations such as those we discuss in Part I.

**XI. A USEFUL REPRESENTATION OF HAAR MEASURE AND DETAILS ON, AND FURTHER CONSEQUENCES OF, LUBKIN’S RESULT**

Let $\mathcal{H}$ be an $N$-dimensional Hilbert space and let $\{E_1, \ldots, E_N\}$ be an arbitrary orthonormal basis. Then, as usual, we coordinatize an arbitrary vector, $\psi \in \mathcal{H}$, by the $N$-tuple of complex numbers $(z_1, \ldots, z_N)$ where $\psi = \sum_{a=1}^{N} z_a E_a$. $\psi$ is, of course, then a unit vector if and only if $\sum_{a=1}^{N} |z_a|^2 = 1$. So the set of normalized vectors in our Hilbert space is coordinatized as the unit sphere in $\mathbb{C}^N$. (Writing $z_a = a_1 + i a_2$ etc. we see that this is obviously the ‘same thing’ as the real unit $(2N-1)$-sphere.) Next we change to polar coordinates in each copy of $\mathbb{C}$ by setting $z_a = r_a e^{i \theta_a}$, whereupon the usual volume element $dz_1 \ldots dz_N$ on $\mathbb{C}^N$ takes the form $r_1 \ldots r_N dr_1 \ldots dr_N d\theta_1 \ldots d\theta_N$. Changing coordinates further from $(r_1, \ldots, r_N; \theta_1, \ldots, \theta_N)$ to $(r_1, \ldots, r_{N-1}; R; \theta_1, \ldots, \theta_N)$, where

$$R^2 = \sum_{a=1}^{N} r_a^2, \quad (88)$$

this volume element is easily seen to become $r_1 \ldots r_{N-1} Rdr_1 \ldots dr_{N-1} R d\theta_1 \ldots d\theta_N$. Next we note that, in these latter coordinates, the unit sphere in $\mathbb{C}^N$ is defined by the condition $R = 1$. Thus the usual area element, $dA$, on our unit sphere is obtained by setting $R = 1$ and removing the term $dR$ from this formula. i.e.

$$dA = r_1 \ldots r_{N-1} dr_1 \ldots dr_{N-1} d\theta_1 \ldots d\theta_N.$$ 

It is now convenient to replace the coordinates $r_a$ ($a = 1, \ldots, N - 1$) by $w_a$ ($a = 1, \ldots, N - 1$) where $w_a = r_a^2$, whereupon clearly

$$dA = 2^{N-1} dw_1 \ldots dw_{N-1} d\theta_1 \ldots d\theta_N.$$ 

In view of the relation between the $w_a$ and the first $N - 1$ of the $r_a$ and the fact that the $r_a$ satisfy \[88\] with $R = 1$, the $(N-1)$-tuple $(w_1, \ldots, w_{N-1})$ clearly takes values which range over the simplex defined by the inequalities $0 \leq w_a$ for each $a$-value from $1$ to $N - 1$, together with the inequality $\sum_{a=1}^{N-1} w_a \leq 1$. (We shall call this the standard $(N-1)$-simplex.) We remark that we can think of the quantity $1 - \sum_{a=1}^{N-1} w_a$ as ‘$w_N$’ for we will then have $w_N^2 = r_N^2$. So the latter inequality can then be expressed as $0 \leq w_N$. The $\theta_a$ values each, of course, range over $[0, 2\pi)$. So the area of our unit sphere is the integral of $dA$ over the above ranges for our variables which is $2^{N-2}(2\pi)^{N-1}$ times the volume of our simplex. But the latter is easily seen to be $1/(N - 1)!$. So the area of our unit sphere is $2\pi^{N}/(N - 1)!$ which, of course, is the well-known value for the surface area of the real $(2N-1)$-sphere. We want our Haar measure to be a probability measure, so we need to normalize it by dividing by this surface area. In conclusion, (up to an irrelevant set of measure zero) we have coordinatized the set of normalized vectors in our $N$-dimensional Hilbert space by products of $(N-1)$-tuples $(w_1, \ldots, w_{N-1})$ whose values range over our standard $(N-1)$-simplex, with $N$-tuples $(\theta_1, \ldots, \theta_N)$ whose values
Let $\Psi$ be an arbitrary unit vector in $\langle(83) \text{ in Section } X\rangle$ over Haar measure (with the result (97) below) and also to (re-) derive Lubkin’s formula (83) for $w$ remarked above, with an obvious corresponding notation.

(The reason for the ‘check’ ‘ˇ’ is that we will also want, below, to talk about the \(\hat{\Psi}\) defined as
\[
J = \int d\theta_1 \ldots d\theta_N.
\]
and
\[
d\text{Torus} = (2\pi)^{-N} d\theta_1 \ldots d\theta_N.
\]

We next define, and record, the easy-to-check values of, certain integrals of certain products of $w$’s over our simplex w.r.t. $d\text{Simplex}$:

\[
J_1 := \int w_1 d\text{Simplex} = 1/N,
\]
\[
J_{11} := \int w_1^2 d\text{Simplex} = 2/N(N + 1),
\]
\[
J_{12} := \int w_1 w_2 d\text{Simplex} = 1/N(N + 1).
\]

Obviously we assume here, for $J_1$ and $J_{11}$, that $N$ is at least 2, and for $J_{12}$, that $N$ is at least 3. We note that $J_p$ (defined as $J_1$ but with $w_1$ replaced by $w_p$) will equal $J_1$ for any other value of $p$ between 1 and $N$. Similarly (and with an obvious corresponding notation) $J_{pp} = J_1$ for any other $p$ between 1 and $N$, and $J_{pp} = J_{12}$ for any pair of different $q$ and $p$ between 1 and $N$. We reiterate that all this holds even if $q$ or $p$ is equal to $N$, in which case, as we remarked above, $w_N$ is taken to mean $1 - w_1 - \cdots - w_{N-1}$.

We now use this coordinatization to compute the average, $\langle\rho_m^{mn}\rangle$, of $\rho_m^{mn}$ (see the paragraph before Equation (83) in Section X) over Haar measure (with the result (97) below) and also to (re-)derive Lubkin’s formula (83) for $\langle\text{tr}(\rho_m^{mn})^2\rangle$. (Here, as in Section X, we indicate averages with respect to Haar measure with angle-brackets $\langle \cdot \rangle$.) Let $\Psi$ be an arbitrary unit vector in $\mathcal{H}_m \otimes \mathcal{H}_n$ and choose (arbitrary) bases, $\{e_1, \ldots, e_m\}$ for $\mathcal{H}_m$ and $\{f_1, \ldots, f_n\}$ for $\mathcal{H}_n$. Then we may write

\[
\Psi = c_{ak} e_a \otimes f_k \quad \text{(summed over } a \text{ and } k),
\]

where $c_{ak} e_a^k$ [summed over $a$ and $k$] = 1) and the reduced density operator (see Sections I and X) $\hat{\rho}_m^{mn}$ on $\mathcal{H}_m$ takes the form $(\hat{\rho}_m^{mn})_{a\bar{a}} |e_a\rangle \langle e_{\bar{a}}|$ (summed over $a$ and $\bar{a}$ from 1 to $m$) where

\[
(\hat{\rho}_m^{mn})_{a\bar{a}} = c_{ak} e_a^k \quad \text{(summed over } k).
\]

