Typical entanglement entropy in the presence of a center: Page curve and its variance

Eugenio Bianchi\textsuperscript{1, 2} and Pietro Donà\textsuperscript{1, 2}

\textsuperscript{1}Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, Pennsylvania 16802, USA
\textsuperscript{2}Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

In a quantum system in a pure state, a subsystem generally has a nonzero entropy because of entanglement with the rest of the system. Is the average entanglement entropy of pure states also the typical entropy of the subsystem? We present a method to compute the exact formula of the entanglement entropy probability distribution $P(S_A)\,dS_A$ that a subsystem has entanglement entropy $S_A$. The method applies to subsystems defined by a subalgebra with a trivial center. In the case of a trivial center, we reobtain the well-known result for the average entropy and the formula for the variance. In the presence of a nontrivial center, the Hilbert space does not have a tensor product structure and the well-known formula does not apply. We present the exact formula for the average entanglement entropy and its variance in the presence of a center. We show that for large systems the variance is small, $\Delta S_A/\langle S_A \rangle \ll 1$, and therefore the average entanglement entropy is typical. We compare exact and numerical results for the probability distribution and comment on the relation to previous results on concentration of measure bounds. We discuss the application to physical systems where a center arises. In particular, for a system of noninteracting spins in a magnetic field and for a free quantum field, we show how the thermal entropy arises as the typical entanglement entropy of energy eigenstates.

\textbf{Introduction.---}In a seminal paper \cite{Page1993}, Page showed that, when an isolated quantum system is in a random pure state, the average entropy of a subsystem is close to maximal. This result plays a central role in the analysis of the black hole information puzzle \cite{Srednicki1993, Susskind1995, Susskind1995b, Susskind1995c, Schwinger1995, Page1995, Page1996, Page1997}, in the quantum foundations of statistical mechanics \cite{Bennett1988, Deutsch1993, Deutsch1993b, Deutsch1993c}, in quantum information theory \cite{Hayashi2004, Horodecki2003, Horodecki2003b, Horodecki2003c, Horodecki2003d} and in the study of the quantum nature of spacetime geometry \cite{Deutsch1991, Bousso1999, Bousso2000, Bousso2000b, Bousso2000c, Bousso2000d}. The entanglement entropy, expressed as a function of the subsystem size, is often called the \textit{Page curve} (Fig. 1).

In a quantum system in a pure state, a subsystem $A$ generally has a nonzero entropy because of entanglement with the rest of the system. In this paper we address the question: \textit{Is the average entanglement entropy of pure states $\langle S_A \rangle$ also the typical entropy of the subsystem?} To illustrate the significance of this question, let us consider for instance the gas in a room held at fixed temperature. The canonical ensemble allows us to compute the average energy of the gas. However, the configuration of molecules in one room is one realization of this ensemble—we are not averaging over rooms. How close to the average is the energy of this realization? In other words, is the average energy typical? In statistical mechanics, we answer this question by computing the variance $\langle \Delta E \rangle^2$. In the canonical ensemble, we find $\langle \Delta E \rangle^2 \ll 1$ and we conclude that the average energy is typical. Here we investigate the typicality of the entanglement entropy $S_A$ of a subsystem by studying its average and variance.

A well-studied special case of subsystem $A$ is described by a subalgebra with a trivial center. In this case the Hilbert space of the system is simply given by a tensor product of the subsystem Hilbert space $\mathcal{H}_A$ and its complement $\mathcal{H}_B$, i.e. $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. This case was originally considered by Page \cite{Page1993, Page1995b, Page1995c, Page1995d, Page1995e, Page1995f} who computed the average in $\mathcal{H}$ of the entanglement entropy. The variance was computed recently in \cite{Staudacher2011, Staudacher2012}. In this paper we present an algorithm that reproduces the average and variance, together with the skewness and higher moments of the entanglement entropy probability distribution $P(S_A)\,dS_A$ for a trivial center.

In many physically relevant cases, the Hilbert space over which we average the entanglement entropy does not have a tensor product structure, $\mathcal{H} \neq \mathcal{H}_A \otimes \mathcal{H}_B$. In fact, in general, the subalgebra that identifies the subsystem $A$ has a nontrivial center. As a result, the Hilbert space of the system is a direct sum of tensor products \cite{Bueno2011, Bueno2012}. In this work we extend the typicality results on the entanglement entropy in the presence of a center. Ex-
amps of subsystems with a nontrivial center are subsets of a spin chain with fixed boundary conditions; lattice fermions with fixed boundary conditions [48]; compact scalar fields, Abelian and non-Abelian gauge fields on a lattice [46]; loop quantum gravity [35, 36, 49, 50].

A simple example of subsystems with a center is provided by the energy eigenspace in a noninteracting system with Hamiltonian $\hat{H} = \hat{H}_A + \hat{H}_B$. In this case the eigenspace $\mathcal{H}(E) \subset \mathcal{H}$ has the structure of a direct sum of tensor products

$$\mathcal{H}(E) = \bigoplus_{j=1}^d \left( \mathcal{H}_A(\varepsilon_j) \otimes \mathcal{H}_B(E - \varepsilon_j) \right),$$

where $\mathcal{H}_A(\varepsilon_j)$ and $\mathcal{H}_B(\varepsilon_k)$ are eigenspaces of given energy for the subsystems $A$ and $B$. This structure is relevant for the study of thermal properties of isolated quantum systems. See Fig. 1 for an example.

Building on [1, 38–43], we develop methods to compute the exact formula for the average entanglement entropy and the moments $m_n(\lambda) = \langle S^a \rangle$ for all pure states in the Hilbert space $\mathcal{H}$ with Hamiltonian $\hat{H}$ and the moments $m_n(\lambda) = \langle S^a \rangle$ for all pure states in the Hilbert space $\mathcal{H}(E)$ with Hamiltonian $\hat{H}(E)$ with a nontrivial center. We compare our results to concentration of measure bounds [24].

By studying the average entanglement entropy $\langle S_A \rangle$ and the variance

$$\langle (\Delta S_A)^2 \rangle = \langle S_A^2 \rangle - \langle S_A \rangle^2,$$  (2)

we show that, both in the case of a trivial and a nontrivial center, for large systems the average entanglement entropy of a subsystem is also typical.

After presenting a detailed derivation of the Page curve and its variance, we discuss the application of our results to a model system where a nontrivial center arises: we determine the Page curve of energy eigenstates of a parafermions with fixed boundary conditions [48]; compact Abelian and non-Abelian gauge fields on a lattice [46]; loop quantum gravity [35, 36, 49, 50].

Average entropy and variance with a trivial center.— Let us consider a quantum system with a bipartite Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. The subsystems $A$ and $B$ have dimension $d_A = \dim \mathcal{H}_A$ and $d_B = \dim \mathcal{H}_B$, with $A$ the smaller of the two subsystems, $1 < d_A \leq d_B$, and the dimension of $\mathcal{H}$ given by $d = d_Ad_B$. The restriction of a pure state $|\psi\rangle \in \mathcal{H}$ to the subsystem $A$ defines the reduced density matrix $\rho_A = \text{Tr}_B|\psi\rangle\langle\psi|$. The entanglement entropy $S_A(\psi) = -\text{Tr}(\rho_A \log \rho_A)$ is the von Neumann entropy of the reduced density matrix.

