Microscopic origin of thermodynamic entropy in isolated systems

J. M. Deutsch,[1] Haibin Li,[1,2] and Auditya Sharma[3]

[1]*Department of Physics, University of California, Santa Cruz, California 95064, USA*
[2]*Department of Applied Physics, Zhejiang University of Technology, Hangzhou 310023, P. R. China*
[3]*International Institute of Physics, Federal University of Rio Grande do Norte, Brazil*

The quantum entropy is usually defined using von Neumann's formula, which measures lack of information and vanishes for pure states. In contrast, we obtain a formula for the entropy of a pure state as it is measured from thermodynamic experiments, solely from the self-entanglement of the wave function, and find strong numerical evidence that the two are in agreement for nonintegrable systems, both for energy eigenstates and for states that are obtained at long times under the evolution of more general initial conditions. This is an extension of Boltzmann's hypothesis for classical systems, relating microscopic motion to thermodynamics.

PACS number(s): 05.70.−a, 89.70.Cf

I. INTRODUCTION

The emergence of thermodynamics from underlying microscopic motion is still unclear. One of the hardest concepts to understand is that of entropy. Classically, the microscopic state of a system can be thought of as a single point in phase space and as it evolves, this point wanders in this space, filling up more of it as time progresses. Boltzmann's hypothesis relates the entropy to the volume that can be filled, and shows how microscopic motion determines large-scale thermodynamic properties [1]. For a generic, that is, a nonintegrable, system, ergodicity implies that this volume will be the hypersurface of constant energy, but for integrable systems, the volume accessible depends on the initial conditions so that such systems will not thermalize. There are many different approaches and definitions for understanding entropy [2], but here we will be concerned with the experimental thermodynamic definition that is obtained, for example, by doing calorimetry measurements as related to the flow of heat. Classically, these thermodynamic measurements are not influenced by knowledge of the microscopic state of a system. For example, computer simulations are often used to predict thermodynamic quantities, such as entropy. However, microscopic knowledge of the system allows one to extract work out of it by the well-known example of Maxwell's demon. This would violate the second law of thermodynamics unless we correspondingly reduce the entropy to take into account this microscopic information [2]. Therefore classically there are two distinct uses of the word entropy: (i) obtained by performing macroscopic thermodynamic measurements, which we will call the thermodynamic entropy, and (ii) a measure of the lack of available information about a system's state.

Even at the classical level, there are very few systems where ergodicity can be proved and the quantum mechanical situation is far less clear. Knowledge of the system through observation does alter the state of the system, so that it is not apparent if the thermodynamic entropy is altered by this process. For a system with density matrix ρ, the von Neumann entropy $S_{\mathrm{VN}} = -\mathrm{Tr}(\rho \ln \rho)$, which is identically *zero* for any system in a pure state, including an energy eigenstate. Because the state of the system is known completely, this reduces its entropy to zero as in case (ii). But what does this do to entropy (i)? If we do thermodynamic measurements on systems in pure states, that is, where the state is known completely to the experimenter, does this alter the *thermodynamic entropy* that is measured?

Furthermore, the quantum ensemble definition of entropy, case (i), is simply related to the density of states of the energy. But if the system is in a single energy eigenstate, it might appear to be impossible to get the density of states. This would argue that thermodynamic entropy measured for systems in pure states would be different than for a system coupled to an external environment. Such questions until recently were purely theoretical but now with the emergence of experiments that study the thermalization of isolated quantum systems or lack thereof [3,4], a proper understanding of the microscopic origins of entropy has become increasingly important.

Here we find convincing evidence that for pure states, there is a way of defining an entropy as in case (i) for a nonintegrable quantum mechanical system with a large number of degrees of freedoms, which is identical to the thermodynamic definition, and give an explicit prescription for its construction from a knowledge of the eigenstate alone, thus successfully linking the thermodynamic entropy with its underlying microscopic origins. Throughout this work we will be concerned with the thermodynamic limit of a large number of degrees of freedom N and will not consider nonextensive corrections to the entropy for finite systems. We will show that despite the limited sizes attainable numerically, it is possible to probe this large N limit. Furthermore in this limit, the microcanonical and canonical ensembles give identical results [1]. Therefore we can freely choose which ensemble to consider. In both cases the thermodynamic entropy will coincide with the statistical mechanical definition of entropy measured in either ensemble.