(The reason for the ‘check’ ‘ˇ’ is that we will also want, below, to talk about the \((m \times m)\) matrix whose components are $(\hat{\rho}_m^{mn})_{a\bar{a}}$. And we call this ‘\(\hat{\rho}_m^{mn}\)’ to distinguish it from the operator $\rho_m^{mn}$ on $\mathcal{H}_m$.) We want to average this over the unit sphere in $\mathbb{C}^N$ for $N = mn$ where each factor of $\mathbb{C}$ accommodates one of the $N = mn$ components of $c_{ak}$. So we replace $c_{ak}$ in (96) by $r_{ak} e^{i\theta_{ak}}$ and then by $w_{ak}^{1/2} e^{i\theta_{ak}}$ and similarly for $c_{\bar{a}k}$, obtaining

\[
(\hat{\rho}_m^{mn})_{a\bar{a}} = w_{ak}^{1/2} w_{\bar{a}k}^{1/2} e^{i(\theta_{ak} - \theta_{\bar{a}k})} \quad \text{(summed over } k)
\]
and we integrate this over Haar measure (89). Integrating over the $\theta$s first (with $d\text{Torus}$ (91)) will obviously give a factor of $\delta_{a\bar{a}}$ for each $k$ in the sum (from 1 to $n$) over $k$. We are thus left with a sum (over $k$) of $n$ integrals,

\[
\int w_{ak} d\text{Simplex}
\]
for each $a$, each of which takes the form of $J_1$ (92) for $N = mn$. So we conclude that

\[
\langle(\hat{\rho}_m^{mn})_{a\bar{a}}\rangle = \frac{n\delta_{a\bar{a}}}{mn} = \frac{1}{m} \delta_{a\bar{a}}
\]
and hence, obviously,
\[ \langle \rho_m^{mn} \rangle = \frac{1}{m} I_m \] (97)
where \( I_m \) denotes the identity operator on \( \mathcal{H}_m \). Similarly, denoting, by \( \rho_n^{mn} \) the reduced density operator on \( \mathcal{H}_n \) we will have
\[ \langle \rho_n^{mn} \rangle = \frac{1}{n} I_n. \] (98)

We remark that we don’t strictly need this result for the qualitative consequences of Lubkin’s results we discuss below, but it is anyway interesting and also serves as a useful preliminary to the recalculation of Lubkin’s result to which we will turn next – see especially the remark after equation (106). More importantly, we will need the counterpart to this result in our argument, below, for the closeness of \( \rho_S^{\text{modern}} \) and \( \rho_S^{\text{modapprox}} \).

Proceeding similarly, it is straightforward to see that
\[ \text{tr}((\rho_m^{mn})^2) = c_{ak}c_{ak}^*c_{al}c_{al}^* \quad \text{(summed over k, l, a and} \ \hat{a}) \]

\[ = w_{ak}^{1/2}w_{ak}^{1/2}w_{al}^{1/2}w_{al}^{1/2}e^{i(\theta_{ak}-\theta_{ak})}e^{-i(\theta_{a}l-\theta_{a}l)} \quad \text{(summed over k, l, a and} \ \hat{a}). \]

We integrate this over \( d\text{Haar} \), again doing the \( \theta \)-integrals first. Clearly the latter will vanish unless either \( a = \hat{a} \) or \( k = l \) (or both) whereupon, for fixed values of \( a, \hat{a}, k \) and \( l \), the complex exponential will integrate (with \( d\text{Torus} \)) to 1. Moreover, (i) If \( a = \hat{a} \) and \( k = l \), then the \( w \)-integral over the simplex will equal \( J_{11} \) for \( N = mn \) – and there are \( mn \) such cases; (ii) If \( a \neq \hat{a} \) and \( k = l \), then the \( w \)-integral over the simplex will equal \( J_{12} \) for \( N = mn \) – and there are \( mn(m-1) \) such cases; and finally (iii) If \( a = \hat{a} \) and \( k \neq l \), then the \( w \)-integral over the simplex will equal \( J_{12} \) for \( N = mn \) again, and there are \( (n-1)m \) such cases. Thus we conclude that
\[ \langle \text{tr}((\rho_m^{mn})^2) \rangle = \frac{2mn + mn(m-1) + n(n-1)m}{mn(mn+1)} = \frac{m+n}{mn+1} \] (99)
in agreement with (93).

Lubkin’s result (93) is important for us because of two qualitative consequences: First, as we mentioned in Section X and as Lubkin himself essentially argued, if \( n \gg m \) then \( \langle \text{tr}((\rho_m^{mn})^2) \rangle \) will be close to \( 1/m \). The only way this can happen is if most (in the sense we clarify below) totem states have reduced system density operators \( \rho_m^{mn} \) close to the maximally mixed density operator \( (1/m)I_m \). To see this, notice that (adopting the convention of counting each eigenvalue, \( \lambda, \nu \) times when \( \nu \) is its multiplicity) amongst density operators, \( \rho \), on an \( m \)-dimensional Hilbert space, the eigenvalues of \( \rho \) have to satisfy both \( \sum_{a=1}^m \lambda_a = 1 \) (since every density operator has unit trace) as well as \( \sum_{a=1}^m \lambda_a^2 = \text{tr}(\rho^2) \) and one easily sees from these two conditions that the minimum value of \( \text{tr}(\rho^2) \) is \( 1/m \) and that this minimum value is attained only when each of the \( \lambda_a \) equals \( 1/m \). If we next consider the set of such \( \rho \) for which \( \text{tr}(\rho^2) \) is equal to \( 1/m + \eta \) where \( \eta \) denotes a (small) positive number, then one easily sees (again by considering the sum of the eigenvalues and the sum of their squares) that each of the \( \lambda_a \) must take the form \( 1/m + \delta_a \) where \( \sum_{a=1}^m \delta_a^2 = \eta \). Applying this result to each of our reduced density operators \( \rho_n^{mn} \), now writing the eigenvalues of each of these in the form \( 1/m + \delta_a \), then we immediately see that if \( \langle \text{tr}((\rho_n^{mn})^2) \rangle = 1/m + \eta \) (which will hold with \( \eta = (m+n)/(mn+1) - 1/m = (m^2 - 1)/(mn+1) \) which will be small if \( n \gg m \)) then the statement in words:

For \( n \gg m \), \( \rho_n^{mn} \) will probably be close to \( \frac{1}{m} I_m \) (100)

will hold in the sense that \( \sum_{a=1}^m \delta_a^2 = \eta \).

Similar results will obviously hold for \( \rho_n^{mn} \):

For \( m \gg n \), \( \rho_n^{mn} \) will probably be close to \( \frac{1}{n} I_n \) (101)
in a similar sense to above.

Our second qualitative consequence of Lubkin’s result for \( \rho_n^{mn} \) arises as a corollary to the above statement about \( \rho_n^{mn} \). To explain what it is, note first that, just as we had the formula (96) for the components, \( (\rho_n^{mn})_{a\hat{a}} \), of the \( m \times n \) matrix \( \rho_n^{mn} \), so we clearly have that \( \rho_n^{mn} = (\rho_n^{mn})_{kk} |f_k\rangle \langle f_k| \) where, in the notation of (95), the \( n \times n \) matrix \( \rho_n^{mn} \) is given by
\[ (\rho_n^{mn})_{kk} = c_{ak}c_{ak}^* \quad \text{(summed over} \ a). \] (102)
So, denoting by $C$ the $(m \times n)$ matrix whose components are the $c_{ak}$ and by $C^+$ its $(n \times m)$ adjoint matrix, we clearly have

$$\rho_{mn}^* = C C^+ \text{ and } \rho_{mn}^{*m} = C^+ C. \tag{103}$$

It easily follows from (103) that $x \in \mathbb{C}^m$ can be an eigenvector of $\rho_{mn}^*$ with a non-zero (positive) eigenvalue, $\lambda$, if and only if $y = \lambda^{-1/2}(C^+ x)^*$ $(\in \mathbb{C}^n)$ is an eigenvector of $\rho_{mn}$ with the same eigenvalue. (The factor of $\lambda^{-1/2}$ is easily seen to be needed if we want to ensure that $y$ is normalized whenever $x$ is normalized.) Moreover we note for future reference (in our digression on the Schmidt decomposition below) that we then have $C y^* = \lambda^{1/2} x$ – i.e.

$$e_{ak} y^{k*} = \lambda^{1/2} x^a \tag{104}$$

– the left hand side being summed over $k$. We conclude (continuing to adopt the convention of counting any eigenvalue $\nu$ times if it has multiplicity $\nu$) that, if $m > n$ and if $\rho_{mn}$ has eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\rho_{mn}^*$ will have this same set of eigenvalues together with $m - n$ more, all of which will, however, be zero! Moreover (cf. the discussion after equation (100)), since $\langle \text{tr}(\rho_{mn}^*)^2 \rangle = 1/n + \eta$ for $\eta$ now equal to $(m + n)/(mn + 1) - 1/n = (n^2 - 1)/(mn + 1)$ – which will be small if $m \gg n$ – we will have $\lambda_k = 1/n + \delta_k$ where $\sum_{k=1}^n (\delta_k^2) = \eta$. So, we may say that

For $m \gg n$, $\rho_{mn}^*$ will probably be close to

$$\frac{1}{n} \sum_{k=1}^n |\hat{e}_a|^2 \langle \hat{e}_a | \hat{e}_a |$$

where $\{\hat{e}_1, \ldots, \hat{e}_n\}$ is a basis for an $n$-dimensional subspace of $\mathcal{H}_m$ (which will depend on $\Psi$). This is the second qualitative consequence of Lubkin’s result we promised to arrive at at the outset. As far as we are aware, it does not appear to have been pointed out before. But, for our purposes, it will be of equal importance to the first consequence. Similarly, of course:

For $n \gg m$, $\rho_{mn}^*$ will probably be close to

$$\frac{1}{m} \sum_{a=1}^m |\hat{f}_a|^2 \langle \hat{f}_a | \hat{f}_a |$$

where $\{\hat{f}_1, \ldots, \hat{f}_m\}$ is a basis for an $m$-dimensional subspace of $\mathcal{H}_n$ (which will again depend on $\Psi$).

