Tutorial

Volume-Law Entanglement Entropy of Typical Pure Quantum States

Eugenio Bianchi,[1,2,*] Lucas Hackl,[3,4,†] Mario Kieburg,[3,‡] Marcos Rigol,[1,§] and Lev Vidmar[5,6,¶]

[1] *Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[2] *Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[3] *School of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria 3010, Australia*
[4] *School of Physics, The University of Melbourne, Parkville, Victoria 3010, Australia*
[5] *Department of Theoretical Physics, J. Stefan Institute, Ljubljana SI-1000, Slovenia*
[6] *Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana SI-1000, Slovenia*

(Received 17 December 2021; revised 4 May 2022; published 26 July 2022)

The entanglement entropy of subsystems of typical eigenstates of quantum many-body Hamiltonians has recently been conjectured to be a diagnostic of quantum chaos and integrability. In quantum chaotic systems it has been found to behave as in typical pure states, while in integrable systems it has been found to behave as in typical pure Gaussian states. In this tutorial, we provide a pedagogical introduction to known results about the entanglement entropy of subsystems of typical pure states and of typical pure Gaussian states. They both exhibit a leading term that scales with the volume of the subsystem, when smaller than one half of the volume of the system, but the prefactor of the volume law is fundamentally different. It is constant (and maximal) for typical pure states, and it depends on the ratio between the volume of the subsystem and of the entire system for typical pure Gaussian states. Since particle-number conservation plays an important role in many physical Hamiltonians, we discuss its effect on typical pure states and on typical pure Gaussian states. We prove that, while the behavior of the leading volume-law terms does not change qualitatively, the nature of the subleading terms can change. In particular, subleading corrections can appear that depend on the square root of the volume of the subsystem. We unveil the origin of those corrections. Finally, we discuss the connection between the entanglement entropy of typical pure states and analytical results obtained in the context of random matrix theory, as well as numerical results obtained for physical Hamiltonians.

DOI: 10.1103/PRXQuantum.3.030201

CONTENTS

I. INTRODUCTION 2
A. Ground-state entanglement 2
B. Excited-state entanglement 3
C. Random matrix theory in physics 4
D. Local spectral statistics 4
E. Global spectral statistics and eigenvector statistics 6
F. Typicality and entanglement 6
G. Outline 7
II. GENERAL PURE STATES 8
A. Arbitrary number of particles 8
1. Statistical ensemble of states 9
2. Average and variance 10
B. Fixed number of particles 11
1. Statistical ensemble of states 12
2. Average and variance 13
3. Weighted average and variance 17
III. PURE FERMIONIC GAUSSIAN STATES 18
A. Definition of fermionic Gaussian states 18
B. Arbitrary number of particles 20
1. Statistical ensemble of states 20
2. Average and variance 21
C. Fixed number of particles 22
1. Statistical ensemble of states 22

*ebianchi@psu.edu
†lucas.hackl@unimelb.edu.au
‡m.kieburg@unimelb.edu.au
§mrigol@phys.psu.edu
¶lev.vidmar@ijs.si

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

2. *Average and variance* 24
3. *Weighted average and variance* 26

IV. EXACT RELATION TO RANDOM HAMILTONIANS 28
A. Random many-body Hamiltonians 28
B. Random quadratic Hamiltonians 30
C. Dynamical averages 31

V. RELATION TO PHYSICAL HAMILTONIANS 32
A. Quantum-chaotic interacting model 32
B. Quantum-chaotic quadratic model 34
C. Translationally invariant noninteracting fermions 35

VI. SUMMARY AND OUTLOOK 36
A. Summary 36
B. Outlook 37
1. *Typical eigenstate entanglement entropy as a diagnostic of quantum chaos and integrability* 37
2. *Beyond qubit-based systems* 37
3. *Other ensembles and entanglement measures* 38

ACKNOWLEDGMENTS 38

APPENDIX A: GENERAL PURE STATES WITH A FIXED NUMBER OF PARTICLES 39
1. Average entanglement entropy 39
a. Asymptotic expansion of the logarithm of the ratio of dimensions 40
b. Computation of N_{crit} 41
c. Moments of ϱ_{N_A} 44
d. Average over dimension ratios 48
e. Resulting formula 51
f. Resolving the critical regimes 53
2. Weighted average over sectors 55

APPENDIX B: INTEGRATION FORMULA FOR JACOBI POLYNOMIALS 57

APPENDIX C: PURE FERMIONIC GAUSSIAN STATES 59
1. Average entanglement entropy 59
2. Variance 60

APPENDIX D: PURE FERMIONIC GAUSSIAN STATES WITH A FIXED NUMBER OF PARTICLES 60
1. Average entanglement entropy 60
2. Variance 63
a. Weighted average over sectors 64
b. Resulting formula 64
c. Resolving the critical regimes 67

REFERENCES 71

I. INTRODUCTION

Entanglement is a defining property of quantum theory, and plays a crucial role in a broad range of problems in physics, ranging from the black hole information paradox [1] to the characterization of phases in condensed matter systems [2]. Put simply, entanglement refers to quantum correlations between different parts of a physical system that cannot be explained classically [3,4]. Over the years, a wide range of *entanglement measures* have been devised to quantify entanglement [5]. Prominent among those are the *bipartite* entanglement measures, which involve splitting the system in two parts.

For the special case of globally pure quantum states $|\psi\rangle$ (our interest here) and a bipartition, the von Neumann entanglement entropy, also known as the entropy of entanglement or just the *entanglement entropy*, is one of the simplest measures of quantum entanglement. It vanishes if and only if there is no quantum entanglement between the two parts, in which case the state must be a product state. We study the entanglement entropy in Hilbert spaces with a tensor product structure $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ [6]. To compute the entanglement entropy of subsystem A (with volume V_A) of $|\psi\rangle$, one traces out the complement subsystem B (with volume $V - V_A$, where V is the total volume) to obtain the mixed density matrix $\hat{\rho}_A = \mathrm{Tr}_{\mathcal{H}_B} |\psi\rangle \langle\psi|$. The entanglement entropy S_A of subsystem A is then

$$S_A = -\mathrm{Tr}(\hat{\rho}_A \ln \hat{\rho}_A), \tag{1}$$

while the nth Rényi entropy is defined as

$$S_A^{(n)} = -\ln[\mathrm{Tr}(\hat{\rho}_A^n)]. \tag{2}$$

The second-order Rényi entropy $S_A^{(2)}$ has already been measured in experiments with ultracold atoms in optical lattices [7,8].

We stress that the focus of this tutorial is in pure quantum states. Quantifying entanglement in globally mixed states is more challenging. In particular, the von Neumann and Rényi entanglement entropies are not entanglement measures for globally mixed states. Several of the bipartite entanglement measures defined for mixed states (e.g., distillable entanglement, entanglement cost, entanglement of formation, relative entropy of entanglement, and squashed entanglement) reduce to the entanglement entropy when evaluated on pure states [5].

A. Ground-state entanglement

In general, one is interested in understanding the behavior of measures of entanglement in physical systems, and in determining what such a behavior can tell us about the physical properties of the system. Much progress in this direction has been achieved in the context of many-body ground states of local Hamiltonians, for which a wide range of theoretical approaches are available [2,9–11]. Such ground states usually exhibit a leading term of the entanglement entropy that scales with the area, or with the logarithm of the volume, of the subsystem. Identifying and understanding universal properties

of the entanglement entropy in ground states of local Hamiltonians has been a central goal [12–15].

In one-dimensional systems of spinless fermions or $\frac{1}{2}$ spins, the leading (in the volume V_A) term in the entanglement entropy has been found to distinguish ground states of critical systems from those of noncritical ones [15–17]. In the former the leading term exhibits a logarithmic scaling with the volume (when described by conformal field theory, the central charge is the prefactor of the logarithm [15,16,18]), while in noncritical ground states the leading term is a constant (which, in one dimension, reflects an area-law scaling). Subleading terms have also been studied, specially in the context of states that are physically distinct but exhibit the same leading entanglement entropy scaling. An example in the context of quadratic Hamiltonians in two dimensions are ground states that are critical with a pointlike Fermi surface versus noncritical, which both exhibit a leading area-law entanglement entropy [19–23]. Remarkably, the subleading term in the former scales logarithmically with V_A while it is constant for noncritical ground states [24]. Also, in two-dimensional systems, critical states described by conformal field theory [25] and states with a spontaneously broken continuous symmetry [26,27] have been found to exhibit a universal subleading logarithmic term.

B. Excited-state entanglement

In recent years, interest in understanding the far-from-equilibrium dynamics of (nearly) isolated quantum systems and the description of observables after equilibration [28–30] have motivated many studies of the entanglement properties of highly excited eigenstates of quantum many-body systems (mostly in the context of lattice systems) [31–64]. Because of the limited suit of tools available to study entanglement properties of highly excited eigenstates of model Hamiltonians, most of the results reported in those works were obtained using exact diagonalization techniques, which are limited to relatively small system sizes.