A random state $|\psi\rangle$ in the Hilbert space $\mathcal{H}$ can be generated by choosing an orthonormal basis $|n\rangle$ with $n = 1, \ldots, d$ and picking a vector $|\psi\rangle = \sum_n \psi_n |n\rangle$ at random with respect to the uniform measure $d\mu(\psi) = Z^{-1} \delta(|\psi|^2 - 1) d\psi/d\mathbf{v}$ on the unit sphere in $\mathbb{C}^d$, with the constant $Z$ defined so that the measure is normalized to unity, $\int d\mu(\psi) = 1$. The average of a function $f$ over all states $|\psi\rangle$ can be computed by integrating uniformly over states, $\langle f \rangle = \int f(\psi) d\mu(\psi)$. As the entanglement entropy $S_A$ depends on the reduced density matrix $\rho_A$ only via its eigenvalues $\lambda_i$, (the entanglement spectrum, with $a = 1, \ldots, d_A$), to compute the average over $|\psi\rangle$ we need only the induced measure over the eigenvalues $\lambda_i$. This measure $d\mu(\lambda_1, \ldots, \lambda_{d_A})$ was computed by Lloyd and Pagels in [39], (see also App. A).

Computing the average entanglement entropy $\langle S_A \rangle$ and its variance $\Delta S_A$ is not immediate because they are not polynomial functions of $\rho_A$. Here we follow a strategy that generalizes to the computation of higher moments of $S_A$. We first consider the average of the function $\langle \text{Tr}_A(\rho_A^k) \rangle$ with $r \geq 0$,

$$\langle \text{Tr}(\rho_A^k) \rangle = \int (\sum_{\lambda=1}^{d_A} \lambda_r) d\mu(\lambda_1, \ldots, \lambda_{d_A}).$$  (3)

To compute the integral over $\lambda_a$ it is useful to introduce the quantity $X_{ij}(r)$, defined as an integral of generalized Laguerre polynomials $L^{(d_b-d_A)}_j(q)$ with $i = 0, 1, \ldots, d_A - 1$:

$$X_{ij}(r) = \int_0^\infty q^r L^{(d_b-d_A)}_j(q) L^{(d_b-d_A)}_j(q) q^{d_b-d_A e^{-q}} dq \Gamma(i+1)\Gamma(d_b-d_A+i+1) \Gamma(i+1)\Gamma(d_b-d_A+i+1)$$

$$= \sum_{p=0}^{d_A-1} \frac{\Gamma(d_b-d_A+r+1+p)\Gamma(j+1)\Gamma(r+j+1)\Gamma(j+1)\Gamma(r+j+1)}{\Gamma(1)\Gamma(j+1)\Gamma(r+j+1)\Gamma(r+j+1)}$$  (4)

Computing the integral (4) is nontrivial. We obtained the result by using the generating function for the Laguerre polynomials, as we describe in more detail in App. A. Formula (4) for $X_{ij}(r)$ provides the main technical tool for our derivation.

The average (3) takes a simple form when expressed in terms of the matrix $X_{ij}(r)$,

$$\langle \text{Tr}(\rho_A^k) \rangle = \frac{\Gamma(d_b d_A)}{\Gamma(d_b d_A + r)} \text{Tr} X(r),$$  (5)

where we treat $X_{ij}(r)$ as a $d_A \times d_A$ matrix $X(r)$. We note that (5) is a smooth continuous function of $r$. The average entanglement entropy can be obtained immediately by taking a derivative with respect to $r$,

$$\langle S_A \rangle = -\lim_{r\to 1} \langle \text{Tr} \rho_A \log \rho_A \rangle = -\lim_{r\to 1} \partial_r \langle \text{Tr}(\rho_A^k) \rangle.$$  (6)

The result of this computation, expressed in terms of the digamma function $\Psi(x) = \Gamma'(x)/\Gamma(x)$ (i.e., the logarithmic derivative of the gamma function), is the formula
\[ \langle S_A \rangle = \Psi(d_A d_B + 1) - \Psi(d_B + 1) - \frac{d_A - 1}{2d_B}. \]  
(7)

This formula was first conjectured by Page [1] and then proved with different techniques here and in [40–43].

The technique described above applies directly to the calculation of the average of \( S_A^2 \). The strategy is to first consider the average of \( \langle \text{Tr}(\rho_A^r \rho_A^s) \rangle \) and express it in terms of matrices \( X(r) \):

\[ \langle \text{Tr}(\rho_A^r) \text{Tr}(\rho_A^s) \rangle = \frac{\Gamma(d_Ad_B)}{\Gamma(d_A d_B + r_1 + r_2)} \times \]  
\[ \times ( \text{Tr}(X(r_1 + r_2) + \text{Tr}(X(r_1) \text{Tr}(X(r_2) - \text{Tr}(X(r_1) X(r_2)))). \] 

(8)

Its derivatives with respect to \( r_1 \) and \( r_2 \) can again be expressed in terms of derivatives of the gamma function, and simplified with the help of Wolfram’s Mathematica. The average of \( S_A^2 \) is obtained as

\[ \langle S_A^2 \rangle = \lim_{r_1, r_2 \to 1} \partial_{r_1} \partial_{r_2} \langle \text{Tr}(\rho_A^r \text{Tr}(\rho_A^s)) \rangle. \]  
(9)

Using the definition (2) and the formula (7), we obtain the exact formula:

\[ (\Delta S_A)^2 = \frac{d_A + d_B}{d_B d_A + 1} \Psi'(d_B + 1) + \]  
\[ - \Psi'(d_A d_B + 1) - \frac{(d_A - 1)(d_A + 2d_B - 1)}{4d_B (d_A d_B + 1)}, \] 

(10)

discussed also in [44, 45]. Using the same technique we can determine also the average of higher powers of \( S_A \). In particular, we report the exact formula for the third moment \( m_3 \) of the entanglement entropy distribution (App. A)

\[ m_3 = \Psi''(d_A d_B + 1) - \frac{d_A^2 + 3d_A d_B + d_B^2 + 1}{(d_A d_B + 1)^2 (d_A d_B + 2)} \Psi'(d_B + 1) + \]  
\[ + \frac{(d_A - 1)(d_A + 2d_B - 1)}{4d_B (d_A d_B + 1)^2 (d_A d_B + 2)} \Psi'(d_B + 1) + \]  
\[ - \frac{(d_A - 1)(2d_A^2 + 3d_A d_B + d_B^2 - 4d_A d_B + 4d_A d_B^2 - 3d_A d_B)}{4d_B (d_A d_B + 1)^2 (d_A d_B + 2)} - \]  
\[ - \frac{(d_A - 1)(2d_A^2 + 8d_A d_B + 10d_B^2 - 4d_A - 6d_B + 2)}{4d_B (d_A d_B + 1)^2 (d_A d_B + 2)}. \]  
(11)

We observe that the formulas for the moments \( m_n \) are exact: they provide the mean (7), the variance (10) and the skewness (11) of the entropy for small systems, as well as for large systems. In the latter case, a Taylor series in \( 1/d_B \) provides asymptotic expressions for large systems and any subsystem: the average entropy is approximated by the expression

\[ \langle S_A \rangle \approx \log d_A - \frac{1}{d_A d_B} \frac{d_A^2 - 1}{2} \quad \text{for} \quad d_B \gg 1. \]  
(12)

As the entropy of \( A \) can be at most \( S_{\text{max}} = \log d_A \), this formula shows that for a large system the average entropy of a subsystem is close to maximal. The asymptotic expression for the variance of \( S_A \) is:

\[ \Delta S_A \approx \frac{1}{d_A d_B} \sqrt{\frac{d_A^2 - 1}{2}} \quad \text{for} \quad d_B \gg 1. \]  
(13)

We observe that the variance vanishes in the limit of \( d_B \to \infty \). Therefore we conclude that, in a large system, any subsystem has \( S_A \langle \langle S_A \rangle \rangle \ll 1 \). In other words, in a large system, the average entropy of any subsystem is also its typical entropy. We report also the asymptotic expression for the third moment \( m_3 \) (11)

\[ m_3 \approx -\frac{d_A^2 - 1}{d_A d_B^2} \quad \text{for} \quad d_B \gg 1. \]  
(14)

This formula shows that the skewness \( m_3/\sigma^3 \approx -\sqrt{8/(d_A^2 - 1)} \) is negative, which results in a right tilt in the distribution that does not vanish as \( d_B \to \infty \).