We remark that, in preparation for the argument we give below for the claim that $\rho_{m}^\text{modern}$ is well-approximated by $\rho_{m}^\text{modapprox}$, it is useful to observe that/how (105) and (101) are consistent with (97) and (98).

Further insight into the origin of (100), (101), (105) and (106) can be had by recalling that a given vector $\Psi \in \mathcal{H}_m \otimes \mathcal{H}_n$ – which we have written so far in the form (95) – can also be written as a single sum

$$\Psi = \sum_{i=1}^{\min(m,n)} \lambda_i^{1/2} \hat{e}_i \otimes \hat{f}_i \tag{107}$$

for suitable choices of bases $\{\hat{e}_1, \ldots, \hat{e}_n\}$ on $\mathcal{H}_m$ and $\{\hat{f}_1, \ldots, \hat{f}_n\}$ on $\mathcal{H}_n$. This is the well-known Schmidt decomposition (cf. e.g. [29] and/or the next paragraph) and the $\lambda_i$ are the same as those discussed above. (100) and (106) may then be viewed as (easy) consequences of the fact that, when $n \gg m$, the $\lambda_i$ in (107) are probably close to $1/m$ for $i = 1, \ldots, m$, while they are zero for $i > m$. Similarly (101) and (105) may be viewed as consequences of the fact that, when $m \gg n$, the $\lambda_i$ in (107) are probably close to $1/n$ for $i = 1, \ldots, n$, while they are zero for $i > n$.

The Schmidt decomposition in the form (107) can actually be derived easily from the results following Equation (103). In this paragraph, we digress to point out how, treating the cases where $n \geq m$: Denote by $\{x_1, \ldots, x_m\}$ a complete set of orthonormal eigenvectors of $\rho_{mn}^*$, and denote by $x_i^a$ the $a$th component of $x_i$. In view of the sentence following Equation (103), we can clearly find a complete set, $\{y_1^*, \ldots, y_n^*\}$, of orthonormal eigenvectors of $\rho_{mn}$ so that, denoting by $y_i^{k*}$ the $k$th component of $y_i^*$, we have, by (104), $c_{ak} y_i^{k*} = \lambda_i^{1/2} x_j^a$ (the left hand side being summed over $k$). (We only need to make sure that the $i$-value of every $x_j^a$ belonging to each non-zero eigenvalue of $\rho_{mn}$ matches the $j$-value of a $y_i^*$ belonging to an equal non-zero eigenvalue of $\rho_{mn}^*$ – there being necessarily an equal number of each; for any other $y_j^a$ [and of course there have to be others whenever $n > m$] the right hand side will anyway vanish.) Also introduce a new basis $\{\hat{e}_1, \ldots, \hat{e}_n\}$ for $\mathcal{H}_m$ such that $c_a = x_i^a \hat{e}_i$ (summed over $i$) and, similarly, introduce a new basis $\{\hat{f}_1, \ldots, \hat{f}_n\}$ for $\mathcal{H}_n$ such that $f_k = y_i^k \hat{f}_j$ (summed over $j$). Then we have (see (95))

$$\Psi = c_{ak} e_a \otimes f_k \text{ (summed over } a \text{ and } k) = c_{ak} x_i^a y_j^k \hat{e}_i \otimes \hat{f}_j \text{ (summed over } i, j, a \text{ and } k)$$

= by (104) \[ \lambda_j^{1/2} x_i^a x_j^a \hat{e}_i \otimes \hat{f}_j \text{ (summed over } i, j \text{ and } a) = \lambda_j^{1/2} \delta_{ij} \hat{e}_i \otimes \hat{f}_j \text{ (summed over } i \text{ and } j) \]
So we have
\[
\Psi = \lambda_i^{1/2} \hat{\epsilon}_i \otimes \tilde{f}_i \quad \text{(summed over } i)\]
thus establishing \(\text{[107]}\) in cases where \(n \geq m\). The cases where \(m \geq n\) are obviously similar. This ends our digression.

XII. MAIN ARGUMENT FOR THE VALIDITY OF THE APPROXIMATION \(\text{[15]}\) OF \(\rho_S^{\text{modern}}\) BY \(\rho_S^{\text{modapprox}}\)

Let us now turn to consider the set of unit vectors, \(\Psi\), in the Hilbert space of a totem as specified in Section \(\text{[1C]}\) — i.e. in the subspace of states with total energies in the range \([E, E + \Delta]\) of \(H_S \otimes H_B\). Allowing ourselves to make the slight distortion explained before equations \(\text{[1]}\) and \(\text{[2]}\), we may assume \(H_S\) has an orthonormal basis consisting of vectors \(|\epsilon_S, i\rangle\), where \(\epsilon_S\) ranges from \(\Delta\) to \(E\) in steps of \(\Delta\), while, for each \(\epsilon_S\), the integer, \(i\), ranges from 1 to \(n_S(\epsilon_S)\); and similarly \(H_B\) has an orthonormal basis consisting of vectors \(|\epsilon_B, j\rangle\), where \(\epsilon_B\) ranges from \(\Delta\) to \(E\) in steps of \(\Delta\), while, for each \(\epsilon_B\), \(j\) ranges from 1 to \(n_B(\epsilon_B)\). Below, and as in Section \(\text{[4]}\) we shall sometimes drop the \(S\) and \(B\) subscripts on the \(\epsilon_S\) when no ambiguity can arise.) Then, we are interested in the set of unit vectors, \(\Psi\), in the subspace, which we shall call below \(H_M\), of \(H_S \otimes H_B\) with total energy exactly \(E\). The reason for the name \(H_M\) is that \(H_M\) will clearly have dimension \(M\), where \(M\) is as in \(\text{[7]}\).

Each such \(\Psi \in H_M\) is writeable in the form
\[
\Psi = \sum_{\epsilon=\Delta}^{E} \sum_{i=1}^{n_S(\epsilon)} \sum_{j=1}^{n_B(\epsilon - \epsilon)} c_{ij}^{|\epsilon, i\rangle \otimes |E - \epsilon, j\rangle} \quad (108)
\]
where we recall (see above and cf. before Equation \(\text{[6]}\)) that the sum over \(\epsilon\) goes up in integer multiples of \(\Delta\). We also note that since \(\Psi\) is a unit vector, the sum (with the above indicated ranges) over \(\epsilon, i\) and \(j\) of \(|c_{ij}|^2\) equals 1. For such a \(\Psi\), the partial trace of \(|\Psi\rangle \langle \Psi|\) over \(H_B\), i.e. the reduced density operator, \(\rho_S^{\text{modern}}\), on \(H_S\), will then clearly be given by
\[
\rho_S^{\text{modern}} = \sum_{\epsilon=\Delta}^{E} \sum_{i=1}^{n_S(\epsilon)} \sum_{j=1}^{n_B(\epsilon - \epsilon)} |\epsilon, i\rangle \langle \epsilon, i| \quad (109)
\]
where
\[
r^{\text{S}\epsilon}_{\epsilon} = \sum_{j=1}^{n_B(\epsilon - \epsilon)} c_{ij}^* \hat{\epsilon}_i^* \quad (110)
\]
We shall find it useful sometimes to think of \(H_S\) as a direct sum \(\oplus_{\epsilon=\Delta}^{E} H^S_{\epsilon}\) (and similarly for \(H_B\)) where \(H^S_{\epsilon}\) is spanned by the \(|\epsilon, i\rangle\) for fixed \(\epsilon\) as \(i\) varies from 1 to \(n_S(\epsilon)\) and we shall call the restriction of \(\rho_S^{\text{modern}}\) to \(H^S_{\epsilon}\) simply \(r^{\text{S}\epsilon}_{\epsilon}\) — its \(i\)\(\tilde{i}\) components in the basis consisting of the \(|\epsilon, i\rangle\) being obviously the \(r^{\text{S}\epsilon}_{\epsilon}\) introduced above.