II. EIGENSTATE THERMALIZATION HYPOTHESIS AND ITS EXTENSION

Statistical mechanics is concerned mainly with computing the time average of observables that depend on only a few variables, and relates this time average to an average over an ensemble. Thus for a system with average energy $\langle E \rangle$, statistical mechanics posits the relation between the time average of some observable $\hat{O}$ and its average over the microcanonical ensemble of states, the latter being far easier to compute

$$\overline{\langle \hat{O} \rangle} = \mathrm{Tr}(\rho_{\mathrm{micro},\langle E \rangle} \hat{O}), \tag{1}$$

where the bar on the left-hand side denotes a time average and the microcanonical density matrix at energy E_0 is

$$\rho_{\text{micro},E_0} \equiv \frac{1}{\mathcal{N}_{\text{states}}} \sum_{E_0 < E < E_0 + \Delta E} |E\rangle\langle E|. \quad (2)$$

Here ΔE is much greater than the average distance between neighboring energy levels, but much less than the macroscopic energy scale E_0, and $\mathcal{N}_{\text{states}}$ is the number of terms in the sum. In quantum mechanics, the fluctuation in the energy can be large [5], but in practice is taken to be small so that $\langle E \rangle$ is a good measure of the energy. The way to see the connection between the underlying quantum evolution needed on the left-hand side of Eq. (1) of a generic system and the ensemble methods of statistical mechanics has recently become much better understood. The idea is the "eigenstate thermalization hypothesis" (ETH) [5–7]: for large N, the expectation value of an observable in an energy eigenstate becomes equal to the microcanonical average at the same energy, that is,

$$\text{Tr}(\rho_E \hat{O}) = \text{Tr}(\rho_{\text{micro},E} \hat{O}), \quad (3)$$

where ρ is the density matrix for the wave function $|E\rangle$ at energy E, $\rho_E = |E\rangle\langle E|$. Put more simply, the expectation value of $\hat{O}$ will vary very little between neighboring energy levels for large N, implying that the expectation value of $\hat{O}$ in any energy eigenstate is the microcanonical average (at that energy). There is strong numerical evidence and analytic arguments to support this for a variety of systems [5–10]. Not all systems obey this hypothesis [11,12]. In particular, integrable systems [13] do not do so. Note that $\rho_E \neq \rho_{\text{micro},E}$, as there are a large number of constants of motion, even for generic Hamiltonians, such as projection operators [14] that violate Eq. (3). This is why the choice of observables satisfying this equation is restricted.

Below we shall extend these ideas to understanding entropy. In the thermodynamic limit statistical mechanics posits the formula for the thermodynamic entropy $S_{\text{thermo}} = -\text{Tr}(\rho_{\text{micro}} \ln \rho_{\text{micro}})$ where ρ_{micro} could be equivalently replaced with the thermal density matrix in the canonical ensemble in the limit that we are considering. We want to see if this can be calculated from the properties of the wave function itself without any recourse to ensembles.

We consider homogeneous systems for simplicity with short-range interactions and start by following the standard textbook scenario: the system is divided into two parts, the larger part B can act as a bath in contact with the smaller one A that is our system of interest. If the complete system starts out in a pure state, then by doing a partial trace over B, the reduced density matrix for A, $\rho_A = \text{Tr}_B \rho$ becomes mixed because it is entangled with B. This is part of a common but nonrigorous argument for how a canonical distribution of energies emerge for the smaller system A, where the temperature T is given by the statistical mechanical relation between T and the average energy. If the complete system was actually integrable this argument fails and thermalization does not occur.

The *entanglement entropy* [15] is defined as

$$S_{\text{Ent}}(A,B) = -\text{Tr}(\rho_A \ln \rho_A). \quad (4)$$

In what follows we shall show how to relate $S_{\text{Ent}}(A,B)$, which is obtained for the system $A + B$ in isolation, to the statistical mechanical entropy of the subsystem A. Using the extensivity of entropy, we can use this relation to obtain the total statistical mechanical (and hence thermodynamic) entropy.