In contrast to the ground states, typical highly excited many-body eigenstates of local Hamiltonians have a leading term of the entanglement entropy that scales with the volume of the subsystem. Also, in contrast to the ground states, the leading volume-law term exhibits a fundamentally different behavior depending on whether the Hamiltonian is nonintegrable (the generic case for physical Hamiltonians) or integrable. In the former case the coefficient has been found to be constant, while in the latter case it depends on the ratio between the volume of the subsystem and the volume of the entire system.

Many-body systems that are integrable are special as they have an extensive number of local conserved quantities [65]. As a result, their equilibrium properties can in many instances be calculated analytically, and their near-equilibrium properties can be "anomalous," e.g., they can exhibit transport without dissipation (ballistic transport). Also, isolated integrable systems fail to thermalize if taken far from equilibrium. Interested readers can learn about the effects of quantum integrability in the collection of reviews in Ref. [66].

There is a wide range of quadratic Hamiltonians in arbitrary dimensions (which include a wide range of noninteracting models), e.g., translationally invariant quadratic Hamiltonians, that can be seen as a special class of integrable models. A class in which the nondegenerate many-body eigenstates are Gaussian states, while their degenerate eigenstates can always be written as Gaussian states. This means that those many-body eigenstates are fully characterized by their one-body density matrix or their covariance matrix. The entanglement entropy of highly excited eigenstates of some of those "integrable" quadratic Hamiltonians was studied in Refs. [36,42,44,50,53]. Other quadratic Hamiltonians in arbitrary dimensions that will be of interest to us here are quadratic Hamiltonians in which the single-particle sector exhibits quantum chaos (to be defined in the next subsections). We refer to such Hamiltonians as quantum-chaotic quadratic Hamiltonians. The entanglement entropy of highly excited eigenstates of quantum-chaotic quadratic Hamiltonians (on a lattice) was studied in Refs. [61,62]. It was found to exhibit a typical leading volume-law term that is qualitatively similar to that found in eigenstates of integrable quadratic Hamiltonians (in which the single-particle sector does not display quantum chaos), such as translationally invariant quadratic Hamiltonians (on a lattice) [42,50].

In the presence of interactions, many-body integrable systems mostly exist in one dimension [67,68]. They come in two "flavors," Hamiltonians that can be mapped onto noninteracting ones (a smaller class), and Hamiltonians that cannot be mapped onto noninteracting ones. Remarkably, both "flavors" have been found to describe pioneering experiments with ultracold quantum gases in one dimension [69–85]. The entanglement entropy of highly excited eigenstates of lattice Hamiltonians that can be mapped onto noninteracting ones (which exhibit the same leading volume-law terms as their noninteracting counterparts) was studied in Refs. [48,50], while the entanglement entropy of highly excited eigenstates of a Hamiltonian (the spin-$\frac{1}{2}$ XXZ chain) that cannot be mapped onto a noninteracting one was studied in Ref. [55]. Remarkably, in all the quadratic and integrable systems studied so far, the coefficient of the leading volume-law term of typical eigenstates has been found to depend on the ratio between the volume of the subsystem and the volume of the entire system.

Analytical progress understanding the previously mentioned numerical results has been achieved in some special cases. One such case is translationally invariant quadratic Hamiltonians, or models that can be mapped onto them in one dimension [67], for which tight bounds were obtained

for the leading (volume-law) term in the average entanglement entropy [42,50], and some understanding was gained about subleading corrections [48]. This was possible thanks to the Gaussian nature of the eigenstates. Another case is nonintegrable models under the assumption that their eigenstates exhibit eigenstate thermalization [33,45,46,52].

C. Random matrix theory in physics

Random matrix theory has provided a more systematic approach to gaining an analytical understanding of the entanglement properties of many-body eigenstates in nonintegrable models [40,43,63,86–89]. Such an approach is justified by the fact that many studies (see, e.g., Ref. [29] for a review) have shown that nonintegrable models exhibit "quantum chaos." By quantum chaos what is meant is that statistical properties of highly excited eigenstates of such models, e.g., level spacing distributions, are described by the Wigner surmise [29]. This was conjectured by Bohigas, Giannoni, and Schmit (BGS) [90] for quantum systems with a classical counterpart, in which case "quantum chaos" usually occurs when the classical counterparts are K chaotic, where K stands for Kolmogorov, and it is the class of systems that exhibit the highest degree of chaos. Remarkably, even statistical properties of eigenvectors such as the ratio between the variance of the diagonal and the off-diagonal matrix elements of Hermitian operators have been shown to agree with random matrix theory predictions [91–94]. Recently, two of us (M.R. and L.V., in collaboration with P. Łydżba) used random matrix theory in the context of quantum-chaotic quadratic Hamiltonians to obtain a closed-form expression that describes the average entanglement entropy of highly excited eigenstates of quadratic models whose single-particle spectrum exhibits quantum chaos, such as the three-dimensional Anderson model [61,62].

The application of random matrix theory to many-body systems goes back to works by Wigner [95–98] as well as Landau and Smorodmsky [99] in the 1950s, who aimed at finding a statistical theory that described the excitation spectra in nuclei for elastic scattering processes. Their novel idea was that a sufficiently complicated operator like the Hamilton, or the lattice Dirac operator, can be replaced by a random matrix (whose entries are, preferably, Gaussian distributed as those are easier to deal with analytically) with the appropriate symmetries. For this to hold, it is not important that the physical operator has matrix entries that are all occupied with nonzero entries. In condensed matter models [29], as well as in lattice QCD [100–104], numerical evidence has shown that very sparse matrices can also exhibit spectral characteristics of a random matrix with Gaussian distributed entries. It is the concept of universality that has made random matrices so versatile. Like in the central limit theorem, in which an infinite sum of independently and identically distributed random variables leads to a Gaussian random variable under very mild conditions, it happens that, for many spectral quantities, it does not matter how the random matrix is actually distributed.

Over the years, random matrix theory has found many more applications in physics; for example, the local level density about Dirac points (also known as hard edges in random matrix theory) has been used to classify operators such as Hamiltonians and Dirac operators, and to discern global symmetries of a system. By global symmetries, it is meant those that are described by a linear involution (operators that square to unity) in terms of unitary and antiunitary operators. Well-known examples in physics are time reversal, parity, charge conjugation, and chirality. Global symmetries play a central role when classifying systems in the context of quantum chaos [105], in superconductors and topological insulators [106,107], in quantum-chromodynamics-like theories in the continuum and on a lattice [104,108], and in Sachdev-Ye-Kitaev (SYK) models [109,110].

D. Local spectral statistics

There are two spectral scales that are usually discussed in the context of random matrix theory, and to which different kinds of universalities apply. Those are the local and the global spectral scales.

The microscopic or local spectral scale is given by the local mean level spacing where the fluctuations of the individual eigenvalues are resolved. This scale is often of more physical interest as it analyses the level repulsion of eigenvalues that are very close to each other. Such a level repulsion is usually algebraic for very small distances s. Namely, the level spacing distribution $p(s)$, which is the distribution of the distance of two consecutive eigenvalues, is of the form s^β (where β is the Dyson index) for small distances.

While the symmetry of a Hamiltonian, such as time reversal, chirality, or charge conjugation, is not very important for the global spectral scale, it is very important for the local spectral statistics as it influences the value of β. Wigner [97] derived the distribution for two-level Gaussian random matrices with Dyson index $\beta = 1$, which was soon generalized to $\beta = 2, 4$,

$$p(s) = 2\frac{(\Gamma[(\beta+2)/2])^{\beta+1}}{(\Gamma[(\beta+1)/2])^{\beta+2}} s^\beta \times \exp\left[-\left(\frac{\Gamma[(\beta+2)/2]}{\Gamma[(\beta+1)/2]}\right)^2 s^2\right] \tag{3}$$

with the gamma function $\Gamma[x]$. This distribution is nowadays called Wigner's surmise. The corresponding random matrices are known as the Gaussian orthogonal ensemble (GOE; $\beta = 1$), the Gaussian unitary ensemble (GUE; $\beta = 2$), and the Gaussian symplectic ensemble (GSE; $\beta =$

4). Those are usually compared with the level spacing distribution of independently distributed eigenvalues ($\beta = 0$), which gives the Poisson distribution

$$p(s) = e^{-s}, \tag{4}$$

and with the level spacing distribution of the one-dimensional quantum harmonic oscillator (also known as the picket fence statistics), which is a simple Dirac delta function

$$p(s) = \delta(1 - s). \tag{5}$$

All five benchmark distributions are shown in Fig. 1(a).