Our aim is to give an argument in favour of the claimed correctness of the proposition, which we state in Section \(\text{[1D]}\) that, in situations of interest, \(\rho_S^{\text{modern}}\) will be well-approximated by the \(\rho_S^{\text{modapprox}}\) of \(\text{[15]}\) and, in the course of giving this argument, to make clear how the \(n_B(\epsilon - \epsilon)\)-dimensional subspace of the \((n_S(\epsilon))\)-dimensional energy-\(\epsilon\) subspace of \(H_S\) spanned by the \(|\epsilon, i\rangle\) depend on \(\Psi\). (We should amplify on this statement by explaining that, when we say that \(\rho_S^{\text{modern}}\) is well-approximated by \(\rho_S^{\text{modapprox}}\), what we mean is that the values of physical quantities of interest, such as the mean energy and the entropy of the system \(S\), calculated using \(\rho_S^{\text{modapprox}}\) will be close to the values of the same quantities calculated using \(\rho_S^{\text{modern}}\).)

The main ingredients in our argument concern the average, \(\langle r^{\text{S}\epsilon}_{\epsilon}\rangle\), of \(r^{\text{S}\epsilon}_{\epsilon}\) and also the average, \(\langle tr((r^{\text{S}\epsilon}_{\epsilon})^2)\rangle\), of \(tr((r^{\text{S}\epsilon}_{\epsilon})^2)\); the calculations of these averages are closely similar to the preliminary calculations we carried out above for the average of the density operator \(\rho_m^{\text{modapprox}}\) and the average of the trace of its square; the difference being that we are now averaging over all unit \(\Psi\) in our full \(M\)-dimensional Hilbert space \(H_M\) (with \(M\) as in \(\text{[7]}\)) even though what we are averaging is only the restriction, \(r^{\text{S}\epsilon}_{\epsilon}\), (and the trace of the square of the restriction) of \(\rho_S^{\text{modern}}\) to \(H^S_{\epsilon}\) for fixed \(\epsilon\). As a result, while the counterpart of the product, \(mn\), in the denominator in our preliminary calculation would just be the single product \(n_S(\epsilon)n_B(E - \epsilon)\) were our average only to be over \(H^S_{\epsilon} \otimes H^B_{E - \epsilon}\), since we average over the unit vectors of the full Hilbert space \(H_M\), the counterpart will be turn out to be \(M\). Aside from this difference, to calculate \(\langle r^{\text{S}\epsilon}_{\epsilon}\rangle\) one proceeds
very similarly to the passage, above, between equations \((96)\) and \((97)\); the reader can easily supply the details simply by replacing \((96)\) by \((110)\) etc. One clearly obtains (instead of \((97)\))

\[
\langle r_{r}^{S} \rangle = \frac{n_{B}(E - \epsilon)}{M} I_{r}^{S}
\]

\((111)\)

where \(I_{r}^{S}\) denotes the identity on \(H_{r}^{S}\). Similarly, proceeding as in the passage between equations \((97)\) and \((99)\) (but again it will turn out that one needs to replace \(mn\) in the denominator by \(M\)) we easily find:

\[
\langle \text{tr}((r_{r}^{S})^{2}) \rangle = \frac{2n_{B}(E - \epsilon)n_{S}(\epsilon) + n_{B}(E - \epsilon)ns(\epsilon)n_{S}(\epsilon) - 1 + n_{B}(E - \epsilon)(n_{B}(E - \epsilon) - 1)n_{S}(\epsilon)}{M(M + 1)}
\]

\[
= \frac{n_{B}(E - \epsilon)n_{S}(\epsilon)(n_{B}(E - \epsilon) + n_{S}(\epsilon))}{M(M + 1)}
\]

\((112)\)

Now, rather as in our arguments for the two qualitative consequences of Lubkin’s result, (but now our arguments will involve both the counterpart, \((111)\), to \((97)\) as well as the counterpart, \((112)\), to \((99)\)) we observe from \((112)\) that, whenever \(n_{S}(\epsilon) \gg n_{B}(E - \epsilon)\), \(\langle \text{tr}((r_{r}^{S})^{2}) \rangle\) will be very close to \(M^{-2}(n_{B}(E - \epsilon))^{2}\), which, in the presence of \((111)\), easily implies (i.e. by similar reasoning to that used above in our derivation of \((100)\) and \((101)\)) that \(r_{r}^{S}\) must be very close to \(M^{-1}n_{B}(E - \epsilon)\sum_{i=1}^{n_{B}(E - \epsilon)}|\epsilon, i\rangle\langle \epsilon, i|\). Moreover, whenever \(n_{B}(E - \epsilon) \gg n_{S}(\epsilon)\), \(\langle \text{tr}((r_{r}^{S})^{2}) \rangle\) will be very close to \(M^{-2}(n_{S}(\epsilon))^{2}\), which, again in the presence of \((111)\), easily implies (i.e. by similar reasoning to that used above in our derivation of \((105)\) and \((106)\)) that \(r_{r}^{S}\) will be very close to \(M^{-1}n_{S}(\epsilon)\sum_{i=1}^{n_{S}(\epsilon)}|\epsilon, i\rangle\langle \epsilon, i|\) where \(|\epsilon, i\rangle\) denote the elements of an orthonormal basis for an \(n_{B}(E - \epsilon)\)-dimensional subspace of the \((n_{S}(\epsilon)\)-dimensional) energy-\(\epsilon\) subspace, \(H_{r}^{S}\) of \(H_{S}\) which will depend on \(\Psi\). Comparing these conclusions with the form of Equation \((15)\) we immediately see that, if it were the case that for all \(\epsilon\), either \(n_{S}(\epsilon) \gg n_{B}(E - \epsilon)\) or \(n_{B}(E - \epsilon) \gg n_{S}(\epsilon)\), then \((15)\) would obviously be a good approximation (at least each term in the sum over \(\epsilon\) will be) for all \(\epsilon\). However, of course, in typical situations of interest, there will be a region of \(\epsilon\) values around the value \(E_{c}\) – see the definitions of terms immediately after Equation \((15)\) – where neither of these statements will hold and \(n_{S}(\epsilon)\) and \(n_{B}(E - \epsilon)\) will be of comparable size. (We remark in passing, though, that, typically, [one will be able to choose \(\Delta\) so that] each of these quantities will be much greater than 1 for all or very nearly all \(\epsilon\) which are multiples of \(\Delta\) and in the range \([0, E]\).)

Nevertheless for the sort of situations of interest to us – and, in particular, for the densities of states which increase according to the power law, \((18)\), as considered in Section \(\text{II}\) or which increase exponentially, \((34)\), as considered in Sections \(\text{III}\) and \(\text{V}\), or which increase as quadratic exponentials, \((38)\), as considered in Section \(\text{VI}\) – and assuming the totem energy \(E\) and our choice of energy-increment, \(\Delta\) (see Section \(\text{I}\)) are such that \(M \gg 1\) (to ensure that the system [and bath] has access to a very large number of states) – one can check that the region of \(\epsilon\)-values around \(E_{c}\) where \(n_{B}(E - \epsilon)\) and \(n_{S}(\epsilon)\) are of comparable size will always be very small in size compared to \(E\), while the sum over this region of \(n_{S}(\epsilon)n_{B}(E - \epsilon)\) will be very small compared to \(M\). (In other words, the integral over this energy-region of the energy-probability density \(P_{S}(\epsilon)\) [see \((10)\)] will be very much less than 1.) Moreover, as \(\epsilon\) decreases towards zero, or increases towards \(E\) from \(E_{c}\), then for all three densities of states, \((18), (34), (38)\), one may check that the ratio \(n_{S}(\epsilon)/n_{B}(E - \epsilon)\), respectively \(n_{B}(E - \epsilon)/n_{S}(\epsilon)\), and hence the counterparts (i.e. with \(n_{S}(\epsilon)\) replacing \(m\) and \(n_{B}(E - \epsilon)\) replacing \(n\)) to the quantities which we called \(\eta\) before Equation \((100)\), respectively Equation \((105)\), will get rapidly smaller and hence the relevant notion of closeness (i.e. as in \((100), (105)\)) will get rapidly stronger. It is then straightforward to argue from these statements that quantities of interest such as \((\epsilon_{S}^{\text{modapprox}})^{2}\) itself, and (cf. \((41)\)) \(S_{\text{modapprox}}^{\text{modapprox}} = -\text{tr}(\rho_{S}^{\text{modapprox}} \log \rho_{S}^{\text{modapprox}})\) will be closely approximated (respectively) by \(\epsilon_{S}^{\text{modapprox}} = \text{tr}(\rho_{S}^{\text{modapprox}} H_{S})\), \(\text{tr}(\rho_{S}^{\text{modapprox}} H_{S})\), \(\text{tr}(\rho_{S}^{\text{modapprox}} H_{S})\), \((\rho_{S}^{\text{modapprox}})^{2}\), and (cf. \((41)\)) \(S_{\text{modapprox}}^{\text{modapprox}} = -\text{tr}(\rho_{S}^{\text{modapprox}} \log \rho_{S}^{\text{modapprox}})\).