If instead we considered the complete system to already be described by the microcanonical density matrix, then $\rho_{A,\text{micro}} \equiv \text{Tr}_B \rho_{\text{micro}}$ must behave with canonical statistics even for integrable systems, and the entanglement entropy in that case is $S_{\text{Ent,micro}}(A,B) \equiv -\text{Tr}(\rho_{A,\text{micro}} \ln \rho_{A,\text{micro}})$, which is the statistical mechanical expression for the entropy. This usual microcanonical, or equivalently canonical, entropy is what one measures experimentally in thermodynamic experiments. Thus

$$S_{\text{thermo}}(A) = S_{\text{Ent,micro}}(A,B) = -\text{Tr}(\rho_{A,\text{micro}} \ln \rho_{A,\text{micro}}). \quad (5)$$

This argument is very similar to most treatments of entropy in textbooks [16], however, there is more evidence than this to support it. Rigorous analysis [17] gives an canonical entanglement entropy for one-dimensional models in agreement with the thermodynamic results [18].

This *does not* connect thermodynamic entropy with the wave function of a system because the right-hand side of Eq. (5) involves the microcanonical average and is similar to the right-hand side of the ETH as written in Eq. (3), but differs because of the nonlinear logarithmic factor which is not an observable operator. Thus we ask whether we can extend the ETH to quantities of this form; that is, for a generic quantum system in an energy eigenstate can we replace $\rho_{A,\text{micro}}$ with ρ_A? In other words is

$$S_{\text{thermo}}(A) = -\text{Tr}(\rho_A \ln \rho_A)? \quad (6)$$

This form was recently hypothesized using other less direct analytical arguments [19]. Although this is the thermodynamic entropy only for a subsystem of $A + B$, the entropy is extensive so we can obtain the entropy for the full system by adding the entropies for individual subsystems together. In the particular case of a homogeneous system, the above equation can easily give us the thermodynamic entropy per unit volume. This can be extended to a heterogeneous system by dividing it into many parts, for example, a division into three, A, B, C. The properties of the system in isolation would give the statistical mechanical (and thus thermodynamic) entropy as $S_{\text{Ent}}(A,BC) + S_{\text{Ent}}(B,CA) + S_{\text{Ent}}(C,AB)$. Thus entropy defined this way is a measure of *self-entanglement*.

III. NUMERICAL EXACT DIAGONALIZATION

Now we employ exact diagonalization on a number of systems to determine if Eq. (6) is satisfied, and how it scales with system size. Because system sizes are still far from the thermodynamic limit, finite size effects are prohibitively large. However, instead of using extensivity to directly compute the left-hand side of Eq. (6) from the full thermodynamic entropy $S_{\text{thermo}}(A + B)$, we can minimize finite size effects by asking, equivalently, using Eq. (5), does

$$-\text{Tr}(\rho_A \ln \rho_A) = -\text{Tr}(\rho_{A,\text{micro}} \ln \rho_{A,\text{micro}})? \quad (7)$$

In other words, although the thermodynamic limits of these two approaches are equivalent, finite size effects are much smaller in the formulation of the second approach. Qualitatively, the

finite size corrections with this approach are related to the size of the Hilbert space, not the number of particles in the system, which is also why numerical work on ETH [5,7] also works so well.

We study systems that have been previously shown to give rise to energy eigenstate thermalization. Following some of these previous studies [8–10] we consider spinless fermions and hard core bosons on a one-dimensional lattice with nearest neighbor (NN) and next nearest neighbor (NNN) hopping, and interaction. The hopping strengths for NN and NNN are t and t', likewise we denote the interaction strengths V and V'. We study these systems for parameters where they are known to obey ETH and for integrable parameters where they do not. Following previous work, we use periodic boundary conditions in the subspace with wave vector $k = 1$, rather than $k = 0$ to avoid a parity symmetry. Throughout this work we set the energy scale to have $t = V = \hbar = 1$.