The use of the Wigner surmise as a diagnostic of quantum chaos and integrability followed fundamental conjectures by BGS [90] (mentioned before) and Berry and Tabor [111], respectively. The latter states that, for an integrable bounded system with more than two dimensions and incommensurable frequencies of the corresponding tori, the spectrum should follow the Poisson statistics. However, both conjectures have to be understood with the following care as the eigenvalue spectrum must be prepared appropriately.

(i) The spectrum must be split into subspectra with fixed "good" quantum numbers such as the spin, parity, and conserved charges. This requires knowledge of all the symmetries of the model. This step must be taken since a direct sum of independent GUE matrices can yield a level spacing distribution that resembles the Poisson statistics; see Fig. 1(b).

(ii) One needs to unfold the spectra, meaning that the distance between consecutive eigenvalues must be in average equal to one. This second step is crucial as only then the level spacing distributions are comparable and universal statistics can be revealed. The eigenvalue spectrum of an irregularly shaped drum, a complex molecule, and that of a heavy nucleus have completely different energy scales. After the unfolding of their spectra these scales are removed and show common behavior. Yet, the procedure of unfolding is far from trivial for empirical spectra. There are other means such as the study of the ratio between the two spacings of three consecutive eigenvalues [112]. But this observable also has its limitations as this kind of "automatic unfolding" only works in the bulk of the spectrum. It fails at spectral edges and other critical points in the spectrum.

In the context of the Wigner surmise, we should stress that even though the statistics of the spectral fluctuations are well described at the level of the mean level spacing [113–115] (even beyond the context of many-body systems; see, e.g., the reviews and books [116–119] and the references therein), it was soon realized that there are statistical properties of the spectral fluctuations of many-body Hamiltonians that cannot be described using full random matrices; see Refs. [120–123]. This is due to the fact that usually only one-, two- and maybe up to four-body interactions represent the actual physical situation. Random matrices that reflect these sparse interactions are called embedded random matrix ensembles [116,122,124–126]. In the past decades, they have experienced a revival due to studies of the SYK model [127–132], and two-body interactions [133–135]. A full understanding of how these additional tensor structures, which arise naturally in quantum many-body systems, impact the entanglement of the energy eigenstates is currently missing.

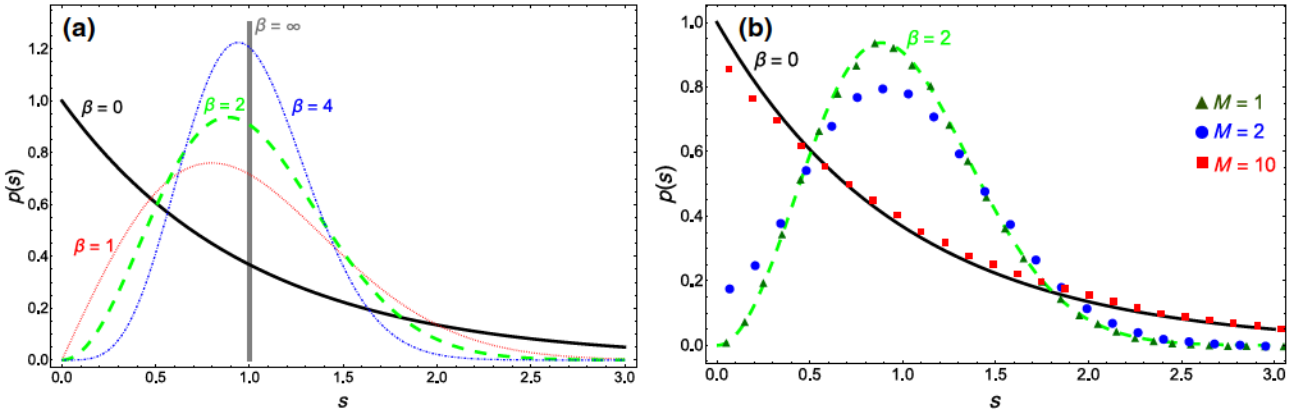


FIG. 1. (a) The level spacing distributions of the Poisson distribution (solid line; $\beta = 0$), the Wigner surmise of the GOE (dotted line; $\beta = 1$), of the GUE (dashed line; $\beta = 2$), of the GSE (dash-dot line; $\beta = 4$), and the picket fence statistics (vertical line; $\beta = \infty$). (b) Three Monte Carlo simulations (symbols) of the spacing between eigenvalues $(50 \cdot M)$ and $(50 \cdot M + 1)$ of the direct sum of M GUEs with a matrix dimension $N = 100$ (in total, the matrix dimension is $100^M \times 100^M$), compared to the Poisson distribution (solid line), and the Wigner surmise of the GUE (dashed line). The ensemble size is 10^5 such that the statistical error is about 1%. The bin size is about 0.1, but varies as the unfolding slightly changes their actual value.

E. Global spectral statistics and eigenvector statistics

The second scale is the macroscopic or global spectral scale, which is usually defined as the average distance between the largest and the smallest eigenvalues. For this scale, Wigner [95,98] derived the famous Wigner semicircle, which describes the level density of a Gaussian distributed real symmetric matrix. He was also the first to show, again under mild conditions, that the Gaussian distribution of the independent matrix entries can be replaced by an arbitrary distribution, and nevertheless one still obtains the Wigner semicircle. One important feature of this kind of universality is that it does not depend on the symmetries of the operators. For instance, whether the matrix is real symmetric, Hermitian, or Hermitian self-dual has no impact on the level density, which is in all those cases a Wigner semicircle [136]. The global spectral scale also plays a crucial role in time series analysis [137] and telecommunications [138], where instead of the Wigner semicircle the Marčenko-Pastur distribution [139] describes the level density.

The global scale is always important when considering the so-called linear spectral statistics, meaning an observable that is of the form $\sum_{j=1}^{N} f(\lambda_j)$, where the λ_j are the eigenvalues of the random matrix. This is the situation that we encounter when computing the entanglement entropy, where the λ_j are the eigenvalues of the density matrix; cf. Eq. (1). Therefore, we expect that the leading terms in the entanglement entropy are insensitive to the Dyson index β, so that the entanglement entropy can serve as an excellent diagnostic for integrable or chaotic behavior.

A related diagnostic for the amplitude A of vector components of eigenstates is the Porter-Thomas distribution [140], which is used to decide whether a state is localized or delocalized. The Porter-Thomas distribution is a χ^2 distribution,

$$\mathcal{I}(A) = \left(\frac{\beta N}{2}\right)^{\beta/2} \frac{A^{\beta/2-1}}{\Gamma[\beta/2]} \exp\left[-\frac{\beta N}{2} A\right], \qquad (6)$$

where the normalization of the first moment is chosen to be equal to $1/N$. Note that in the quaternion case one defines the amplitude as the squared modulus of a quaternion number. Hence, as a sum of four squared real components, similar to the squared modulus of a complex number (which is the sum of the square of the real and imaginary parts). Actually, the application of random matrices for computing the entanglement entropy is based on this idea. We can only replace a generic eigenstate by a Haar-distributed vector on a sphere after assuming that the state is delocalized. Unlike the Porter-Thomas distribution, as previously mentioned, the leading terms in the entanglement entropy are expected to be independent of the Dyson index β (which has yet to be proved).

The relation between certain quantum informational questions and random matrix theory also has a long history, and the techniques developed are diverse (see, e.g., the review [141] and Chapter 37 of Ref. [118]). Questions about generic distributions and the natural generation of random quantum states have been a focus of attention [142,143]. The answers to those questions are still debated as there are several measures of the set of quantum states and each has its benefits and flaws; for instance, two of those are based on the Hilbert-Schmidt metric and the Bures metric [142,144]. Those measures define some kind of "uniform distribution" on the set of all quantum states and, actually, generate random matrix ensembles that have been studied to some extent [142,143,145–149]. In this tutorial, we encounter one of the aforementioned ensembles, namely, the one related to the Hilbert-Schmidt metric, which naturally arises from a group action so that the states are Haar distributed according to this group action.

F. Typicality and entanglement

An important question that one can ask, which relates to the latest observations made in the context of random matrix ensembles, is what are the entanglement properties of typical pure quantum states. This was the earliest question to be addressed. Following work by Lubkin [150] and Lloyd and Pagels [151], Page [152] obtained a closed analytical formula for the average entanglement entropy (over all pure quantum states) as a function of the system and subsystem Hilbert space dimensions. This formula was rigorously proven later in Refs. [153–155]. In lattice systems in which the dimension of the Hilbert space per site is finite, one can show that Page's formula results in a "volume-law" behavior, i.e., the entanglement entropy scales linearly in the volume V_A of the subsystem, $S_A \propto V_A$ (for a large system of volume V and a subsystem with $V_A < V/2$). The prefactor of this volume law is the same that was later found in studies of highly excited eigenstates of nonintegrable Hamiltonians and, separately, within random matrix theory. Deviations from Page's result have been discussed in the context of highly excited eigenstates of number-preserving Hamiltonians away from half-filling [43,46]. The entanglement entropy of pure quantum states with particle-number conservation was also studied in Refs. [47,49].