One might expect that these asymptotic formulas can be obtained in a simpler way, for instance by first expanding the entropy \( S_A \) around the maximally mixed state and then taking the average. The expansion in \( \Delta p_A = p_A - 1/d_A \) was first proposed by Lubkin and Lukin in [38] and is commonly found in reviews [7]. However, it was shown by Dyer in [43] that the series in \( \Delta p_A \) for \( \langle S_A \rangle \) does not converge. This has the consequence that, truncating the expansion in \( \Delta p_A \), only the first order in Page’s formula (12) is accidentally reproduced. Similarly, the leading order variance \( \Delta S_A \) (13) cannot be obtained by truncating the expansion in \( \Delta p_A \).

Determining the full probability distribution \( P(S_A) dS_A \) of a random pure state is not immediate. The methods introduced here allow us to determine its average \( \langle S_A \rangle \), its variance \( \sigma^2 = \langle \Delta S_A \rangle^2 \) and higher order moments \( m_n \) such as the skewness (14), (see also App. A). The normal distribution

\[ P_N(S_A) \ dS_A = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(S_A - \langle S_A \rangle)^2}{2\sigma^2}\right) \]  
(15)

is the distribution with the largest Shannon entropy at fixed average \( \langle S_A \rangle \) and variance \( \sigma^2 \). In Fig. 2 we compare this distribution to a numerical sample and find that it characterizes well the support of the probability distribution \( P(S_A) dS_A \) of random pure states. In particular, the numerical sample and the analytic formulas (12), (13) show that it is unlikely to find a state with maximum entropy \( S_{\text{max}} \). The numerical sample also shows a small right tilt with respect to the normal distribution, in accordance with the negative skewness (14).

Previous analysis of typicality have used concentration of measure techniques to provide upper bounds on the probability of finding a state with entropy lower than the average entropy [24]. In particular, using Levy’s lemma, Theorem III.3 in [24][51] states that the cumulative distribution \( P[S_A < \alpha] \) is bounded from above by a Gaussian function,

\[ P[S_A < \alpha] = \int_0^\alpha P(S_A) dS_A \leq \exp\left(-\frac{(\alpha - \langle S_A \rangle)^2}{2\sigma^2}\right) \]  
(16)
The algebra of the subsystem $A$ Hamiltonian $H_A$ is the commutant of a composite system with Hilbert space $\mathcal{H}_A \oplus \mathcal{H}_B$. Where the sum is over the spectrum of $Z$ subsystem subalgebra of observables $A$. We define a subsystem $A$ by choosing a subalgebra of observables $A_A$. The complement of the subsystem $A$ is denoted $B$ and its algebra of observables is the commutant of $A_A$, i.e., $A_B = \{ b \in A \mid [b, a] = 0 \ \forall a \in A_A \}$. The intersection of the two subalgebras, $A_A \cap A_B = \mathcal{Z}_A$, is called center of $A_A \subset A$. In the presence of a center, the Hilbert space of the system decomposes as a direct sum of tensor products $\mathcal{H} = \bigoplus_\zeta \left( \mathcal{H}_A(\zeta) \otimes \mathcal{H}_B(\zeta) \right)$, where the sum is over the spectrum of $\mathcal{Z}_A$.

Where we consider a composite system with Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ and Hamiltonian $H = H_A + H_B$ having energy eigenvalues $E_{jk} = \varepsilon_{jA} + \varepsilon_{kB}$. On the energy eigenspace $\mathcal{H}(E) \subset \mathcal{H}$ the algebra of the subsystem $A$ has a nontrivial center $\mathcal{Z}_A = H_A$. Therefore, $\mathcal{H}(E)$ has the structure (17),

$$\mathcal{H}(E) = \bigoplus_{j=1}^d \left( \mathcal{H}_A(\varepsilon_j) \otimes \mathcal{H}_B(E - \varepsilon_j) \right),$$

where the sum over $j$ is such that $\varepsilon_{kB} = E - \varepsilon_{jA}$.

Energy eigenspaces of the subsystem $A$ are denoted $\mathcal{H}_A(\varepsilon_j)$ and have dimension $d_{jA} = \dim \mathcal{H}_A(\varepsilon_j)$. Similarly for subsystem $B$. The energy eigenspaces of the system have then the direct sum structure $\mathcal{H}(E) = \oplus_j \mathcal{H}_j(E)$ where the sector $\mathcal{H}_j(E) = \mathcal{H}_A(\varepsilon_j) \otimes \mathcal{H}_B(E - \varepsilon_j)$ has definite energy in each subsystem. We denote $d_j = \dim \mathcal{H}_j(E)$ the dimension of each sector, with $d_j = d_{jA} d_{jB}$ and $d_E = \sum_j d_j$ the dimension of $\mathcal{H}(E)$.

Due to the direct-sum structure (18), Page’s formula (7) does not apply. Other instances of systems where the relevant Hilbert space has the form (1) are subsystems in lattice gauge theory [46], in spin chains and in lattice fermions with fixed boundary conditions [48], in loop quantum gravity [35, 36, 49, 50], and in general in presence of an additive constraint. The formulae that we derive below apply equally to all these cases.

To investigate typicality of the entropy in the energy eigenspace $\mathcal{H}(E)$, we determine the uniform measure over pure states belonging it. We note that a state $|\psi, E\rangle$ in $\mathcal{H}(E)$ can be written as a superposition $|\psi, E\rangle = \sum_j \sqrt{p_j} |\phi_j\rangle$ of normalized states $|\phi_j\rangle \in \mathcal{H}_j(E)$, with weights $p_j \geq 0$ satisfying the normalization condition $\sum_j p_j = 1$. The reduced density matrix for the subsystem $A$ is given by $\rho_A = \sum_j p_j \rho_{jA}$ with $\rho_{jA} = \text{Tr}_B |\phi_j\rangle \langle \phi_j|$. The entanglement entropy of the subsystem $S_A(\psi) = -\text{Tr}(\rho_A \log \rho_A)$, splits into the sum of two terms [52]:

$$S_A(\psi) = \sum_j p_j S_{jA}(\phi_j) - \sum_j p_j \log p_j,$$

where $S_{jA}(\phi_j) = -\text{Tr}(\rho_{jA} \log \rho_{jA})$ is the entanglement entropy in the sector $\mathcal{H}_j(E)$. We note that the entanglement entropy of $A$ is the $p_j$-weighted sum of the entanglement entropy $S_{jA}$ in each sector, plus the Shannon entropy of the weights $p_j$.