For both the fermions and hard core boson cases, $t' = V' = 0.96$ which is nonintegrable and should obey the ETH. We calculate the entanglement entropy for each energy eigenstate for a range of different bath and system sizes. Our largest size was $N = 27$ lattice sites with $N_p = 6$ particles. We trace over the bath sites and calculate the entanglement entropy Eq. (4), and do this as a function of the number of lattice sites of A, denoted m. For these cases, we plotted the entanglement entropy as a function of m. As is well known [20] $S_{\text{ent}}(m) = S_{\text{ent}}(N - m)$ so for $m > N/2$ the entanglement entropy must go back down to zero. As we display in the supplementary information [21] in the case of hard core bosons, it is very close to linear behavior over a substantial range of m, for the nonintegrable case, as is expected due to the extensive nature of entropy.

However, the behavior for integrable systems $t' = V' = 0$ is more erratic and the linearity depends much more strongly on the eigenvector. Similar results were also obtained for the fermion model as well as is shown in the supplementary material.

We directly checked to see if Eq. (7) is supported numerically. We calculated the left-hand side of this equation for different values of m and because of the linearity of the entanglement entropy with m, the results are insensitive to this value. In Fig. 1 it is shown for $m = 4$. We also calculate the microcanonical reduced density matrix by taking the partial trace of Eq. (2). We average the reduced density matrix over 100 neighboring eigenvectors, and use that to calculate the entanglement entropy as done on the right-hand side of Eq. (7). The result for 6 hard code bosons on 27 sites is shown in Fig. 1(a). The solid line is no microcanonical averaging ($n = 1$) that has been shifted down by 0.02 because it coincides very well with the other plots in this graph. The circles show the microcanonical average over $n = 100$ neighbors. The two data sets are very close to each other in most of the range where the density of states is high and one has eigenstate thermalization. The square points show the entanglement entropy with no microcanonical averaging (as in the solid curve) and average this over 100 neighboring energy levels. The "$+$" plot will be discussed below. The value of 100 neighbors was used because the density of energy levels is more than several hundred per unit energy, over the range shown, and so the energy levels in the averages are indeed close together. We have verified
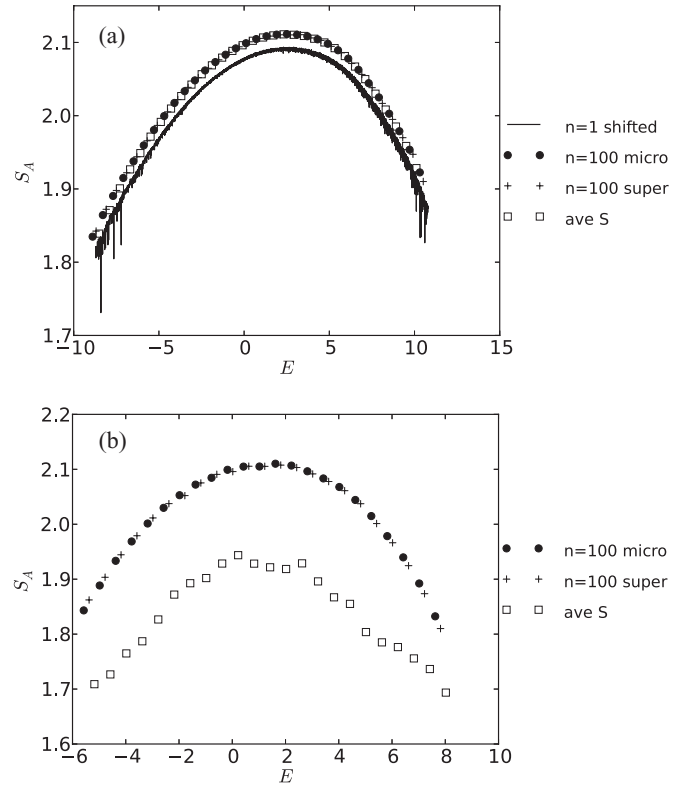

FIG. 1. (a) The entanglement entropy $S_{\text{Ent}}(A,B)$ [Eq. (4)], for a nonintegrable system of 6 hard core bosons on 27 sites where A has 4 sites and B has 23. The solid line shows the entanglement entropy of individual energy eigenvectors $n = 1$, *which has been shifted by* 0.02 *from the other plots* for clarity. The circles show the microcanonical entropy obtained by averaging the density matrix over $n = 100$ neighboring eigenvectors. The $+$ symbols show the entanglement entropy of a wave function that is a random superposition of 100 neighboring levels. The squares show the entanglement entropy for single eigenstates ($n = 1$) averaged over 100 neighboring energy levels. (b) The same as in (a) but for an integrable system $t' = V' = 0$. The solid line has been omitted because the variance in these values is very large, see Fig. 2.

that using a fixed energy window instead of a fixed number of neighbors leads to very similar results.