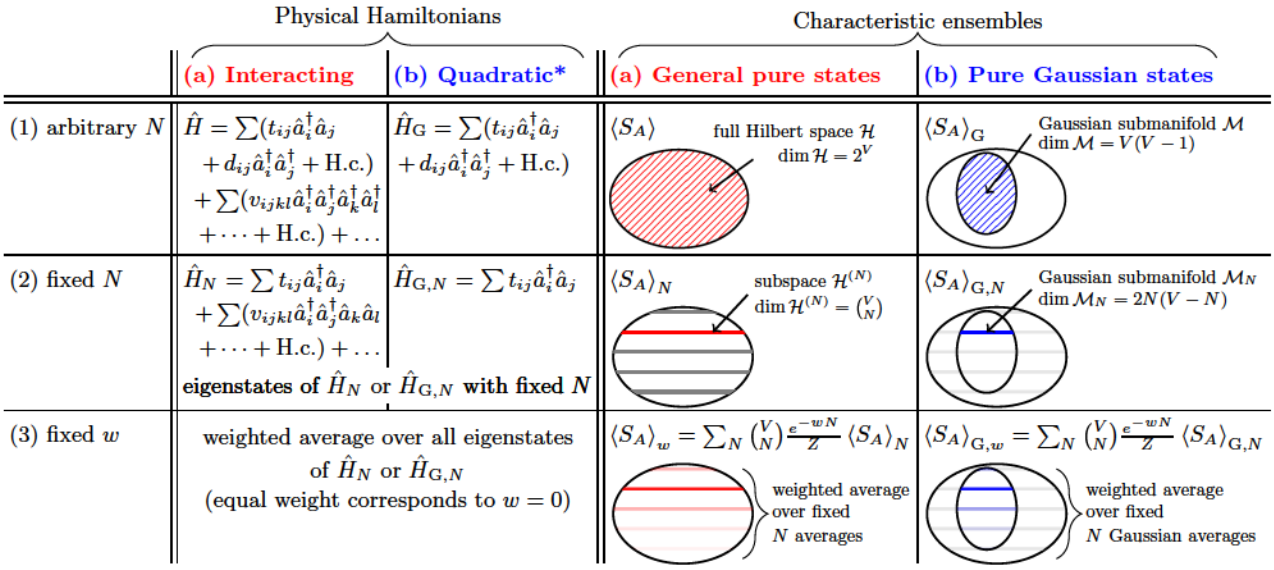
In many physically relevant situations, constraints such as particle-number conservation are present and, as a result, the Hilbert space of the system does not factor into the tensor product of Hilbert spaces of subsystems. Other notable examples are gauge theories that have a Gauss constraint, and quantum gravitational systems where the Hamiltonian itself is a quantum constraint [156]. When the constraint is additive over subsystems, i.e., of the form $C_A + C_B = 0$, one can resolve the Hilbert space as a direct sum over a tensor product of simultaneous eigenspaces

of C_A and C_B that solve the constraint. An example of this phenomenon is the previously mentioned systems with a fixed total number of particles, where C_A is the number of particles in the subsystem. For this class of systems, a formula for the typical entanglement entropy of pure constrained states (and its variance) was obtained recently by one of us (E.B.) in collaboration with P. Donà [157].

Another important class of pure states are fermionic Gaussian states. They are of special interest because, as mentioned before, the many-body eigenstates of quadratic Hamiltonians are Gaussian states. In the quantum computing community those states are known as matchgate states, and they are used to implement classical computations on a quantum computer [158]. The average entanglement entropy over translationally invariant Gaussian states was studied by four of us (E.B., L.H., M.R., and L.V.) in Refs. [42,48,50], where tight bounds were derived. The average over all fermionic Gaussian states was studied later by three of us (E.B., L.H., and M.K.) in Ref. [159], and a closed-form expression was derived as a function of the total (V) and the subsystem (V_A) volumes. To leading order in V_A, that expression agreed with the one previously derived in Ref. [61] for the average entropy over all eigenstates of particle-number-preserving random quadratic Hamiltonians that, in turn, was shown to agree with the average over random quadratic Hamiltonians without particle-number conservation in Ref. [62]. Entanglement entropies of Gaussian states with particle-number conservation were also studied in Refs. [86,88] in the context of SYK models. We connect all these results throughout this tutorial.

G. Outline

We provide a detailed understanding of the behavior of the entanglement entropy of typical pure quantum states in (a) the entire Hilbert space (Sec. II) and (b) the subset of Gaussian states (Sec. III). Both for pure quantum states in the entire Hilbert space, and within the subset of Gaussian states, we consider case (1) in which the number of particles is not fixed separately from cases (2) in which it is fixed and (3) in which we take a weighted sum over all fixed particle-number sectors. Overall, we thus consider *six* characteristic ensembles, which we can compare with the respective typical eigenstate entanglement entropy found in physical Hamiltonians, as illustrated in Fig. 2. For general states without fixed particle number, the results for the average over all quantum states are well known [152], while for general Gaussian states, the results were recently obtained by three of us (E.B., L.H., and M.K.) [159]. For quantum states with a fixed number of particles, all the results discussed are derived here (the derivations are explained in detail in the appendices). We identify and explain the qualitative changes that occur in subleading terms when one fixes the number of



* Numerical results have shown that integrable interacting systems show qualitatively similar behavior [55].

FIG. 2. Physical Hamiltonians versus characteristic ensembles. This tutorial relates the typical entanglement entropy of eigenstates of physical Hamiltonians (left) to the typical entanglement entropy computed analytically for six characteristic ensembles constructed from random matrix theory (right). On the left we show four representative physical Hamiltonians, (a) generic interacting and (b) quadratic, (1) without and (2) with number conservation. On the right we show the six characteristic ensembles, which are obtained considering three sets of eigenstates [(1) arbitrary N, (2) fixed N, and (3) weighted average] for (a) general pure states and (b) pure Gaussian states.

particles in the quantum states. We explore the behavior of the average entanglement entropy as a function of the system volume V, the subsystem volume V_A, and the total number of particles N. We show that general pure states and Gaussian states share common properties for $f = V_A/V \ll V$, but become increasingly distinct as $f \to \frac{1}{2}$.

In Sec. IV, we examine the relation between typical entanglement entropies in the Hilbert space and typical entanglement entropies generated by random matrices. In Sec. V, we examine the relation between typical entanglement entropies in the Hilbert space and typical entanglement entropies of eigenstates of specific model Hamiltonians. While we mostly have fermionic lattice systems (with and without spin) in mind, our results apply to any fermionic system with a well-defined bipartition, hard-core bosons, and spin-$\frac{1}{2}$ systems. In the outlook, we discuss how the same methods can be used to study regular (soft-core) bosons, large spins, and systems of distinguishable particles.

It was recently conjectured that the entanglement entropy of typical eigenstates of quantum many-body Hamiltonians can be used as a diagnostic of quantum chaos and integrability [55]. This conjecture was motivated by the finding that the leading volume-law term of the entanglement entropy in typical eigenstates of an integrable model that is not mappable onto a noninteracting one behaves qualitatively (and quantitatively) like in typical eigenstates of translationally invariant quadratic Hamiltonians, and in stark contrast to the behavior in typical eigenstates of quantum-chaotic Hamiltonians. With that conjecture in mind, we expect that the analytical expressions derived in the sections to follow can be used as benchmarks for numerical results obtained for physical Hamiltonians (and analytical results obtained using different techniques). The entanglement entropy of typical eigenstates of physical Hamiltonians is complementary to diagnostics of integrability and quantum chaos currently in use, such as the previously mentioned Wigner surmise that is based on the statistical properties of the eigenenergies.

II. GENERAL PURE STATES

We consider the general setting of a system with V fermionic modes (potentially arranged in a lattice of arbitrary dimension) and a bipartition of the system into a subsystem A (with V_A fermionic modes) and its complement B (with $V - V_A$ fermionic modes). Hence, the Hilbert space has the structure $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ and dimension $\dim \mathcal{H} = d_A d_B$, where $d_A = \dim \mathcal{H}_A$ and $d_B = \dim \mathcal{H}_B$.