The uniform measure $d\mu_E(\psi)$ over pure states in $\mathcal{H}(E)$
decomposes as
\[ d\mu_E(\psi) = d\nu(p_1, \ldots, p_J) \prod_j d\mu(\phi_j), \]
where \( d\mu(\phi_j) \) is the uniform measure over pure states in each sector \( H_j(E) \). We derive the measure over the weights \( p_j \) in App. B and find
\[ d\nu(p_1, \ldots, p_J) = \frac{1}{Z} \delta(\sum_j p_j - 1) \prod_j (d^{d_j} - 1) d\nu_j. \]
This measure defines the Dirichlet distribution, also known as the multivariate beta distribution [53]. The constant \( Z \) is defined so that the measure is normal-
ized to unity: \( \int d\nu(p_1, \ldots, p_J) = 1 \). The average weight is \( \langle p_j \rangle = d_j/E \) and its second moments are \( \langle p_i p_j \rangle = (d_i d_j + \delta_{ij} d_i)/(d_i E(d_i E + 1)) \).

Using the technique described in (6), we write the average in \( H(E) \) of the entanglement entropy as
\[ \langle S_A \rangle_e = -\lim_{t \to 1} \sum_j \langle \partial_t (\sum_j p_j \text{Tr}(p_j^R A) + \sum_j p_j^j) \rangle, \]
and find the exact formula
\[ \langle S_A \rangle_e = \sum_j \frac{d_j}{E} \langle \langle S_j A \rangle + \Psi(d_j + 1) - \Psi(d_j) \rangle \], (23)
where \( \langle S_j A \rangle \) is given by (7). Using the same technique, we compute the exact formula for the average of \( S_A^2 \).

See App. B for a detailed derivation. In the limit of large dimension of each sector, \( d_j \gg 1 \), the average entanglement entropy takes the form
\[ \langle S_A \rangle_e = \sum_j \langle p_j \rangle \langle S_j A \rangle - \sum_j \langle p_j \rangle \log \langle p_j \rangle + O(1/E). \] (25)
In the same limit, the variance \( \langle \Delta S_A \rangle_e^2 = \langle S_A^2 \rangle_e - \langle S_A \rangle_e^2 \) goes to zero and the ratio of the two quantities scales as
\[ \frac{\langle \Delta S_A \rangle_e}{\langle S_A \rangle_e} \sim \frac{1}{\sqrt{E}}. \] (26)
Therefore the average entropy of a subsystem is also its typical entropy. Fig. 2 shows that the normal distribution with mean (23) and variance \( \langle \Delta S_A \rangle_e^2 \) characterizes well the support of the distribution of the entanglement entropy of random pure states of energy \( E \).

Discussion and applications.—We can now answer the question posed in the introduction. As, for large systems, the standard deviation \( \Delta S_A \) is much smaller than the average, we conclude that the average entropy \( \langle S_A \rangle \) is also the typical value of the entanglement entropy of a subsystem. The conclusion holds both in the presence of a trivial and of a nontrivial center.

The result applies to all systems where a subalgebra with a center arises. A notable case is the one of a free isolated quantum system prepared in an energy eigenstate. While interaction between subsystems is necessary for thermalization [10–23], an arbitrarily small interaction is sufficient to select a typical energy eigenstate. We illustrate the relevance of our result with two examples.

A system of \( N \) noninteracting spins with magnetic moment \( \mu \) in a magnetic field \( B \) has Hamiltonian \( H = \sum_i \mu_i B \sigma_i^z \). The energy eigenspace \( H(E) \) has dimension \( d_E = N!/(\binom{N/2}{n}\binom{N/2}{n}) \) with \( n = E/\mu B \). We consider a subsystem \( A \) consisting of \( N_A \) spins. The eigenspace \( \mathcal{H}(E) \) has the structure (1) where the direct sum is over eigenvalues \( \varepsilon \) of the Hamiltonian \( H_A \) of the subsystem \( A \). The average energy in \( A \) is \( \bar{E} = \langle \text{Tr}(H_A^R \rho_A) \rangle \), which shows equipartition of the energy in a typical state.

As a second example, we consider a free quantum field: the quantum electromagnetic field in a cubic box of volume \( L^3 \). Previous analysis have focused on the geometric entanglement entropy of a region of space [56–58] and its renormalization [8, 59–61]. Here we identify a different subalgebra of observables that better characterizes what can be measured [62]. The Hilbert space of the quantum field \( \mathcal{H} = \bigotimes_{A} \mathcal{H}_A \) is the tensor product of Hilbert spaces of discrete wavelength \( \lambda = (2L/k) / L/k \) with \( k_x, k_y, k_z \in \mathbb{N} \). A measuring device, such as an antenna of length \( \ell \), defines a subsystem \( A \) corresponding to the dis-
crete wavelengths $\lambda_A = (2\ell/k_x, 2\ell/k_y, 2\ell/k_z)$. If $\ell/L \leq 1$ is a rational number, the wavelengths $\lambda_A$ are a subset of the wavelengths $\lambda$. The Hilbert space $\mathcal{H}(E)$ of eigenstates with energy $E$ inherits the structure (1), where $\varepsilon$ is the energy of the antenna $A$. When the box is large, i.e. when $EL \gg \hbar c$, we can compute the dimension $d_E$ of $\mathcal{H}(E)$ from the density of states and the occupation numbers of photons. We find $d_E \approx \exp \left( \frac{4\sqrt{\pi}}{3(15)/4\pi} \left( \frac{E\ell}{\hbar c} \right)^{3/4} \right)$. Analogously, the dimension of $\mathcal{H}(\varepsilon)$ for $\varepsilon \ell \gg \hbar c$ is $d_A(\varepsilon) \approx \exp \left( \frac{4\sqrt{\pi}}{3(15)/4\pi} \left( \frac{\varepsilon \ell}{\hbar c} \right)^{3/4} \right)$.

The average energy in the subsystem $A$ is $\bar{E} = \sum_\varepsilon \varepsilon \langle p(\varepsilon) \rangle \approx E\ell^3/L^3$, which shows that in a typical state the energy of the subsystem is extensive and $E/L^3$ is the energy per unit volume. The average entanglement entropy of a subsystem with $\ell^3 \leq L^3/2$ is

$$\langle S_A \rangle \approx \log d_A(\varepsilon) \approx \frac{4\sqrt{\pi}}{3(15)/4\pi} \left( \frac{L}{\hbar c} \right)^{3/4}. \quad (29)$$

The result follows from (23) together with the fact that $\langle p(\varepsilon) \rangle$ is sharply peaked at $\bar{\varepsilon}$ and therefore the Shannon entropy contribution is negligible and the dominating term is the average entanglement entropy $\langle S_{AE} \rangle$. As the variance is exponentially small in $EL/\hbar c$, this is also the typical value of the entropy in an energy eigenstate.

For $\ell \ll L$ we can also define a temperature from entanglement:

$$kT = \left( \frac{\partial \langle S_A \rangle_E}{\partial \varepsilon} \right)^{-1} \approx \frac{1}{N} \left( \frac{3\pi}{2} \right)^{3/4} \left( \frac{\hbar c}{L^3} \right)^{1/4} \left( \frac{E_L}{L^3} \right)^{3/4}. \quad (30)$$

We note that this temperature does not depend on the subsystem size. In terms of this temperature, the entanglement entropy assumes the form of the familiar extensive formula for the canonical entropy of black body radiation. On the other hand, for large subsystems, the typical entanglement entropy is smaller than the thermal entropy and follows a Page curve qualitatively similar to the one in Fig. 1,

$$\langle S_A \rangle \approx \frac{4\pi^2}{3(4\pi)^{3/4}} \left( \frac{kT}{\hbar c} \right)^3 \min \left( \ell^3, L^3 - \ell^3 \right). \quad (31)$$

This entropy is arising from the entanglement between the modes that the antenna can measure, i.e. the wavelengths $\lambda_A$, and the modes that it cannot measure. The unmeasured modes include both longer wavelengths and wavelengths shorter than $\ell$ that the antenna cannot couple to.