Concerning the latter two quantities – i.e. the trace of the square of the reduced density operator of \(S\) and its von Neumann entropy – there are reasons to expect the approximation of \(S_{S}^{\text{modapprox}}\) by \(S_{S}^{\text{modapprox}}\) to be even better than the approximation of \(\text{tr}(\rho_{S}^{\text{modapprox}})^{2}\) by \(\text{tr}(\rho_{S}^{\text{modapprox}})^{2}\).

XIII. FURTHER CHECKS AND DETAILS ON THE VALIDITY OF THE APPROXIMATION \((15)\)

As a partial check of various aspects of all of the above argument, and in justification of our latter remark, it is instructive first to consider the case where, for all \(\epsilon = 0, \Delta, 2\Delta, \ldots\) in the range \([0, E]\), we have \(n_{S}(\epsilon) = 1 = n_{B}(E - \epsilon)\) where, of course it is \textit{never} true that \(n_{S}(\epsilon) \gg n_{B}(E - \epsilon)\) or that \(n_{S}(\epsilon) \ll n_{B}(E - \epsilon)\) (nor that each of these quantities is very much greater than 1) so we can think of this as one sort of ‘worst case scenario’. Of course this is not an example that interests us in Part 1, but it would apply e.g. to a totem consisting of a pair of weakly coupled quantum
harmonic oscillators (with equal spring constants) for a total energy much greater than the level spacing (and a choice of $\Delta$ equal to the level spacing). For this model, we may clearly write

$$\Psi = \sum_{\epsilon=\Delta}^E c_\epsilon |E - \epsilon\rangle$$

so that the reduced density operator of the system, $S$, will be

$$\rho_S^{\text{modern}} = \sum_{\epsilon=\Delta}^E |c_\epsilon|^2 |\epsilon\rangle\langle \epsilon|,$$

while the formula, (15), for $\rho_S^{\text{modapprox}}$ becomes simply

$$\rho_S^{\text{modapprox}} = \frac{1}{M} \sum_{\epsilon=\Delta}^E |\epsilon\rangle\langle \epsilon|,$$

where (cf. (7)) $M = E/\Delta$. Clearly, (cf. (38)) the approximate mean energy,

$$\bar{\epsilon}_S^{\text{modapprox}} := \text{tr}(\rho_S^{\text{modapprox}} H_S) = \frac{1}{M} \sum_{\epsilon=\Delta}^E \epsilon$$

$$= \frac{1}{2M} \left( \frac{E}{\Delta} + 1 \right) = \frac{E}{2}.$$ 

(In calculating the value of the above sum, we need of course to recall that the sum is over values of $\epsilon$ which are integer multiples of $\Delta$.) Therefore, since this doesn’t depend on the $c_\epsilon$, its average over Haar measure (indicated with “$\langle \rangle$”) takes the same value:

$$\langle \bar{\epsilon}_S^{\text{modapprox}} \rangle = \frac{E}{2}.$$ 

On the other hand, the average over Haar measure of the exact mean energy, $\bar{\epsilon}_S^{\text{modern}}$, may be calculated as follows:

$$\langle \bar{\epsilon}_S^{\text{modern}} \rangle = \langle \sum_{\epsilon=\Delta}^E \epsilon |c_\epsilon|^2 \rangle$$

$$= \sum_{\epsilon=\Delta}^E \epsilon \int w_\epsilon d\text{Haar}$$

where the integral is over the complex $M$-dimensional sphere of unit vectors in the Hilbert space, $H_M$, spanned by the vectors of form (113), coordinatized with $w_\epsilon$ ranging over the $(M-1)$-simplex and $\theta_\epsilon$ ranging over the $M$-torus as explained at the beginning of Section XI where $w_\epsilon = |c_\epsilon|^2$ etc. Obviously the torus factor of the integral just gives 1, so the integral has, by (92), the value $1/M$ for each $\epsilon$. So we conclude that $\langle \bar{\epsilon}_S^{\text{modern}} \rangle$ has the same value as $\langle \bar{\epsilon}_S^{\text{modapprox}} \rangle$, i.e.

$$\langle \bar{\epsilon}_S^{\text{modern}} \rangle = \frac{E}{2},$$

which, of course, has to be the correct value by the symmetry under the interchange of $S$ and $B$ in this case.

Turning to averages over Haar measure of the trace of the square of the reduced density operator of $S$, we have, on the one hand,

$$\langle \text{tr}((\rho_S^{\text{modapprox}})^2) \rangle = \langle \sum_{\epsilon=\Delta}^E \frac{1}{M^2} \rangle = \sum_{\epsilon=\Delta}^E \frac{1}{M^2} = \left( \frac{\Delta}{E} \right)^2 \left( \frac{E}{\Delta} \right)$$
\[
\langle \text{tr}(\rho_S^{\text{modern}})^2 \rangle = \frac{E}{M} \quad (\text{= } \frac{\Delta}{E}).
\]

Whereas, on the other hand,

\[
\langle \text{tr}(\rho_S^{\text{modapprox}})^2 \rangle = \frac{E}{M} \frac{2}{M(M+1)} \simeq \frac{2}{M} \quad (\simeq \frac{2\Delta}{E})
\]

(where we have used \([33]\) in calculating the integral) which differs from the approximate value by a factor of 2!

However, if we turn to calculate the averages over Haar measure of the von Neumann entropies of the approximate and exact reduced density operator of S, we find, on the one hand,

\[
\langle S^{\text{modapprox}} \rangle = \langle -\text{tr}(\rho_S^{\text{modapprox}} \log \rho_S^{\text{modapprox}}) \rangle = \langle \log M \rangle = \log M \quad (= \log(E/\Delta)).
\]

On the other hand,

\[
\langle S^{\text{modern}} \rangle = \langle -\text{tr}(\rho_S^{\text{modern}} \log \rho_S^{\text{modern}}) \rangle = -M \int_{\text{Unit Sphere in } \mathbb{C}^M} w_1 \log w_1 d\text{Haar}
\]

\[
= \text{(by } [89] \text{ and } [90] \text{)} -M(M-1)! \int_{\text{Simplex}} w_1 \log w_1 dw_1 \ldots dw_{M-1}
\]

\[
= -M(M-1) \int_0^1 w \log(1-w)^{M-2} dw.
\]

One may do this integral by noticing that \(w \log w = (dw^\alpha/d\alpha)|_{\alpha=1} - \text{obtaining for its value, } d(B(\alpha+1, M-1))/d\alpha|_{\alpha=1}\) where \(B\) denotes the beta function (see e.g. \([34]\)). One finds that \(-M(M-1)\) times this simplifies (using \([21]\)) to \(\psi(1 + M) - \psi(2)\) where \(\psi\) denotes the \(\psi\) (or ‘digamma’) function defined by \(\psi(x) = d\log\Gamma(x)/dx\), and this \([50]\), in turn, equals \(\sum_{k=2}^{M} 1/k\) which, by the standard asymptotic expansion of Euler’s constant, \(C (= 0.5772\ldots)\) is equal to \(\log M + C - 1 + 1/2M + O(1/M^2)\). So we conclude that

\[
\langle S^{\text{modern}} \rangle = \log M + C - 1 + O(1/M) \quad (= \log(E/\Delta) + C - 1 + O(\Delta/E))
\]

where \(C\) is Euler’s constant \((0.5772\ldots)\).