In contrast, in Fig. 1(b), for the integrable case $t' = V' = 0$, the fluctuations are much larger and that the microcanonical average is much further from the entanglement entropy for individual eigenstates. The solid curve is not shown because it fluctuates rapidly over a very wide range, resulting in much lower square points that deviate significantly from the other plots. Similar results are seen in the one-dimensional spinless fermion system. To quantify this difference further, we examine the standard deviation of the fluctuation of the entanglement entropy for both the integrable and nonintegrable cases.

We calculated the standard deviation $\sigma(S)$ of the entanglement entropy $S_{\text{ent}}(A,B)$ around the maximum of the curves in Fig. 1 for different lattice sizes, 16, 18, 20, 25, and 27, all with six particles in Fig. 2. $\sigma(S)$ is computed over 100 neighboring energy levels, and this is shown as a function of the density of states on a log-log plot for both the hard core boson models (B) and spinless fermions (F), and for both the nonintegrable
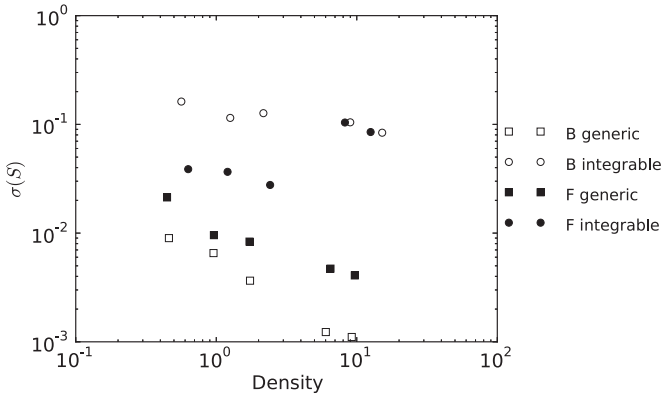
FIG. 2. The fluctuations in the entanglement entropy for both nonintegrable (generic) and integrable systems of six hard core bosons (B) but with a different number of lattice sites, and the same quantity for the spinless fermion model (F). The standard deviation of the entanglement entropy $S_{ent}(A, B)$ is plotted against the density of states. This is done around the point of maximum entanglement entropy.

and integrable cases. As is apparent, the integrable fluctuations are substantially higher. This is also the case for observables as has been shown previously [7]. The behavior as a function of the density of states and system size is qualitatively different. The fluctuations are diminishing much more rapidly in the nonintegrable case. This is similar to the slower diminution of fluctuations of observables predicted for integrable versus nonintegrable systems, where in the first case we expect a power law decrease as a function of the number of degrees of freedom and in the second case it is predicted to be an exponential [5].

The above studies provide excellent numerical evidence for the equivalence of the thermodynamic entropy of non-integrable systems with the entanglement entropy when the wave function is in an energy eigenstate. We also studied two-dimensional hard core bosons and fermions and reached the same conclusion (see supplementary materials [21]). Now we ask what happens with more general initial states. If one starts the system off in a state with fluctuations in energy in a window, then we can ask how this evolves for long times. We expect the thermalization of observables, and that the entropy should also be the thermodynamic entropy. We can test to see if this is the case numerically.

Over long times, the energy eigenvector components of nonintegrable wave functions will have random phases [22].