Our results show that, in small noninteracting systems prepared in a typical energy eigenstate, thermal properties can arise from entanglement. Recent experimental developments on measurements of thermalization in small isolated quantum systems, such as ultracold atoms in optical lattices [63–66], might provide access to the deviations from statistical mechanics predicted by the exact formulas for the average entropy (23) and its variance (24) at fixed energy.

**Acknowledgments.**— The authors thank Neev Khera for discussions on the convergence of the $\delta \rho$ expansions, Abhay Ashtekar for discussions on the structure of direct-sum decompositions, Pierpaolo Vivo and Lu Wei for correspondence. E.B. is supported by the NSF Grant No. PHY-1806428. P.D. is supported by the NSF Grants No. PHY-1505411, No. PHY-1806356 and the Eberly research funds of Penn State.

**Appendix A: Detailed computation of the moments of the distribution $P(S_A)dS_A$**

Given a Hilbert space $\mathcal{H}$ of dimension $d$ and an orthonormal base $|n\rangle$ of $\mathcal{H}$, a normalized state $|\psi\rangle \in \mathcal{H}$ can be decomposed as $|\psi\rangle = \sum_{n=1}^d \psi_n |n\rangle$ with $\psi \in \mathbb{C}^d$ and $||\psi|| = \sum_{n=1}^d |\psi_n|^2 = 1$. The uniform measure on $\mathcal{H}$ is then the measure on the unit sphere on $\mathbb{C}^d$,

$$\text{d}\mu(\psi) = \frac{1}{Z} \delta \left( 1 - \sum_{n=1}^d |\psi_n|^2 \right) \prod_{n=1}^d \text{d}\psi_n \text{d}\overline{\psi}_n, \quad (A1)$$

where $Z$ is a normalization constant. The average over states in $\mathcal{H}$ of functions of the reduced density matrix of a subsystem $A$ [1] is computed more efficiently using the measure over the eigenvalues of the density matrix induced by $\text{d}\mu(\psi)$, first computed by Lloyd and Pagels in [39]:

$$\text{d}\mu(\psi) \rightarrow \text{d}\mu(\lambda_1, \ldots, \lambda_{d_A}) \equiv Z \delta \left( 1 - \sum_{i=1}^{d_A} \lambda_i \right) \Delta^2 (\lambda_1, \ldots, \lambda_{d_A}) \prod_{k=1}^{d_A} \lambda_k^{d_B - d_A} \text{d}\lambda_k, \quad (A2)$$

where $\Delta (\lambda_1, \ldots, \lambda_{d_A}) = \prod_{1 \leq a < b \leq d_A} (\lambda_b - \lambda_a)$ is the Vandermonde determinant and $Z$ a normalization constant.

To compute the average of the entanglement entropy $S_A$ over the states in $\mathcal{H}$ we start by computing the integral

$$\langle \text{Tr}(\rho_A^2) \rangle = \int \text{Tr}(\rho_A^2) \text{d}\mu(\psi) = \int \left( \sum_{a=1}^{d_A} \lambda_a^2 \right) \text{d}\mu(\lambda_1, \ldots, \lambda_{d_A}). \quad (A3)$$
It is convenient to multiply (A3) by a factor of 1 written using the definition of the gamma function \( \Gamma \),

\[
\frac{1}{\Gamma(d_A d_B + r)} \int_0^\infty \zeta^{d_A d_B + r + 1} e^{-\zeta} d\zeta = 1, \tag{A4}
\]

and perform a change of variables \( q_a = \zeta \lambda_a \). Integrating over \( \zeta \), we eliminate the delta function in the measure. The integration is over \( d_A \) copies of the real positive line, \( q_a \in [0, \infty] \),

\[
\langle \text{Tr} \rho_A^\alpha \rangle = \frac{\hat{Z}}{\Gamma(d_A d_B + r)} \int \left( \sum_{a=1}^{d_A} q_a^r \right)^2 (q_1, \ldots, q_{d_A}) \prod_{k=1}^{d_A} q_k^{d_B - d_A} e^{-q_k} dq_k. \tag{A5}
\]

Each term of the sum over \( a \) gives an equal contribution to \( \langle \text{Tr} \rho_A^\alpha \rangle \). Therefore, we write

\[
\langle \text{Tr} \rho_A^\alpha \rangle = \frac{\hat{Z} d_A}{\Gamma(d_A d_B + r)} \int q_1^r (q_1, \ldots, q_{d_A}) \prod_{k=1}^{d_A} q_k^{d_B - d_A} e^{-q_k} dq_k. \tag{A6}
\]

The Vandermonde determinant can be computed using any monic polynomial. As noticed in [42, 43], it is convenient to compute the Vandermonde determinant using the generalized Laguerre polynomials \( L_i^{(d_B - d_A)} (q_a) \) since they form a complete and orthogonal basis of \( L^2 (\Re^+; q^{d_B - d_A} e^{-q} dq) \). We note that our convention for the Laguerre polynomials includes a factor \((-1)^k k! \),

\[
L_k^{(d_B - d_A)} (q) = (-1)^k k! \sum_{r=0}^{k} \binom{k + d_B - d_A}{r + d_B - d_A} \frac{(-1)^r q^r}{r!}. \tag{A7}
\]

Writing the determinant in terms of the fully antisymmetric tensor \( \epsilon \) with \( d_A \) indices \( i_k = 0, \ldots, d_A - 1 \), the integrals factorize

\[
\langle \text{Tr} \rho_A^\alpha \rangle = \frac{\hat{Z}}{\Gamma(d_A d_B + r)} \sum_{i_k, j_k} \frac{d_A! \epsilon_{i_1 i_2 \cdots i_d_A} \epsilon_{j_1 j_2 \cdots j_d_A}}{\Gamma(d_A d_B + r)} \int q_1^r \prod_{k=1}^{d_A} L_{i_k}^{(d_B - d_A)} (q_k) L_{j_k}^{(d_B - d_A)} (q_k) q_k^{d_B - d_A} e^{-q_k} dq_k. \tag{A8}
\]

All the integrals over \( q_k \) with \( k > 1 \) can be immediately computed using the orthogonality relation of the generalized Laguerre polynomials

\[
\int_0^\infty L_i^{(d_B - d_A)} (q) L_j^{(d_B - d_A)} (q) q^{d_B - d_A} e^{-q} dq \propto \delta_{ij}. \tag{A9}
\]

The two \( \epsilon \) tensors are then contracted on all indices but the first one. Using the relation

\[
\sum_{i_2, \ldots, i_d_A} \epsilon_{i_1 i_2 \cdots i_d_A} \epsilon_{j_1 j_2 \cdots j_d_A} \propto \delta_{i_1 j_1}, \tag{A10}
\]

and introducing a new normalization constant \( \hat{Z} \) to take care of the normalization of the measure (A2) and of other proportionality factors, we write \( \langle \text{Tr} \rho_A^\alpha \rangle \) as

\[
\langle \text{Tr} \rho_A^\alpha \rangle = \frac{\hat{Z}}{\Gamma(d_A d_B + r)} \sum_{i_j} \delta_{ij} X_{ij} (r), \tag{A11}
\]

where we denoted \( X_{ij} \) with \( i, j = 0, \ldots, d_A - 1 \) the integral

\[
X_{ij} (r) = \frac{1}{\Gamma(i + 1) \Gamma(d_B - d_A + i + 1)} \int_0^\infty q^r L_i^{d_B - d_A} (q) L_j^{d_B - d_A} (q) q^{d_B - d_A} e^{-q} dq. \tag{A12}
\]