Comparison of \([115]\) and \([114]\) shows that the use of \([15]\) for this ‘worst case scenario’ leads to a von Neumann entropy which, for large \(M\), is very close to the average over Haar measure of its actual value. In view of the fact that both of these values are very close to the maximum possible value, \(\log M\), of the entropy of any density operator on \(\mathcal{H}_S\) (which is of course \(M\)-dimensional in this case) we conclude both that most totem states, \(\Psi\), for this model must have a reduced density operator on \(S\) whose von Neumann entropy is close to \(\log M\); and that the use of \([15]\) leads to a good approximation for this value. And both of these things hold even though, as we saw above, our general arguments do not apply to this case and even though, for this case, as we saw above, \([15]\) leads to a trace of the square of the reduced density operator of \(S\) which is only half of the average over Haar measure of its actual value.

We will next use the Lubkin-Page asymptotic formula, \([87]\), to obtain a result which tends to confirm the accuracy of our general formula \([11]\), obtained using \([15]\), for the von Neumann entropy for our densities of states of interest, \([18]\), \([34]\), \([58]\). Our result will show that the value of the von Neumann entropy obtained with \([15]\) well-approximates a certain restricted average of the exact von Neumann entropy. Before we present this result, we shall find it helpful to first explain what we mean here by a ‘restricted average’ in a different context:

Let us look back at the result essentially due to Lubkin, \([99]\), which we (re-)obtained above, for the average over Haar measure on vectors, \(\Psi\), belonging to the tensor product, \(\mathcal{H}_m \otimes \mathcal{H}_n\), of two Hilbert spaces, of the trace of the square of the reduced density operator, \(\rho_m^{mn}\) on \(\mathcal{H}_m\). Averaging over all totem vectors, \(\Psi \in \mathcal{H}_m \otimes \mathcal{H}_n\), amounts, as we explained above, to averaging with the invariant measure on the complex \(mn\)-sphere over the coefficients, \(c_{ak}\), in the basis-expansion, \([95]\), of \(\Psi\), which, in turn, writing \(c_{ak}\) as \(w_{ak}^{1/2} e^{i\theta_{ak}}\), we saw, amounts to integrating w.r.t. the \(w_{ak}\) over the \((mn-1)\)-simplex and w.r.t. the \(\theta_{ak}\) over the \(mn\)-torus. What we now wish to point out is that, if we restrict to \(c_{ak}\) which take the form \((1/\sqrt{mn}) e^{i\theta_{ak}}\) and just average over these (i.e. by integrating with respect to the \(\theta_{ak}\) over the \(mn\)-torus) one easily finds – denoting our restricted average with the symbol ‘[ ]’ – that

\[
[\text{tr}(\rho_m^{mn})^2] = \frac{m + n - 1}{mn},
\]
and this is not a bad approximation to the value, \( (m + n)/(mn + 1) \), of the full average, \( \text{tr}((\rho^{mn})^2) \), of \( \text{tr}(\rho^{mn})^2 \) – the two expressions differing, in fact, only by terms of order \( 1/mn \). In terms of our geometrical picture, in which averaging w.r.t. Haar measure amounts to integrating over the \((mn - 1)\)-simplex times the \(mn\)-torus, in passing from the unrestricted average, \( \langle \cdot \rangle \), to our restricted average, \( \langle [\cdot] \rangle \), we have replaced the integral over the \((mn - 1)\)-simplex by the value at its centroid (where \( w_{ak} = 1/mn \) for all \( a \) and \( k \)), but continue to integrate over the \(mn\)-torus. Of course this restricted average, \( \langle [\cdot] \rangle \), is a basis-dependent notion, but what we have learnt is that we can choose any bases \( \{e_1, \ldots, e_m\} \) and \( \{f_1, \ldots, f_n\} \) we like and we obtain this reasonably good approximation to the unrestricted average (at least for the quantity \( \text{tr}((\rho^{mn})^2) \)).

We conclude that the corresponding restricted set of totem vectors (cf. [95]),

\[
\Psi = \frac{1}{\sqrt{mn}} e^{i\theta_{ak} e_a \otimes f_k} \text{ (summed over } a \text{ and } k) 
\]

is (for any choice of bases, \( \{e_1, \ldots, e_m\} \) and \( \{f_1, \ldots, f_n\} \) and as the \( \theta_{ak} \) range over the \(mn\)-torus) a sufficiently representative set of totem vectors for the restricted average over this set to indicate sufficiently well the behaviour of a generic totem state, \( \Psi \) (at least as far as \( \text{tr}(\rho^{mn})^2 \) is concerned).

We shall proceed in a similar spirit, but now for the totem of Section 1C. We expect that it won’t make too big a difference if, instead of averaging over the full set of totem vectors, \( \Psi \in \mathcal{H}_M \), we consider a suitable restricted average. To motivate the restriction that we shall make, we notice first, that, if we expand such vectors, \( \Psi \), as in (108), then it follows from (111) that the (unrestricted) average value (i.e. over Haar measure on the set of all \( \Psi \in \mathcal{H}_M \)) of the trace of each \( r^S_\epsilon \) (defined in the paragraph after (110)) is given by

\[
\langle \text{tr}(r^S_\epsilon) \rangle = \left( \left\langle \sum_{i=1}^{ns(\epsilon)} \sum_{j=1}^{n_B(1-\epsilon)} |c_i^j|^2 \right\rangle \right) = \frac{ns(\epsilon) n_B(1-\epsilon)}{M} = \frac{P_3(\epsilon) \Delta}{M}.
\]

\( - \) the equality in parenthesis following from (110).

In view of this, we take our restricted average to be over vectors, \( \Psi \in \mathcal{H}_M \), such that, in the expansion, (108), for each \( \epsilon \), the coefficients \( c_i^j \) are constrained to satisfy exactly

\[
\sum_{i=1}^{ns(\epsilon)} \sum_{j=1}^{n_B(1-\epsilon)} |c_i^j|^2 = \text{tr}(r^S_\epsilon) = \frac{ns(\epsilon) n_B(1-\epsilon)}{M}.
\]

(We remark that, in view of what we explained in the previous two paragraphs, we could alternatively restrict much further and simply average over \( \Psi \) in (108) for which every \( c_i^j \) takes the form \( e^{i\theta_{ak} e_a \otimes f_k} \) and still be able to arrive at similar conclusions to those below. However the restriction we adopt has the advantage of allowing us to directly use the Lubkin-Page approximation in exactly the form (87).) In other words, denoting \( n_S(\epsilon)n_B(1-\epsilon)/M \) by \( \mu_\epsilon \), we average over \( \Psi \in \mathcal{H}_M \) which take the form \( \oplus_{\epsilon=0}^{E=\Delta} \mathcal{H}_\epsilon^M \) (each \( \mathcal{H}_\epsilon^M \) being normalized) where we regard \( \mathcal{H}_M \) as the direct sum, \( \oplus_{\epsilon=0}^{E=\Delta} \mathcal{H}_\epsilon^M \), where, for each \( \epsilon \in [0, \Delta, \ldots, E) \), \( \mathcal{H}_\epsilon^M \) denotes the \((n_S(\epsilon)n_B(1-\epsilon))\)-dimensional Hilbert subspace of \( \mathcal{H}_M \) spanned by the vectors \( \{e_i, i|E-\epsilon, j\} \), \( i = 1 \ldots ns(\epsilon), j = 1 \ldots n_B(1-\epsilon) \) in \( \mathcal{H}_S^B \otimes \mathcal{H}_E^{1-\epsilon} \) – see after equation 110. For such restricted \( \Psi \), \( \rho_\epsilon^{\text{modern}} \) will take the form

\[
\rho_\epsilon^{\text{modern}} = \sum_{\epsilon=\Delta}^{E} \mu_\epsilon R^S_\epsilon
\]

where \( R^S_\epsilon \) is the partial trace of \( \langle \Psi_\epsilon^M \rangle \langle \Psi_\epsilon^M \rangle^\dagger \) over \( \mathcal{H}_E^{1-\epsilon} \) (which will equal \( r^S_\epsilon \) divided by its trace, which is \( \mu_\epsilon \)). Clearly, by the lemma in Section IV we therefore have

\[
S(\rho_\epsilon^{\text{modern}}) = \sum_{\epsilon=\Delta}^{E} \mu_\epsilon S(R^S_\epsilon) = \sum_{\epsilon=\Delta}^{E} \mu_\epsilon S(R^S_\epsilon) - \sum_{\epsilon=\Delta}^{E} \mu_\epsilon \log \mu_\epsilon.
\]