This setup can be used in different contexts. For instance, on a D-dimensional lattice with L sites per space direction and N_{int} fermionic modes per site (representing internal degrees of freedom, such as spin), the total number of fermionic modes is

$$V = L^D N_{\text{int}}. \tag{7}$$

For each mode, we can denote the creation (annihilation) operators as $\hat{f}_i^\dagger$ ($\hat{f}_i$), where $1 \le i \le V$. The bipartition with respect to a subsystem of $V_A \le V$ modes then yields the Hilbert space dimensions

$$d_A = 2^{V_A}, \qquad d_B = 2^{V - V_A}. \tag{8}$$

When we fix the subsystem fraction $f = V_A/V$, our results hold in full generality for arbitrary space dimensions D and fermions with an arbitrary spin. They also hold for hard-core bosons and spin-$\frac{1}{2}$ systems (which have two states per lattice site)—but not for (soft-core) bosons for which the average entanglement entropy diverges due to the infinite local dimension of the Hilbert space.

We note that, in the Introduction, we referred to V, V_A, and $V - V_A$ as volumes. We continue to use that terminology in the rest of the tutorial. Readers should keep in mind that V, V_A, and $V - V_A$ quantify the number of fermionic modes; the volume (or, similarly, the number of sites) being just one of the possible ways in which this is achieved.

A. Arbitrary number of particles

A natural approach to gain an understanding of the entanglement entropy of pure quantum states is to consider randomly chosen vector states $|\psi\rangle \in \mathcal{H}$. Since the Hilbert space is finite dimensional, the set of all states describes a hypersphere that is compact. Thus, the uniform distribution on this set provides a natural unbiased probability distribution over pure states [152]. The corresponding density operator $|\psi\rangle\langle\psi|$ is also called Haar random as it is equal to the induced Haar measure on the coset $U(d_A d_B)/U(d_A d_B - 1)$ with $U(d)$ the d-dimensional unitary group. This measure is the unique one that is normalized and invariant under unitary transformations. In practice, we can construct such a state by first fixing a reference state $|\psi_0\rangle$ and then defining $|\psi\rangle = U|\psi_0\rangle$, where $U : \mathcal{H} \to \mathcal{H}$ is a random unitary transformation drawn from the Haar measure of unitary matrices.

For every such $|\psi\rangle$, the bipartite entanglement entropy with respect to the tensor product decomposition $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is, once again, defined by

$$S_A(|\psi\rangle) = -\mathrm{Tr}_{\mathcal{H}_A}(\hat{\rho}_A \ln \hat{\rho}_A), \quad \text{with } \hat{\rho}_A = \mathrm{Tr}_{\mathcal{H}_B} |\psi\rangle\langle\psi|, \tag{9}$$

where $\mathrm{Tr}_{\mathcal{H}_B}$ refers to the partial trace over the sub-Hilbert space $\mathcal{H}_B$. We are interested in the average and the variance of S_A with respect to the statistical ensemble. In Page's

setting, the ensemble is given by Haar-random density operators $|\psi\rangle\langle\psi|$. We can write those quantities as

$$\langle S_A\rangle = \int S_A(U|\psi_0\rangle)d\mu(U), \tag{10}$$

$$(\Delta S_A)^2 = \int [S_A(U|\psi_0\rangle) - \langle S_A\rangle]^2 d\mu(U), \tag{11}$$

where $d\mu(U)$ represents the normalized Haar measure over the unitary group $U(d_A d_B)$. The average and variance can be computed from the joint probability distribution of the eigenvalues λ_i of $\hat{\rho}_A$, which were derived in Ref. [151], as $S_A(|\psi\rangle) = -\sum_{j=1}^{d_A}\lambda_j \ln(\lambda_j)$.

In the case of Page's setting, where one draws a random uniformly distributed state $|\psi\rangle \in \mathcal{H}$, the average entanglement entropy (and any other statistical quantities) will depend only on the dimensions d_A and d_B. This shows that the ensuing statements are independent of any chosen particle statistics (bosons, fermions) and also not restricted to the special case of qubit-based (two-state per-site) systems in the context of which they are usually used. They apply much more generally.

1. Statistical ensemble of states

Let us outline the derivation of Page's result in Eq. (23). The derivation allows us to draw some relations to random matrix theory, and we use some of the ideas behind it later when studying fermionic Gaussian states.

Let us consider a general state vector in the tensor space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Such a vector can always be decomposed into some factorizing orthonormal basis $|a\rangle \otimes |b\rangle \in \mathcal{H}$:

$$|\psi\rangle = \sum_{a=1}^{d_A}\sum_{b=1}^{d_B} w_{ab}\,|a\rangle \otimes |b\rangle \tag{12}$$

with coefficients $w_{ab} \in \mathbb{C}$. As physical state vectors are normalized, the coefficients satisfy

$$\sum_{a=1}^{d_A}\sum_{b=1}^{d_B} |w_{ab}|^2 = 1. \tag{13}$$

These coefficients and their distribution contain all the information about the random pure state and, hence, about the entanglement entropy. Once we take the partial trace over the subsystem B, the density operator in system A is explicitly given by

$$\hat{\rho}_A = \sum_{a_1,a_2=1}^{d_A}\sum_{b=1}^{d_B} w_{a_1 b} w^*_{a_2 b}\,|a_1\rangle\langle a_2|\,. \tag{14}$$

The coefficients w_{ab} can be seen as the entries of a $d_A \times d_B$ complex matrix W. Thence, we can identify the density operator $\hat{\rho}_A = WW^\dagger$. The density operator for subsystem B, when tracing over subsystem A, is given by $\hat{\rho}_B = \mathrm{Tr}_{\mathcal{H}_A}|\psi\rangle\langle\psi| = W^\dagger W$. This allows us to explicitly see the duality between the two subsystems, and what changes when switching from A to B. In general, the dimensions d_A and d_B are not equal. Thus, one of the two density operators always has zero eigenvalues but otherwise comprises the very same eigenvalues with the same multiplicity as the other density operator.

Equation (13) implies that the matrix W satisfies $\mathrm{Tr}\,WW^\dagger = 1$. Page's setting of uniformly distributed states means that W is distributed uniformly on the unit sphere described by this normalization condition. Such a matrix is a random matrix and the ensemble is known as the fixed trace ensemble [117]. In quantum information theory, it has been used in several studies, e.g., such as those in Refs. [143,145,149,160–162].

With this knowledge at hand, let us compute the entanglement entropy, which can be expressed in terms of the eigenvalues of $WW^\dagger$, i.e., it holds that

$$\begin{aligned} S_A(|\psi\rangle) &= -\mathrm{Tr}[WW^\dagger \ln(WW^\dagger)] \\ &= -\mathrm{Tr}[W^\dagger W \ln(W^\dagger W)] \\ &= S_B(|\psi\rangle). \end{aligned} \tag{15}$$

It is the spectral decomposition theorem that ensures the symmetry of the entanglement entropy between the two subsystems. This symmetry always holds.

To compute the ensemble average of $S_A(|\psi\rangle)$, one implements the normalization condition in terms of a Dirac delta function, which can be written as a Fourier-Laplace transform of the complex Wishart-Laguerre ensemble [117,136],

$$\begin{aligned} \langle S_A\rangle &= -\frac{\int_{\mathbb{C}^{d_A\times d_B}} d[W] S_A(|\psi\rangle)\delta(1-\mathrm{Tr}\,WW^\dagger)}{\int_{\mathbb{C}^{d_A\times d_B}} d[W]\delta(1-\mathrm{Tr}\,WW^\dagger)} \\ &= -\frac{\int_{\mathbb{C}^{d_A\times d_B}} d[W]\int_{-\infty}^{\infty} dt S_A(|\psi\rangle) e^{(1+it)(1-\mathrm{Tr}\,WW^\dagger)}}{\int_{\mathbb{C}^{d_A\times d_B}} d[W]\int_{-\infty}^{\infty} dt e^{(1+it)(1-\mathrm{Tr}\,WW^\dagger)}}, \end{aligned} \tag{16}$$

where $d[W]$ is the product of all differentials of all matrix elements of W and $W^\dagger$. The shift of it to $1+it$ is important since it allows us to rescale $WW^\dagger \to WW^\dagger/(1+it)$, which describes a rotation of the integration contours in the complex plane where the integrand is absolutely integrable. Before rescaling, we use a standard trick to rewrite the entanglement entropy removing the logarithm:

$$S_A(|\psi\rangle) = -\partial_\epsilon\,\mathrm{Tr}(WW^\dagger)^\epsilon|_{\epsilon=1}. \tag{17}$$

We also use this relation when computing the average entanglement entropy of fermionic Gaussian states. After

the aforementioned rescaling of the random matrix $WW^\dagger$, the entanglement entropy reads

$$\langle S_A\rangle = -\partial_\epsilon\left[\frac{\int_{-\infty}^{\infty} dt(1+it)^{-d_Ad_B-\epsilon}e^{(1+it)}}{\int_{-\infty}^{\infty} dt(1+it)^{-d_Ad_B}e^{(1+it)}} \times \frac{\int_{\mathbb{C}^{d_A\times d_B}} d[W]\,\mathrm{Tr}(WW^\dagger)^\epsilon e^{-\mathrm{Tr}\,WW^\dagger}}{\int_{\mathbb{C}^{d_A\times d_B}} d[W] e^{-\mathrm{Tr}\,WW^\dagger}}\right]_{\epsilon=1}. \quad (18)$$

The first factor can be computed using standard techniques in complex analysis, yielding

$$\frac{\int_{-\infty}^{\infty} dt(1+it)^{-d_Ad_B-\epsilon}e^{(1+it)}}{\int_{-\infty}^{\infty} dt(1+it)^{-d_Ad_B}e^{(1+it)}} = \frac{\Gamma[d_Ad_B]}{\Gamma[d_Ad_B+\epsilon]} \quad (19)$$

with $\Gamma[x]$ the gamma function. Once we apply the derivative in ϵ, we see that the first term of Page's result (23) follows from the fixed trace condition. This was also observed and exploited in the original works in which Eq. (23) was computed [152,154].