To compute (A11) we only need the values of \( X_{ij} (r) \) with \( i = j \). However, the computation of all matrix elements \( i \neq j \) is useful for computing higher order moments and therefore is also computed here. The special case \( X_{ij} (0) = \delta_{ij} \) is easy to obtain as it reduces to the orthogonality relation for the Laguerre polynomials. This is sufficient to determine the normalization constant, since

\[
d_A = \langle \text{Tr} \rho_A^0 \rangle = \frac{\hat{Z}}{\Gamma(d_A d_B)} \sum_{i_j} \delta_{ij} X_{ij} (0) = d_A \frac{\hat{Z}}{\Gamma(d_A d_B)}. \tag{A13}
\]
requires \( \tilde{Z} = \Gamma(d_A d_B) \). We then compute the integrals \( X_{ij}(r) \) (A12) for any \( r \geq 0 \) which appear in

\[
\langle \text{Tr} \rho_A^r \rangle = \frac{\Gamma(d_A d_B)}{\Gamma(d_A d_B + r)} \sum_{ij} \delta_{ij} X_{ij}(r).
\]

(A14)

To keep the notation compact we denote \( X(r) \) the \( d_A \times d_A \) matrix with entries \( X_{ij}(r) \). Using the generating function for the generalized Laguerre polynomials

\[
F(t, q) = \sum_{k} (-1)^k \frac{t^k}{k!} L_k^{(d_B - d_A)}(q) = \frac{1}{(1 - t)^{d_B - d_A + 1}} e^{-tq} \quad \text{and} \quad L_k^{(d_B - d_A)}(q) = (-1)^i \left[ \frac{d^i}{dt^i} F(t, q) \right]_{t=0},
\]

we write (A12) as derivatives respect to the parameters \( x \) and \( y \) of the integral of two generating functions:

\[
X_{ij}(r) = \frac{\Gamma(i + 1) \Gamma(r + 1)}{\Gamma(d_B - d_A + i + 1)} \left[ \frac{d^i}{dx^i} \frac{d^j}{dy^j} \int F(x, q) F(y, q) q^{d_B - d_A + r} e^{-q} dq \right]_{x=0, y=0} \quad \text{(A16)}
\]

\[
= \frac{\Gamma(r + 1)^2}{\Gamma(d_B - d_A + r + 1)} \left[ \frac{d^i}{dx^i} \frac{d^j}{dy^j} (1 - xy)^{d_B - d_A - r - 1} (1 - x)^r (1 - y)^r \right]_{x=0, y=0}. \quad \text{(A17)}
\]

We compute the derivatives explicitly by applying successively the Leibniz rule, obtaining a closed form for the integrals \( X_{ij}(r) \)

\[
X_{ij}(r) = \frac{\Gamma(j + 1) \Gamma(r + 1)^2}{\Gamma(d_B - d_A + i + 1)} \sum_{p=0}^{d_A - 1} \frac{\Gamma(d_B - d_A + r + 1 + p)}{\Gamma(r + p - i + 1) \Gamma(j + p + 1) \Gamma(j - p + 1) \Gamma(p + 1)} \quad \text{(A18)}
\]

We conclude the computation for \( \langle S_A \rangle \) noticing that

\[
\langle S_A \rangle = -\langle \text{Tr} \rho_A \log \rho_A \rangle = -\lim_{r \to 1} \partial_r \langle \text{Tr} \rho_A^r \rangle = -\lim_{r \to 1} \partial_r \frac{\Gamma(d_A d_B)}{\Gamma(d_A d_B + r)} \text{Tr} X(r).
\]

(A19)

Taking the derivative and the limit is a long but straightforward calculation that can be done with the help of Wolfram’s Mathematica. The result is the celebrated Page formula,

\[
\langle S_A \rangle = \Psi(d_A d_B + 1) - \Psi(d_B + 1) - \frac{d_A - 1}{2d_B}. \quad \text{(A20)}
\]

The calculation of \( \langle S_A^2 \rangle \) can be done following a similar strategy. We first compute

\[
\langle \text{Tr}(\rho_A^{r_1}) \text{Tr}(\rho_A^{r_2}) \rangle = \int \left( \sum_{a=1}^{d_A} \lambda_a^{r_1} \right) \left( \sum_{b=1}^{d_A} \lambda_b^{r_2} \right) d\mu(\lambda_1, \ldots, \lambda_{d_A}) \quad \text{(A21)}
\]

with \( r_1 > 0 \) and \( r_2 > 0 \). We perform a similar change of variables \( q_a = \zeta \lambda_a \) and eliminate the delta function in the measure by integrating over \( \zeta \). We obtain

\[
\langle \text{Tr}(\rho_A^{r_1}) \text{Tr}(\rho_A^{r_2}) \rangle = \tilde{Z} \frac{1}{\Gamma(d_A d_B + r_1 + r_2)} \left( \sum_{a=1}^{d_A} q_a^{r_1} \right) \left( \sum_{b=1}^{d_A} q_b^{r_2} \right) \Delta^2(q_1, \ldots, q_{d_A}) \prod_{k=1}^{d_A} q_k^{d_B - d_A} e^{-q_k} dq_k. \quad \text{(A22)}
\]

The two sums can be expanded into two terms that can be integrated separately,

\[
\left( \sum_{a=1}^{d_A} q_a^{r_1} \right) \left( \sum_{b=1}^{d_A} q_b^{r_2} \right) = \sum_{a=1}^{d_A} q_a^{r_1 + r_2} + \sum_{a \neq b=1}^{d_A} q_a^{r_1} q_b^{r_2}. \quad \text{(A23)}
\]

The integral of the first term is completely analogous to the computation we just performed resulting in

\[
\langle \text{Tr}(\rho_A^{r_1}) \text{Tr}(\rho_A^{r_2}) \rangle = \frac{\tilde{Z}_1}{\Gamma(d_A d_B + r_1 + r_2)} \sum_{ij} \delta_{ij} X_{ij}(r_1 + r_2), \quad \text{(A24)}
\]
with a proportionality constant \( \tilde{Z}_1 \) to be determined later. The integral of the second term in (A23) reduces to

\[
\tilde{Z}_2 \frac{d_A(d_A-1)}{\Gamma(d_Ad_B+r+1)} \int q_1 r_1 r_2 \Delta^2(q_1, \ldots, q_{d_A}) \prod_{k=1}^{d_A} q_k^{d_B-d_A} e^{-q_k} dq_k. \tag{A25}
\]

Once again, we compute the Vandermonde determinant using Laguerre polynomials. This time the two fully anti-symmetric tensors are contracted on all but two indices:

\[
\sum_{i_3, \ldots, i_{d_A}} \epsilon_{i_1 i_2 i_3 \cdots i_A} \epsilon_{j_1 j_2 j_3 \cdots j_A} \propto \delta_{i_1 j_1} \delta_{i_2 j_2} - \delta_{i_1 j_2} \delta_{i_2 j_1}. \tag{A26}
\]

We recast the integrals in terms of \( X_{ij} (r) \) to obtain

\[
\tilde{Z}_2 \frac{d_A}{\Gamma(d_Ad_B+r_1+1+r_2)} \sum_{ijkl} (\delta_{ij} \delta_{kl} - \delta_{il} \delta_{jk}) X_{ij} (r_1) X_{kl} (r_2). \tag{A27}
\]