But now we notice that, if we identify \( m \) with \( n_S(\epsilon) \) and \( n \) with \( n_B(1-\epsilon) \), then we can identify \( \mathcal{H}_M^M \) with the Hilbert space, \( \mathcal{H}_m \otimes \mathcal{H}_n \), of Sections II and XI and, under this identification, \( R^S_\epsilon \) is identified with \( \rho_m^{mn} \), and \( S(R^S_\epsilon) \) with \( S(\rho_m^{mn}) \). Moreover, averaging \( S(R^S_\epsilon) \) over \( \mathcal{H}_m \), is, under (the reverse of) this identification, then obviously the same as taking the unrestricted average of \( S(\rho_m^{mn}) \) over Haar measure on unit vectors in \( \mathcal{H}_m \otimes \mathcal{H}_n \) and so we may estimate its value using the Lubkin-Page approximation (87). Making these identifications, if we now use \( \langle [\cdot] \rangle \) to denote
our restricted average over our restricted totem vectors, Ψ, and \( \langle \cdot \rangle \) to denote the unrestricted average over Haar measure on unit vectors in \( \mathcal{H}_s^M \) for each \( s \), we may calculate using the formula (117) in our lemma of Section IV:

\[
[S(\rho^\text{mod}_S^\text{modern})] = \left[ S \left( \sum_{\epsilon=\Delta}^{E} \mu_{\epsilon} S(\rho_{\epsilon}^S) \right) \right] = \sum_{\epsilon=\Delta}^{E} \mu_{\epsilon} (S(R_{\epsilon}) - \sum_{\epsilon=\Delta}^{E} \mu_{\epsilon} \log \mu_{\epsilon})
\]

which, recalling that \( \mu_{\epsilon} = n_{S}(\epsilon)n_{B}(E - \epsilon)/M \) and using (87), equals

\[
\sum_{\epsilon=\Delta}^{E} \frac{n_{S}(\epsilon)n_{B}(E - \epsilon)}{M} \left( \log(\min(n_{S}(\epsilon), n_{B}(E - \epsilon))) - \frac{\min(n_{S}(\epsilon), n_{B}(E - \epsilon))}{2\max(n_{S}(\epsilon), n_{B}(E - \epsilon))} \right)
\]

\[- \log \left( \frac{n_{S}(\epsilon)n_{B}(E - \epsilon)}{M} \right) + O \left( \frac{1}{n_{S}(\epsilon)n_{B}(E - \epsilon)} \right)
\]

which easily simplifies to

\[
[S(\rho^\text{mod}_S^\text{modern})] = -M^{-1} \left( \sum_{\epsilon=\Delta}^{E_{c}} n_{S}(\epsilon)n_{B}(E - \epsilon) \log(M^{-1} n_{B}(E - \epsilon)) + \sum_{\epsilon=E_{c}+\Delta}^{E} n_{S}(\epsilon)n_{B}(E - \epsilon) \log(M^{-1} n_{S}(\epsilon)) \right)
\]

\[- M^{-1} \left( \sum_{\epsilon=\Delta}^{E_{c}} \frac{n_{S}(\epsilon)^2}{2} + \sum_{\epsilon=E_{c}+\Delta}^{E} \frac{n_{B}(E - \epsilon)^2}{2} \right) + O(1)
\]

(118)

where \( E_{c} \) is as defined after (15).

Comparing (118) with (41), we notice that the first line of (118) coincides with the formula, (41), \( S_{S}^\text{modapprox} \) for the von Neumann entropy of \( \rho_{S}^\text{modapprox} \) which we derived from (15). Thus we may conclude that our restricted average over totem vectors of \( S_{S}^\text{mod} \) will be given by the formula we gave for \( S_{S}^\text{modapprox} \) in (41) – plus an ‘error term’ given by the last line of (118). Moreover, the close agreement found above between \( S_{S}^\text{mod} \) and \( S_{S}^\text{modapprox} \) in the ‘worst case scenario’ discussed above, strongly suggests that the same statement will be true for the unrestricted average. In order to conclude that this amounts to an independent check of the correctness of the approximate formula \( S_{S}^\text{mod} \) of (41) for our densities of states of interest, (18), (34), (58), it remains to show that (investigate when) the ‘error term’ (i.e. the second line in (118)) is small. To end this section we turn to this last question:

It is in fact easy to see (after converting the sum to an integral, using (8)) that: (a) for our power-law densities of states, (18), with \( A_{S} = A_{B} \) and \( N_{S} = N_{B} = N \) say, the last line of (118) (minus the \( O(1) \) term) is (using Stirling’s approximation – see Section I) \( 1/\sqrt{\pi N} \); (b) for our (equal) exponential densities of states, (34), it is \( (1/bE)(1-e^{-AE}) \); and (c) for our (equal) quadratic densities of states, (58), it is (approximating the integral with the leading term of the asymptotic formula in Endnote (12)) \( \exp(-qE^2/2) \). These terms will all be much smaller than typical values of the first line of (118), provided \( N \) is large: (a), provided \( E \gg 1/b \) (cf. Equation (36)) in (b), and provided \( E \gg 1/\sqrt{q} \) (cf. Equation (59)) in (c).

So in all cases of interest here, and, no doubt, in many others too, the last line of (118) will be negligibly small.

Acknowledgments

I thank Michael Kay for useful discussions about traditional statistical mechanics and also for information about string theory and Richard Hall for showing me how to do integrals involving logarithms.

[1] In the case the energy probability density is sharply peaked, \( T(\epsilon) \) would then be interpretable as the temperature of a small subsystem or of a small system placed in thermal contact with the given system. See Section VII A.
[2] S.W. Hawking, Commun. Math. Phys. 43, 199 (1975)
[3] S.W. Hawking, *Hawking on the Big Bang and Black Holes* (Advanced Series in Astrophysics and Cosmology 8) (World Scientific, Singapore 1993)
[4] G.W. Gibbons and S.W. Hawking (eds.), *Euclidean Quantum Gravity* (World Scientific, Singapore 1993)
[5] G.W. Ford, M. Kac and P. Mazur, J. Math. Phys. 6, 504 (1965)
Entropy defined, entropy increase and decoherence understood, and some black-hole puzzles solved.

[15] Throughout this paper, we shall often consider, side-by-side with our finite-sum formulae, continuum approximations where we make the replacement \( b \rightarrow n \) and \( \kappa \rightarrow \infty \). Usually, the formulae are essentially interchangeable and one can (and it can be helpful to) check they give the same answers. However, we caution that, for example, the continuum versions of the entropy formulae which we obtain, using the prescription in Section 13, can end up having energy-intervals on which the ‘number of energy levels’ is greater than zero but less than 1 and this can result in spurious negative terms which go like the logarithm of (a constant with the dimensions of inverse energy times the) total energy \( E \). An example of such a term occurs in Section 14 which is as mentioned there.

The fact that both the traditional microcanonical approach, and also the more modern approach based on a total pure state along the lines of Goldstein, Lebowitz et al. and Popescu, only explain how a small part of a totem can be thermal also raises obvious puzzles, not only for black holes, but also for cosmology.

E. Lubkin, J. Math. Phys. 19, 1028 (1978)

Actually Lubkin was by no means the first to adopt the natural invariant measure (‘Haar measure’) in a related context.