The remaining integral over W is an average over the complex Wishart-Laguerre ensemble [117,136], which is one of the three classical random matrix ensembles that also include the Gaussian [117,118,136] and the Jacobi [136,163] ensembles. Interestingly, it is the Jacobi ensemble that we encounter when studying fermionic Gaussian states later on.

In the final step, one uses the eigenvalues $x_1,\ldots,x_{d_A}\geq 0$ of $WW^\dagger$ and expresses the average in terms of the level density of the Laguerre ensemble. In this step, one needs to decide whether $d_A\leq d_B$ or $d_A\geq d_B$, as the density is not analytic at $d_A=d_B$. This reflects the fact that either $\hat\rho_A$ or $\hat\rho_B$ has exact zero eigenvalues, and it is the source of the case distinction in Page's result (23). When assuming that $d_A\leq d_B$, the level density is equal to [136]

$$R_{1,\mathrm{Lag}}(x) = \frac{d_A!}{(d_B-1)!}x^{d_B-d_A}e^{-x} \times [L_{d_A-1}^{(d_B-d_A+1)}(x)L_{d_A-1}^{(d_B-d_A)}(x) - L_{d_A-2}^{(d_B-d_A+1)}(x)L_{d_A}^{(d_B-d_A)}(x)], \quad (20)$$

in terms of the generalized Laguerre polynomials $L_n^{(\alpha)}(x)$. Using the series representation of $L_a^{(d_B-d_A+1)}(x)$ in Eq. (18.5.12) of Ref. [164], and the Rodrigues formula for $L_b^{(d_B-d_A)}(x)$ in Eq. (18.5.5) of Ref. [164], one can show for $a\leq b$ that

$$\int_0^\infty x^{\epsilon+\alpha}e^{-x}L_a^{(d_B-d_A+1)}(x)L_b^{(d_B-d_A)}(x)dx = \sum_{l=0}^{a}\frac{\Gamma[a+\alpha+2]\Gamma[\epsilon+l+1]\Gamma[\epsilon+\alpha+l+1]}{\Gamma[l+\alpha+2]\Gamma[\epsilon+l-n+1](a-l)!\,l!\,b!} \times(-1)^{l+b}, \quad (21)$$

which is different from the formula used in Ref. [154]. We use this approach as it parallels our computation for Gaussian states. Putting everything together in Eq. (18), using the identity

$$\frac{\int d[W]\,\mathrm{Tr}(WW^\dagger)^\epsilon e^{-\mathrm{Tr}\,WW^\dagger}}{\int d[W]e^{-\mathrm{Tr}\,WW^\dagger}} = d_A\int_0^\infty x^\epsilon R_{1,\mathrm{Lag}}(x)dx, \quad (22)$$

we arrive at Page's result in Eq. (23). As we have mentioned before, the symmetry in d_A and d_B reflecting the symmetry in the two subsystems $A\leftrightarrow B$ has to be implemented by hand. The random matrix approach underscores this loss of analyticity when going over to the generic nonzero eigenvalues of $WW^\dagger$ and selecting the smaller of the two dimensions.

2. *Average and variance*

The average entanglement entropy of a uniformly distributed pure state in $\mathcal{H}$ restricted to subsystem A is given by the Page formula [152]

$$\langle S_A\rangle = \begin{cases}\Psi(d_Ad_B+1)-\Psi(d_B+1)-\dfrac{d_A-1}{2d_B}, & d_A\leq d_B,\\ \Psi(d_Ad_B+1)-\Psi(d_A+1)-\dfrac{d_B-1}{2d_A}, & d_A>d_B,\end{cases} \quad (23)$$

where $\Psi(x)=\Gamma'(x)/\Gamma(x)$ is the digamma function. In the thermodynamic limit $V\to\infty$ when $V_A, V-V_A\to\infty$ also so that the subsystem fraction

$$f=\frac{V_A}{V} \quad (24)$$

is fixed, Page's formula (23) reduces to

$$\langle S_A\rangle = f\ V\ln 2 - 2^{-|1-2f|V-1}+O(2^{-V}), \quad (25)$$

where we will be careful to consistently use Landau's "big O" and "little o" notation in this manuscript, such that

$$f(V)=O(V^n) \iff \lim_{V\to\infty}\frac{f(V)}{V^n}=c\neq 0, \quad (26)$$

and

$$f(V)=o(V^n) \iff \lim_{V\to\infty}\frac{f(V)}{V^n}=0. \quad (27)$$

The first term in Eq. (25) is a volume law: the average entanglement entropy scales as the minimum between the volumes $V_A=fV$ and $V_B=(1-f)V$. For $f\neq\frac{1}{2}$, the second term is an exponentially small correction. In fact, at fixed f and in the limit $V\to\infty$, the second term $-2^{-|1-2f|V-1}$ becomes $-\frac{1}{2}\delta_{f,1/2}$. We can also resolve precisely how this Kronecker delta arises in the neighborhood

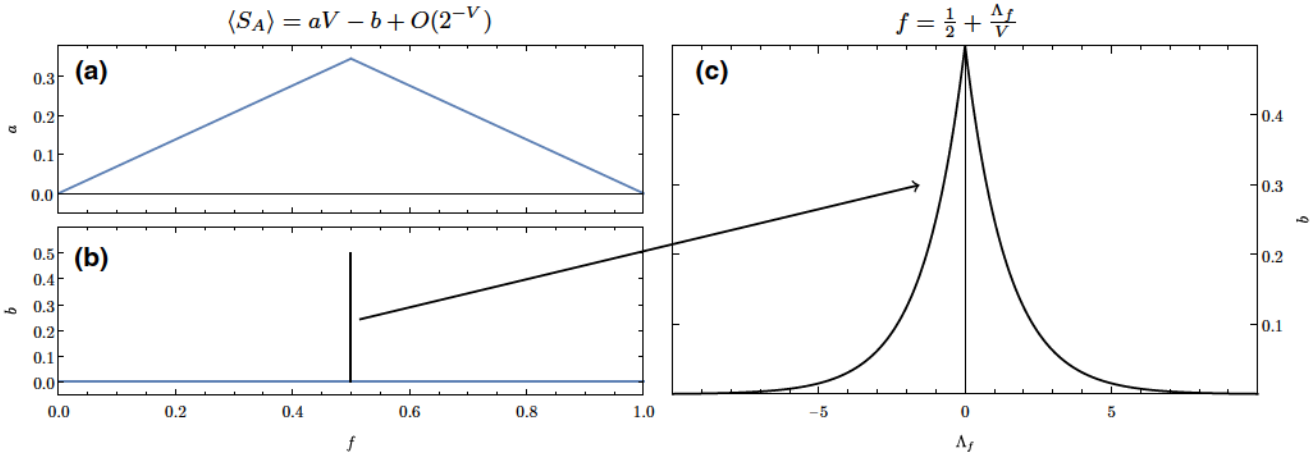


FIG. 3. The average entanglement entropy $\langle S_A\rangle = aV - b + o(1)$ as a function of the subsystem fraction $f = V_A/V$ for large V. (a) Leading-order behavior, also known as the Page curve. (b) The constant correction, which is given by a Kronecker delta $-\frac{1}{2}\delta_{f,1/2}$. This Kronecker delta is resolved in (c) by carrying out a double scaling limit $V \to \infty$ with $f = V_A/V = \frac{1}{2} + \Lambda_f/V$.

of $f = \frac{1}{2}$. As it may be difficult to reach exactly $f = \frac{1}{2}$ in physical experiments, the more precise statement is that we see the correction whenever $f = \frac{1}{2} + O(1/V)$. Formally, we can thus resolve the correction term exactly as $-2^{-|\Lambda_f|-1}$ for $f = \frac{1}{2} + \Lambda_f/V$, as visualized in Fig. 3.