Summing the two contribution together we find

\[
\langle \text{Tr} \rho_A^r \text{Tr} \rho_A^s \rangle = \frac{1}{\Gamma(d_Ad_B+r_1+r_2)} \left( \tilde{Z}_1 \text{Tr} X (r_1 + r_2) + \tilde{Z}_2 \text{Tr} X (r_1) \text{Tr} X (r_2) - \tilde{Z}_2 \text{Tr} X (r_1) X (r_2) \right). \tag{A28}
\]

We fix the proportionality constants requiring \( \langle \text{Tr} \rho_A^r \text{Tr} \rho_A^s \rangle = d_A \langle \text{Tr} \rho_A^r \rangle \). The final expression is

\[
\langle \text{Tr} \rho_A^r \text{Tr} \rho_A^s \rangle = \frac{\Gamma(d_Ad_B)}{\Gamma(d_Ad_B+r_1+r_2)} \left( \text{Tr} X (r_1 + r_2) + \text{Tr} X (r_1) \text{Tr} X (r_2) - \text{Tr} (X (r_1) X (r_2)) \right). \tag{A29}
\]

We conclude the computation for \( \langle S_A^2 \rangle \) noticing that

\[
\langle S_A^2 \rangle = \langle \langle -\text{Tr} \rho_A \log \rho_A \rangle^2 \rangle = \lim_{r_1 \to r_1} \partial_{r_1} \partial_{r_2} \langle \text{Tr} \rho_A^r \text{Tr} \rho_A^s \rangle. \tag{A30}
\]

We take the derivatives and the limits with the help of Wolfram’s Mathematica. The variance of the entanglement entropy of a subsystem is defined as \( (\Delta S_A)^2 = \langle S_A^2 \rangle - \langle S_A \rangle^2 \). Substituting our expressions for \( \langle S_A^2 \rangle \) and \( \langle S_A \rangle \) we obtain the result reported in the main text paper:

\[
(\Delta S_A)^2 = -\Psi'(d_Ad_B + 1) + \frac{d_A + d_B}{d_Ad_B + 1} \Psi' (b_B + 1) - \frac{(d_A - 1)(d_A + 2d_B - 1)}{4d_B^2 (d_Ad_B + 1)}. \tag{A31}
\]

We can compute higher order moments of the entanglement entropy distribution employing the same strategy used to compute average and variance. As an example, we also report the computation for the third momentum \( m_3 = \langle (S_A - \langle S_A \rangle)^3 \rangle \). First, we compute

\[
\langle \text{Tr} \rho_A^r \text{Tr} \rho_A^s \text{Tr} \rho_A^t \rangle = \frac{\Gamma(d_Ad_B)}{\Gamma(d_Ad_B+r_1+r_2+r_3+r_4)} \left( \text{Tr} X (r_1 + r_2 + r_3) + \text{Tr} X (r_1 + r_2) \text{Tr} X (r_3) - \text{Tr} (X (r_1 + r_2) X (r_3)) \right)
\]

\[
\text{Tr} X (r_1 + r_3) \text{Tr} X (r_2) - \text{Tr} (X (r_1 + r_3) X (r_2)) + \text{Tr} X (r_3 + r_2) \text{Tr} X (r_1) - \text{Tr} (X (r_3 + r_2) X (r_1)) \right)
\]

\[
\left( \text{Tr} X (r_1) \text{Tr} X (r_2) \text{Tr} X (r_3) + \text{Tr} (X (r_1) X (r_3) X (r_2)) + \text{Tr} (X (r_1) X (r_2) X (r_3)) - \text{Tr} (X (r_2) \text{Tr} (X (r_1) X (r_3)) - \text{Tr} (X (r_1) \text{Tr} (X (r_2) X (r_3)) - \text{Tr} (X (r_3) \text{Tr} (X (r_1) X (r_2))) \right)
\]

Second, we take the limit of the derivatives respect \( r_1, r_2, \) and \( r_3 \)

\[
\langle S_A^3 \rangle = -\lim_{r_1 \to r_1} \partial_{r_1} \partial_{r_2} \partial_{r_3} \langle \text{Tr} \rho_A^r \text{Tr} \rho_A^s \text{Tr} \rho_A^t \rangle. \tag{A32}
\]

The computation is an herculean task but with the help of Wolfram’s Mathematica we are able to simplify the exact
the dimension of each sector, with |

any loss of generality we can assume that a basis |

Repeating this decomposition on all the subspaces |

The skews of a probability distribution is a measure of the asymmetry of the distribution and is defined as the |

We give an example for a small system consisting of two qubits. The exact formulas (A20), (A31) at $d_A = 2$, $d_B = 2$ evaluate to |

which match the result of the numerical evaluation of the average over random pure states.

Appendix B: Uniform measure over a direct sum of Hilbert spaces

We consider the eigenspace $\mathcal{H}(E) \subset \mathcal{H}$ with fixed energy $E$. $\mathcal{H}(E)$ has the structure of a direct sum of tensor products

where $\mathcal{H}_A(\varepsilon_j)$ and $\mathcal{H}_B(\varepsilon_k)$ are eigenspaces of given energy for the subsystems $A$ and $B$.

Energy eigenspaces of the subsystem $A$ are denoted $\mathcal{H}_A(\varepsilon_j)$ and have dimension $d_{j,A} = \dim \mathcal{H}_A(\varepsilon_j)$. Similarly for subsystem $B$. The energy eigenspaces of the system have then the direct sum structure

where the sector $\mathcal{H}_j(E) = \mathcal{H}_A(\varepsilon_j) \otimes \mathcal{H}_B(E - \varepsilon_j)$ has definite energy in each subsystem. We denote $d_j = \dim \mathcal{H}_j(E)$ the dimension of each sector, with $d_j = d_{j,A} d_{j,B}$ and $d_E = \sum_j d_j$ the dimension of $\mathcal{H}(E)$.

Any state $|\psi, E\rangle \in \mathcal{H}(E)$ can be written as $|\psi, E\rangle = \sum_j \sqrt{p_j} |\phi_j\rangle$ with $|\phi_j\rangle$ normalized to 1 and $\sum_j p_j = 1$. The coefficient $p_j$ can be interpreted as the probability of finding the state $|\psi\rangle$ in the sub-Hilbert space $\mathcal{H}_j(E)$. Without any loss of generality we can assume that a basis $|n, E\rangle$ of $\mathcal{H}(E)$ is adapted to the decomposition in $\mathcal{H}_j(E)$ meaning that $|1, E\rangle, \ldots, |d_1, E\rangle$ is a basis of $\mathcal{H}_1(E), |d_1 + 1, E\rangle, \ldots, |d_1 + d_2, E\rangle$ is a basis of $\mathcal{H}_2(E)$ and so on. Focusing for clarity on $j = 1$, it is easy to see that

Repeating this decomposition on all the subspaces $\mathcal{H}_j(E)$, the uniform measure on $\mathcal{H}(E)$ can be written as the probability distribution of finding $|\psi, E\rangle$ in $\mathcal{H}_j(E)$ times the product of the uniform measures on $\mathcal{H}_j(E)$:

where

\[d\mu(\psi) = d\nu(p) \prod_j d\mu(\phi_j),\] (B4)

\[d\nu(p) = \frac{1}{Z} \delta\left(1 - \sum_j p_j\right) \prod_j p_j^{d_j - 1} dp_j.\] (B5)
The normalization constant $Z$ can be computed using a procedure similar to the one used in (A2),

$$Z = \int \delta\left(1 - \sum_j p_j\right) p_j^{d_j - 1} \prod_j dp_j = \frac{\prod_j \Gamma(d_j)}{\Gamma(d_E)}.$$  \hspace{1cm} (B6)