E. Lubkin, J. Math. Phys. 10, 1827 (1969)

B. S. Kay, Class. Quant. Grav. 15, L89 (1998) arXiv:hep-th/9802172

B. S. Kay, Entropy defined, entropy increase and decoherence understood, and some black-hole puzzles solved. arXiv:hep-th/0511225

B. S. Kay and Y. Abyaneh, Expectation values, experimental predictions, events and entropy in quantum gravitationally decohered quantum mechanics, arXiv:0710.0992

S. Popescu, A. J. Short and A. Winter, Nature Physics 2, 754 (2006) arXiv:quant-ph/0511225

S. Popescu, A. J. Short and A. Winter, Nature Physics 2, 754 (2006) [arXiv:cond-mat/0511091]

E. B. Davies, Quantum Theory of Open Systems (Academic, London 1976)

R. P. Feynman, Statistical Mechanics (W. A. Benjamin Inc., Reading Mass. 1972)

W. Thirring, A Course in Mathematical Physics 1: Quantum Mechanics of Large Systems (Springer, New York 1980)

S. Goldstein, J. L. Lebowitz, R. Tumulka and N. Zanghi, Phys. Rev. Lett. 96, 050403 (2006) [arXiv:cond-mat/0511091]

E. B. Davies, Quantum Theory of Open Systems (Academic, London 1976)

S. Thirring, Mathematical Physics (Springer, New York 1976)

W. Feller, An Introduction to Probability Theory and its Applications. Volume I (Academic, London 1976)
In fact, allowing different values of $c$ ($c_S$ and $c_B$) in the obvious places in (24) will, provided $c_S$ and $c_B$ remain of comparable size, only introduce small corrections to the quantities we calculate. In particular, as may be readily checked. (Hint: write $c_S = ce^b$ and $c_B = ce^{-b}$. Then $E_c$ will be $E/2 - \mu$ the expected energies of $S$ and $B$ will remain exactly unchanged as $E/2$, while there will be only small corrections to the entropies. In particular, $S^{\text{mod}}_B$ generalizes from $bE/4 + \log(cE)$ to $bE/4 + \log(c_S c_B E^2)/2 - (\log(c_S/c_B))^2/4E$. Different values of $b$ will, as might be expected, have a more drastic effect. See also [23] where we will consider densities of states of the form $\sigma(x) = ce^{-h}e^{\nu x}$ i.e. with inverse-power prefactors such as occur in string theory.

[30] I.S. Gradshteyn and I.M. Ryzhik, Table of Integrals, Series and Products Corrected and enlarged edition (Academic, London 2007)

[31] A. Wehrl, Rev. Mod. Phys. 50, 221 (1978)

[32] B.S. Kay, Commun. Math. Phys. 1148 (1994); J. Sánchez-Ruiz, Phys. Rev. E 52, 5653 (1995); S. Sen, Phys. Rev. Lett. 77, 1 (1996)). What will be of relevance for the present paper however, is that the choice of $\Delta$ is ‘harmless’ is confirmed by the fact (see Section IV) that our formulae for mean energy and entropy are independent of $\Delta$.

[33] R.T. Powers and E. Størmer, Commun. Math. Phys. 58, 103 (1979))

[34] R.T. Powers and E. Størmer, Commun. Math. Phys. 58, 103 (1979).

[35] Y. Takahashi and H. Umezawa, Collective Phenomena 379B, Some speculations about black hole entropy in string theory (1992).

[36] T. Smirnov and C. Vafa, Phys. Lett. 379B, 99 (1996).}

[37] A. Wehrl, Rev. Mod. Phys. 50, 221 (1978)

[38] G.T. Horowitz, Quantum states of black holes. In R.M. Wald (editor) Black Holes and Relativistic Stars (University of Chicago Press, Chicago 1998) arXiv:gr-qc/9704072

[39] G.T. Horowitz and J. Polchinski, Phys. Rev. D 55, 6189 (1997)

[40] A. Wehrl, Rev. Mod. Phys. 50, 221 (1978)

[41] B.S. Kay, Commun. Math. Phys. 1148 (1994); J. Sánchez-Ruiz, Phys. Rev. E 52, 5653 (1995); S. Sen, Phys. Rev. Lett. 77, 1 (1996)). What will be of relevance for the present paper however, is that the choice of $\Delta$ is ‘harmless’ is confirmed by the fact (see Section IV) that our formulae for mean energy and entropy are independent of $\Delta$.

[42] D.N. Page, Phys. Rev. Lett. 71, 1291 (1993) arXiv:gr-qc/9305007

[43] G.T. Horowitz and J. Polchinski, Phys. Rev. D 55, 6189 (1997)

[44] A. Wehrl, Rev. Mod. Phys. 50, 221 (1978)

[45] B.S. Kay, Commun. Math. Phys. 1148 (1994); J. Sánchez-Ruiz, Phys. Rev. E 52, 5653 (1995); S. Sen, Phys. Rev. Lett. 77, 1 (1996)). What will be of relevance for the present paper however, is that the choice of $\Delta$ is ‘harmless’ is confirmed by the fact (see Section IV) that our formulae for mean energy and entropy are independent of $\Delta$.

[46] D.N. Page, Phys. Rev. Lett. 71, 1291 (1993) arXiv:gr-qc/9305007

[47] D.N. Page, Phys. Rev. Lett. 71, 1291 (1993).

[48] Page also conjectured the exact general formula $\langle\Sigma\rho_n\rangle = \sum_{k=0}^{m-1} \frac{1}{2^n - \frac{m-1}{2}}$, the correctness of which was, soon afterwards, proven by a number of authors (S.K. Foong and S. Kanno, Phys. Rev. Lett. 72 1148 (1994); J. Sánchez-Ruiz, Phys. Rev. E 52, 5653 (1995); S. Sen, Phys. Rev. Lett. 77, 1 (1996)). What will be of relevance for the present paper however, and particularly for the results we obtain in the remaining sections of Part 2 are mainly the original results due to Lubkin – in particular, his result on the validity of the approximations (55) and (in Section XIII) what we call there the ‘Lubkin-Page’ approximation (57).

[49] The equality of $\Sigma\rho_n\rho_n^\dagger$ and $\Sigma\rho_n\rho_n^\dagger$ follows by the well-known general result that, for a pure total state on any bipartite system, the von Neumann entropies of the two reduced density operators are necessarily always equal. As is well-known, this is easily proven from the Schmidt decomposition which, by the way, we shall recall, in passing, in another context, in Section XI. In fact, by (107), given an arbitrary state vector, $\Psi \in \mathcal{H}_m \otimes \mathcal{H}_n$ (with $\mathcal{H}_m$ and $\mathcal{H}_n$ any $m$- and $n$-dimensional Hilbert spaces) we clearly have that the reduced density operator, $\rho_m^\dagger$, of $|\Psi\rangle$ on $\mathcal{H}_m$ is

$$\sum_{i=1}^n \lambda_i |\tilde{e}_i\rangle\langle\tilde{e}_i|,$$

while the reduced density operator, $\rho_n^\dagger$, of $|\Psi\rangle$ on $\mathcal{H}_n$ is

$$\sum_{i=1}^n \lambda_i |\tilde{f}_i\rangle\langle\tilde{f}_i|,$$

where $\nu = \min(m,n)$. Clearly, the von Neumann entropy of each of these reduced density operators is $-\sum_{i=1}^\nu \lambda_i \log \lambda_i$. The fact that these two distinct probabilities are approximately equal is, arguably, intuitively obvious.

[29] In fact, allowing different values of $c$ ($c_S$ and $c_B$) in the obvious places in (24) will, provided $c_S$ and $c_B$ remain of comparable size, only introduce small corrections to the quantities we calculate. In particular, as may be readily checked. (Hint: write $c_S = ce^b$ and $c_B = ce^{-b}$. Then $E_c$ will be $E/2 - \mu$ the expected energies of $S$ and $B$ will remain exactly unchanged as $E/2$, while there will be only small corrections to the entropies. In particular, $S^{\text{mod}}_B$ generalizes from $bE/4 + \log(cE)$ to $bE/4 + \log(c_S c_B E^2)/2 - (\log(c_S/c_B))^2/4E$. Different values of $b$ will, as might be expected, have a more drastic effect. See also [23] where we will consider densities of states of the form $\sigma(x) = ce^{-h}e^{\nu x}$ i.e. with inverse-power prefactors such as occur in string theory.
[49] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge UK 2006)
[50] G. ’t Hooft, Nucl. Phys. **B256**, 727 (1985)
[51] S. Mukohyama and W. Israel, Phys. Rev. D **58**, 104005 (1998)
[52] B.S. Kay and L. Ortiz, *Brick Walls and AdS/CFT*, arXiv:1111.6429