We find similar Kronecker delta contributions $\delta_{f,1/2}$ in subsequent sections where we discuss the typical entropy at fixed particle number and in the setting of Gaussian states. These terms highlight nonanalyticities in the entanglement entropy that can be resolved by double scaling limits. Those "critical points" occur at symmetry points and along axes. In the present case, this has happened with the dimensions d_A and d_B reflecting whether the density operator $\hat{\rho}_A = WW^\dagger$ or $\hat{\rho}_B = W^\dagger W$ contains generic zero eigenvalues.

The variance of the entanglement entropy of a random pure state is given by the exact formula (for $d_A \leq d_B$) [157,165,166]

$$(\Delta S_A)^2 = \frac{d_A + d_B}{d_A d_B + 1}\Psi'(d_B + 1) - \Psi'(d_A d_B + 1) - \frac{(d_A - 1)(d_A + 2d_B - 1)}{4d_B^2(d_A d_B + 1)}, \tag{28}$$

where $\Psi'(x) = d\Psi(x)/dx = d^2[\ln\Gamma(x)]/dx^2$ is the first derivative of the digamma function. It can be derived using similar techniques as those outlined above for the average. In particular, the fixed trace condition can be separated as before via the trick of the Fourier-Laplace transform, such that one is left with an average over the complex Wishart-Laguerre ensemble. The derivation is tedious and lengthy because one has to deal with double sums, which can be computed as described in Appendix D [167].

In the thermodynamic limit discussed above, Eq. (28) reduces to

$$(\Delta S_A)^2 = \left(\tfrac{1}{2} - \tfrac{1}{4}\delta_{f,\frac{1}{2}}\right)2^{-(1+|1-2f|)V} + o(2^{-(1+|1-2f|)V}). \tag{29}$$

This shows that the variance is exponentially small in V. As a result, in the thermodynamic limit the entanglement entropy of a typical state is given by Eq. (25) [157].

Anew, one could resolve the variance at the critical point $f = \frac{1}{2}$ via a double scaling limit $f = \frac{1}{2} + \Lambda_f/V$. This yields $(\Delta S_A)^2 = 2^{-V}2^{-2|\Lambda_f|-1}(1 - 2^{-2|\Lambda_f|-1})$.

B. Fixed number of particles

Let us go over to a Hilbert space $\mathcal{H}^{(N)}$ with a fixed number of particles, but still carrying over the idea to draw states uniformly from the sphere in this Hilbert space. We further assume that there is a notion of a bipartition into subsystem A and B, such that one can specify for each particle if it is in subsystem A or B. Such a decomposition is not a simple tensor product anymore, but it is a direct sum of tensor products

$$\mathcal{H}^{(N)} = \bigoplus_{N_A=0}^{N} \left(\mathcal{H}_A^{(N_A)} \otimes \mathcal{H}_B^{(N-N_A)}\right). \tag{30}$$

The direct sum is over the occupation number in A (which labels the center of the subalgebra). Each summand represents those states where N_A particles are in subsystem A and $N - N_A$ particles are in subsystem B (assuming indistinguishable particles).

When N_A is larger than dimension V_A of subsystem A, or $N - N_A$ is larger than $V - V_A$, we consider the tensor product $\mathcal{H}_A^{(N_A)} \otimes \mathcal{H}_B^{(N-N_A)}$ as the empty set and, thence,

nonexistent. This is the case as, due to Pauli's exclusion principle, we cannot put more fermions in the system than there are quantum states. We also adapt this understanding for the following discussion where direct sums, ordinary sums, and products are reduced to the components that are actually present.

1. Statistical ensemble of states

Let us consider fermionic creation $\hat{f}_i^\dagger$ and annihilation $\hat{f}_i$ operators, which satisfy the anticommutation relations $\{\hat{f}_i,\hat{f}_j^\dagger\}=\delta_{ij}$, $\{\hat{f}_i,\hat{f}_j\}=0$ with $i,j=1,\ldots,V$. The corresponding number operators are

$$\hat{N}=\sum_{i=1}^{V}\hat{f}_i^\dagger\hat{f}_i,\quad \hat{N}_A=\sum_{i=1}^{V_A}\hat{f}_i^\dagger\hat{f}_i,\quad \hat{N}_B=\sum_{i=V_A+1}^{V}\hat{f}_i^\dagger\hat{f}_i, \tag{31}$$

where one can see that

$$\hat{N}=\hat{N}_A+\hat{N}_B. \tag{32}$$

The Hilbert space of the system can be decomposed as a direct sum of Hilbert spaces at fixed eigenvalue N of $\hat{N}$,

$$\mathcal{H}=\bigotimes_{i=1}^{V}\mathcal{H}_i=\bigoplus_{N=0}^{V}\mathcal{H}^{(N)} \tag{33}$$

with $\mathcal{H}^{(N)}$ given by Eq. (30). The dimension of each N-particle sector is

$$d_N=\dim\mathcal{H}^{(N)}=\frac{V!}{N!\,(V-N)!}. \tag{34}$$

It is immediate to check that $\dim\mathcal{H}=\sum_{N=0}^{V}d_N=2^V$. Similarly, one can use the number operators $\hat{N}_A$ and $\hat{N}_B$ to decompose the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ into sectors

$$\mathcal{H}_A=\bigoplus_{N_A=0}^{V_A}\mathcal{H}_A^{(N_A)},\qquad \mathcal{H}_B=\bigoplus_{N_B=0}^{V-V_A}\mathcal{H}_B^{(N_B)}. \tag{35}$$

Let us stress once again that, while $\mathcal{H}$ is a tensor product over A and B,

$$\mathcal{H}=\left(\bigotimes_{i=1}^{V_A}\mathcal{H}_i\right)\otimes\left(\bigotimes_{i=V_A+1}^{V}\mathcal{H}_i\right)=\mathcal{H}_A\otimes\mathcal{H}_B, \tag{36}$$

the sector at fixed number $N\le V_A$ is not a tensor product. It is the direct sum of tensor products from Eq. (30). The corresponding dimensions of the subsystems are

$$\begin{aligned} d_A(N_A)&=\dim\mathcal{H}_A^{(N_A)}=\frac{V_A!}{N_A!\,(V_A-N_A)!},\\ d_B(N_B)&=\dim\mathcal{H}_B^{(N_B)}=\frac{(V-V_A)!}{N_B!\,[(V-V_A)-N_B]!}. \end{aligned} \tag{37}$$

One can check that the dimensions add up correctly,

$$\sum_{N_A=0}^{N}d_A(N_A)\,d_B(N-N_A)=\frac{V!}{N!\,(V-N)!}=d_N. \tag{38}$$

The formula for d_A, and equivalently that of d_B, follows from a simple counting argument of how many choices there are to place N_A indistinguishable particles on V_A modes. Let us underline that it does not matter what we label particles and what holes. Note that $d_A(N_A)$ or $d_B(N-N_A)$ will vanish for N_A outside of the interval $[\max(0,N+V_A-V),\min(N,V_A)]$, but we will not truncate the sum, as we will soon turn it into a Gaussian integral.

From these dimensions we can readily read off two exact symmetries.

(i) It does not matter whether one considers subsystem A or B. One can exchange $(d_A(N_A),V_A,N_A)\leftrightarrow(d_B(N-N_A),V-V_A,N-N_A)$. This allows us to restrict the discussion to $V_A\le V/2$. However, the dimensions of the two Hilbert spaces are exchanged, which (as we will show) yields nonanalytic points along $V_A=V/2$ due to the two branches of Page curve (23).

(ii) Additionally, there is a particle-hole symmetry since it does not matter whether one counts particles or holes. Actually, the "particles" do not necessarily need to represent particles but they can be, for instance, up spins while the "holes" are down spins (having in mind spin-$\frac{1}{2}$ systems). Any binary structure with fermion statistics (meaning Pauli principle) can be described in this setting. Mathematically, the particle-hole symmetry is reflected in the exchange $(N,N_A)\leftrightarrow(V-N,V_A-N_A)$. We note that in this case the dimensions are not exchanged so one does not switch branches in Page curve (23). Therefore, the symmetry points at $N=V/2$ will be analytic, as we will also show. This symmetry allows us to restrict $N\le V/2$.

In summary, we only need to study the behavior in the quadrant $(V_A,N)\in(0,V/2]^2$. The remaining quadrants are obtained by symmetry.