The average and the variance of $p_j$ can be easily shown to be given by

$$\langle p_j \rangle = \frac{d_j}{d_E}, \quad (\Delta p_j)^2 = \frac{d_j (d_E - d_j)}{d_E^2 (d_E + 1)}.$$  \hspace{1cm} (B7)

The average of the Shannon entropy for the probability distribution (B5) is given by

$$\langle S(p) \rangle = -\sum_j p_j \log p_j = - \left[ \partial_r \sum_j p_j^j \right]_{r=1} = \Psi(d_E + 1) - \sum_j \frac{d_j}{d_E} \Psi(d_j + 1).$$  \hspace{1cm} (B8)

We illustrate the result with two examples. The first one consists in taking all equal dimensions $d_j = d_E/J$. We find:

$$\langle p_j \rangle = \frac{1}{J}, \quad (\Delta p_j)^2 = \frac{J - 1}{J^2} \frac{1}{d_E + 1},$$

$$\langle S(p) \rangle = \Psi(d_E + 1) - \Psi(d_E/J + 1) \approx \log(J) \quad \text{for} \quad J \gg 1. \hspace{1cm} (B9)$$

As a second example, we consider the case where the dimension $d_j$ is much larger than the sum of all the others $d_j \gg d_R = \sum_{i=1}^{J-1} d_i = d_E - d_J$. In this case the exact formulas reduce to

$$\langle p_J \rangle \approx 1 - \frac{d_R}{d_J}, \quad (\Delta p_J)^2 \approx \frac{d_R}{d_J}$$

$$\langle p_i \rangle \approx \frac{d_i}{d_J}, \quad (\Delta p_i)^2 \approx \frac{d_i}{d_J} \quad \text{for} \quad i \neq J,$$

$$\langle S(p) \rangle \approx \frac{1}{d_J} \left( d_R + d_R \log d_J + \sum_{i=1}^{J-1} d_i \Psi(d_i + 1) \right). \hspace{1cm} (B13)$$

Note in particular that the average Shannon entropy goes to zero as $d_j \to \infty$. Furthermore if all the dimensions are large $d_j \gg 1$, we note that the average Shannon entropy equals the Shannon entropy of the average probability,

$$\langle S(p) \rangle \approx \sum_{j=1}^{J} \langle p_j \rangle \log(p_j). \hspace{1cm} (B14)$$

The computation of $\Delta S(p)^2$ is straightforward but its expression is convoluted. We report here its leading order for $d_j \gg 1$,

$$\Delta S(p)^2 \approx \frac{1}{d_E} \sum_j \frac{d_j}{d_E} \left( 1 + \log \frac{d_j}{d_E} \right)^2 - \frac{1}{d_E} \sum_{ij} \frac{d_i d_j}{d_E^2} \left( 1 + \log \frac{d_i}{d_E} \right) \left( 1 + \log \frac{d_j}{d_E} \right). \hspace{1cm} (B15)$$

The variance of the Shannon entropy vanishes as $1/d_E$ for $d_j \to \infty$.

[1] D. N. Page, *Average entropy of a subsystem, Phys. Rev. Lett. 71* (1993) 1291–1294, [gr-qc/9305007].
[2] D. N. Page, *Information in black hole radiation, Phys. Rev. Lett. 71* (1993) 3743–3746, [hep-th/9306083].
[3] S. B. Giddings, *“Black holes, quantum information, and unitary evolution,” Phys. Rev. D 85* (2012) 124063 doi:10.1103/PhysRevD.85.124063 [arXiv:1201.1037].
[4] S. L. Braunstein, S. Pirandola and K. yczkowski, “Better Late than Never: Information Retrieval from Black Holes,” *Phys. Rev. Lett. 110*, no. 10, 101301 (2013) [arXiv:0907.1190].
[5] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, *Black Holes: Complementarity or Firewalls?, JHEP 02* (2013) 062, [arXiv:1207.3123].
[6] D. Marolf, *The Black Hole information problem: past, present, and future, Rept. Prog. Phys. 80* (2017), no. 9
entanglement entropy for gauge fields, Phys. Rev. D89 (2014), no. 8 085012, [arXiv:1312.1183].
[47] C. T. Ma, “Entanglement with Centers,” JHEP 1601, 070 (2016), [arXiv:1511.02671].
[48] J. Lin and D. Radicevic, “Comments on Defining Entanglement Entropy,” [arXiv:1808.05939].
[49] C. Rovelli, Quantum Gravity. Cambridge Monographs on Mathematical Physics. Cambridge, UK, (2004).
[50] W. Donnelly, Entanglement entropy in loop quantum gravity, Phys.Rev. D77(2008) 104006, [arXiv:0802.0880]
[51] A note on conventions: In [24], Theorem III.3 is expressed using the entropy, the log and the exp all taken in base two. On the other hand, in this paper and in Eq. (19) in particular, we always use log(x) and exp(x) to denote the natural logarithm and exponential.
[52] M. Ohya and D. Petz, Quantum Entropy and Its Use. Springer Science & Business Media, 2004.
[53] N. Balakrishnan and V. B. Nevzorov, A Primer on Statistical Distributions. Wiley-Interscience, 2003.
[54] N. F. Ramsey, Thermodynamics and statistical mechanics at negative absolute temperatures, Phys. Rev. 103.1 (1956): 20.
[55] R. Balian, From Microphysics to Macrophysics: Methods and Applications of Statistical Physics, (Vol.I), Springer (2007).
[56] R. D. Sorkin, On the Entropy of the Vacuum outside a Horizon, in Proceedings, 10th International Conference on General Relativity and Gravitation: Padua, Italy, July 4-9, 1983, vol. 2, pp. 734–736, 1984. [arXiv:1402.3589].
[57] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, A Quantum Source of Entropy for Black Holes, Phys. Rev. D 34 (1986) 373–383.
[58] M. Srednicki, Entropy and area, Phys. Rev. Lett. 71 (Aug, 1993) 666–669.
[59] C. Holzhey, F. Larsen, and F. Wilczek, Geometric and renormalized entropy in conformal field theory, Nucl. Phys. B424 (1994) 443–467, [hep-th/9403108].
[60] A. Almheiri and J. Sully, An Uneventful Horizon in Two Dimensions, JHEP 02 (2014) 108, [arXiv:1307.8149].
[61] E. Bianchi and M. Smerlak, Entanglement entropy and negative energy in two dimensions, Phys. Rev. D90 (2014), no. 4 041904, [arXiv:1404.0602].
[62] E. Bianchi and A. Satz, Entropy of a subalgebra of observables and the geometric entanglement entropy, Phys. Rev. D99 (2019), no. 8 085001, [arXiv:1901.06454].
[63] T. Kinoshita, T. Wenger, and D. S. Weiss, A quantum Newton's cradle, Nature 440, 900 (2006).
[64] R. Islam, R. Ma, P. M. Preiss, M. E. Tai, A. Lukin, M. Rispoli, M. Greiner, Measuring entanglement entropy in a quantum many-body system, Nature 528, 77 (2015).
[65] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Quantum thermalization through entanglement in an isolated many-body system, Science 353, 794 (2016).
[66] Y. Tang, W. Kao, K.-Y. Li, S. Seo, K. Mallayya, M. Rigol, S. Gopalakrishnan, and B. L. Lev, Thermalization Near Integrability in a Dipolar Quantum Newton’s Cradle, Phys. Rev. X 8, 021030 (2018).