Therefore we test this case to see if the entanglement entropy still looks like the microcanonical entanglement entropy. In Fig. 1(a) the $+$ line computes the entanglement entropy at each energy by using wave functions that are the superposition of 100 neighboring energy eigenvectors with Gaussian amplitudes and random phases. As can be seen, this matches the predictions of the microcanonical entanglement entropy very well, and the same conclusion is also reached in the fermion case. The closeness of the reduced density matrix to the canonical result is expected for general initial conditions where many energy eigenstates are present [23,24]. Our above analysis shows that for generic systems, when the system is put in an energy eigenstate, we still obtain the microcanonical entropy, which is crucial, as otherwise thermodynamics would fail in this important case. Therefore the equivalence of thermodynamic and the entanglement entropy for long times should be correct, starting from a wide range of initial states.

IV. CONCLUSIONS

We can now answer the questions that we originally posed. Although the statistical mechanical entropy is most simply understood from the density of states, for a generic Hamiltonian, the self-entanglement of a single energy eigenstate can be used to obtain the same result as well. Knowledge of neighboring levels is not necessary as this entropy as we have seen, is a property of a single energy energy eigenvector. Going back to our earlier comparison with classical physics, we can now answer our original question. A complete knowledge of the system's quantum state does not affect its behavior with respect to macroscopic measurements of the entropy. An experimenter who measured a systems's state precisely, will in subsequent measurements obtain the same answers as someone who is not made privy to this information, despite the latter describing the system in a mixed state. For long times, a system in a pure state and one in an statistical ensemble have identical thermodynamic entropies in the limit of large systems. This is because the entropy in such experiments measures the system's self-entanglement, not the lack of knowledge of it.

Note added. After completion of this work [28], related work on weak and strong typicality was published [25].

[1] S.-K. Ma, *Statistical Mechanics* (World Scientific, Singapore, 1985).

[2] R. Balian, in *Poincaré Seminar 2003* (Springer, New York, 2004), pp. 119–144.

[3] T. Kinoshita, T. Wenger, and D. S. Weiss, Nature (London) **440**, 900 (2006).

[4] S. Hofferberth, I. Lesanovsky, B. Fischer, T. Schumm, and J. Schmiedmayer, Nature (London) **449**, 324 (2007).

[5] J. M. Deutsch, Phys. Rev. A **43**, 2046 (1991).

[6] M. Srednicki, Phys. Rev. E **50**, 888 (1994).

[7] M. Rigol, V. Dunjko, and M. Olshanii, Nature (London) **452**, 854 (2008).

[8] L. F. Santos and M. Rigol, Phys. Rev. E **81**, 036206 (2010).

[9] M. Rigol, Phys. Rev. Lett. **103**, 100403 (2009).

[10] M. Rigol, Phys. Rev. A **80**, 053607 (2009).

[11] D. M. Basko, I. L. Aleiner, and B. L. Altshuler, Phys. Rev. B **76**, 052203 (2007).

[12] V. Oganesyan and D. A. Huse, Phys. Rev. B **75**, 155111 (2007).

[13] The definition of integrability in quantum systems is problematic. See, for example, Refs. [26,27].

[14] B. Sutherland, *Beautiful models: 70 years of exactly solved quantum many-body problems* (World Scientific, Singapore, 2004).

[15] C. H. Bennett, Phys. Scr. **1998**, 210 (1998).

[16] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Butterworth-Heinemann, London, 2002).

[17] P. Calabrese and J. Cardy, J. Phys. A: Mathematical and Theoretical **42**, 504005 (2009).

[18] V. E. Korepin, Phys. Rev. Lett. **92**, 096402 (2004).

[19] J. Deutsch, New J. Phys. **12**, 075021 (2010).

[20] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).

[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.87.042135 for ancillary work.

[22] M. Srednicki, J. Phys. A: Math. Gen. **32**, 1163 (1999).

[23] N. Linden, S. Popescu, A. J. Short, and A. Winter, Phys. Rev. E **79**, 061103 (2009).

[24] A. J. Short and T. C. Farrelly, New J. Phys. **14**, 013063 (2012).

[25] L. F. Santos, A. Polkovnikov, and M. Rigol, Phys. Rev. E **86**, 010102 (2012).

[26] J.-S. Caux and J. Mossel, J. Stat. Mech.: Theory Exp. (2011) P02023.

[27] E. A. Yuzbashyan and B. S. Shastry, J. Stat. Phys. **150**, 704 (2013).

[28] J. M. Deutsch, H. Li, and A. Sharma, arXiv:1202.2403.