Like in the setting in which we do not fix the particle number, we can relate the problem to random matrix theory. Here, we briefly recall the most important ingredients from Ref. [157]. A state $|\psi\rangle\in\mathcal{H}^{(N)}$ can be

again written in a basis. We choose the orthonormal basis vectors $|a, N_A\rangle \otimes |b, N - N_A\rangle \in \mathcal{H}_A^{(N_A)} \otimes \mathcal{H}_B^{(N-N_A)}$ so that the state vector has the expansion

$$|\psi\rangle = \bigoplus_{N_A=0}^{N} \sum_{a=1}^{d_A} \sum_{b=1}^{d_B} \tilde{w}_{ab}^{(N_A)} |a, N_A\rangle \otimes |b, N - N_A\rangle , \quad (39)$$

with the abbreviations $d_A = d_A(N_A)$ and $d_B = d_B(N - N_A)$. The normalization is then reflected by the triple sum

$$\sum_{N_A=0}^{N} \sum_{a=1}^{d_A} \sum_{b=1}^{d_B} |\tilde{w}_{ab}^{(N_A)}|^2 = 1. \quad (40)$$

The direct sum over N_A is important as it tells us that the density operator $\hat{\rho}_A = \mathrm{Tr}_{\mathcal{H}_B} |\psi\rangle \langle\psi|$ has a block diagonal form, namely,

$$\hat{\rho}_A = \bigoplus_{N_A=0}^{N} \sum_{a_1,a_2=1}^{d_A} \sum_{b=0}^{d_B} \tilde{w}_{a_1 b}^{(N_A)} (\tilde{w}_{a_2 b}^{(N_A)})^* |b, N_A\rangle \langle a_2, N_A| . \quad (41)$$

Again, we can understand the coefficients $\tilde{w}_{ab}^{(N_A)} \in \mathbb{C}$ as the entries of a $d_A \times d_B$ matrix $\tilde{W}_{N_A}$. The point is that those matrices are coupled by condition (40). In Ref. [157] those matrices were decoupled by understanding their squared Hilbert-Schmidt norms as probability weights, i.e., defining

$$p_{N_A} = \sum_{a=1}^{d_A} \sum_{b=1}^{d_B} |\tilde{w}_{ab}^{(N_A)}|^2 \in [0, 1] \quad (42)$$

such that $\tilde{W}_{N_A} = \sqrt{p_{N_A}}\, W_{N_A}$. This notation allows one to identify the density operator of subsystem A with the block diagonal matrix $\hat{\rho}_A = \mathrm{diag}(p_0 W_0 W_0^\dagger, \ldots, p_N W_N W_N^\dagger)$, as illustrated in Fig. 4.

Thus, the entanglement entropy becomes the sum

$$S_A(|\psi\rangle) = \sum_{N_A=0}^{N} [p_{N_A} \mathrm{Tr}(W_{N_A} W_{N_A}^\dagger \ln[W_{N_A} W_{N_A}^\dagger]) + p_{N_A} \ln(p_{N_A})]. \quad (43)$$

Anew, the symmetry between the two subsystems is reflected by the spectral decomposition theorem since it holds that $\hat{\rho}_B = \mathrm{Tr}_{\mathcal{H}_A} |\psi\rangle \langle\psi| = \mathrm{diag}(p_0 W_0^\dagger W_0, \ldots, p_N W_N^\dagger W_N)$.

Since the norms are encoded in the probability weights p_{N_A}, each matrix $W_{N_A} W_{N_A}^\dagger$ independently describes a fixed trace ensemble, i.e., $\mathrm{Tr}\, W_{N_A} W_{N_A}^\dagger = 1$. Thus, it can be dealt with in the same way as in Page's case, in particular each of those can be traced back to a complex Wishart-Laguerre ensemble of matrix dimension $d_A \times d_B$. The probability weights $p_{N_A} \in [0, 1]$ are also drawn randomly via the joint probability distribution [157]

$$\frac{\delta(1 - \sum_{N_A=0}^{N} p_{N_A}) \prod_{N_A=0}^{N} p_{N_A}^{d_A d_B - 1} dp_{N_A}}{\int \delta(1 - \sum_{N_A=0}^{N} p_{N_A}) \prod_{N_A=0}^{N} p_{N_A}^{d_A d_B - 1} dp_{N_A}}. \quad (44)$$

The Dirac delta function enforces condition (40), while the factors $p_{N_A}^{d_A d_B - 1}$ are the Jacobians for the polar decomposition of the vectors in $\mathcal{H}_A^{(N_A)} \otimes \mathcal{H}_B^{(N-N_A)}$ into their squared norm p_{N_A} and the direction, which is encoded in W_{N_A}. The normalization of the distribution of p_{N_A} was computed in Ref. [157] and can be deduced by inductively tracing the integrals over p_{N_A} back to Euler's beta integrals in Eq. (5.12.1) of Ref. [164].

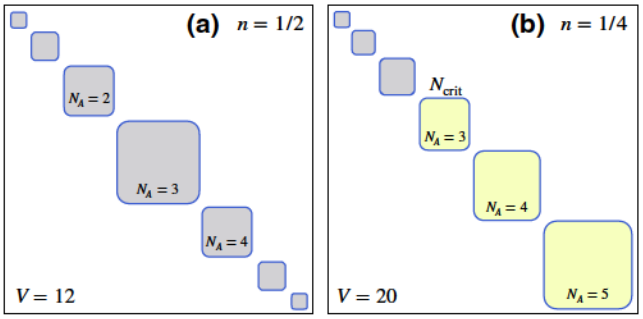


FIG. 4. Sketch of the block dimensions of the reduced density matrix $\hat{\rho}_A$ of subsystem A at the subsystem fraction $f = \frac{1}{2}$. (a) Case $V = 12$ at half filling $n = \frac{1}{2}$, for which $V_A = 6$. The number of particles ranges from $N_A = 0$ to $N_A = 6$, with $N_A = 3$ representing the largest block. (b) Case $V = 20$ at quarter filling $n = \frac{1}{4}$, for which $V_A = 10$. The number of particles ranges from $N_A = 0$ to $N_A = 5$, with $N_A = 5$ representing the largest block. The blocks with $N_A \geq N_{\mathrm{crit}} = 3$ are larger than the corresponding blocks in subsystem B (not shown in the figure).

2. Average and variance

With these definitions and discussions, we are now ready to state the main result in Eq. (23) of Ref. [157]: the average entanglement entropy in system A of a uniformly distributed random state in $\mathcal{H}^{(N)}$ is given by

$$\langle S_A \rangle_N = \sum_{N_A=0}^{\min(N,V_A)} \frac{d_A d_B}{d_N} [\langle S_A \rangle + \Psi(d_N + 1) - \Psi(d_A d_B + 1)], \quad (45)$$

where $d_A = d_A(N_A)$ and $d_B = d_B(N - N_A)$ depend on N_A according to Eq. (37) and $\langle S_A\rangle$ refers to Page's result (23) for given d_A and d_B. Equation (45) follows from the average over $W_{N_A}W_{N_A}^\dagger$ in Eq. (43), which are independent fixed trace random matrices. The prefactor d_Ad_B/d_N, as well as the additional digamma functions, follow from Euler's beta integral in Eq. (5.12.1) of Ref. [164]. In particular, we have used

$$\begin{aligned}\langle p_{N_A}^\epsilon\rangle &= \frac{\int_0^1 p_{N_A}^{\epsilon+d_Ad_B-1}(1-p_{N_A})^{d_N-d_Ad_B-1}dp_{N_A}}{\int_0^1 p_{N_A}^{d_Ad_B-1}(1-p_{N_A})^{d_N-d_Ad_B-1}dp_{N_A}}\\ &= \frac{\Gamma[\epsilon+d_Ad_B]\Gamma[d_N]}{\Gamma[d_Ad_B]\Gamma[\epsilon+d_N]}\end{aligned} \tag{46}$$

for any $\epsilon > -d_Ad_B$. The average on the right-hand side can be obtained by rescaling $p_j \to (1-p_{N_A})p_j$ for any $j \neq N_A$, which decouples the average over p_{N_A} with the remaining probability weights p_j.

We can write Eq. (45) as

$$\langle S_A\rangle_N = \sum_{N_A=0}^{N} \varrho_{N_A}\varphi_{N_A}, \tag{47}$$

by introducing the quantities

$$\begin{aligned}\varrho_{N_A} &= \frac{d_Ad_B}{d_N},\\ \varphi_{N_A} &= \begin{cases}\Psi(d_N+1)-\Psi(d_B+1)-\dfrac{d_A-1}{2d_B}, & d_A \le d_B,\\ \Psi(d_N+1)-\Psi(d_A+1)-\dfrac{d_B-1}{2d_A}, & d_A > d_B\end{cases}\\ &= \Psi(d_N+1)-\Psi(\max(d_A,d_B)+1)-\min\left(\frac{d_A-1}{2d_B},\frac{d_B-1}{2d_A}\right).\end{aligned} \tag{48}$$

The function ϱ_{N_A} can be understood as a probability distribution of having N_A particles in A, with the normalization $\sum_{N_A}\varrho(N_A) = 1$ following from Eq. (38). The function φ_{N_A}, when understood as a continuous function, has a kink at N_{crit}, which refers to the largest integer such that $d_A(N_{\text{crit}}) \le d_B(N - N_{\text{crit}})$. There is only one situation in which N_{crit} is not well defined, namely, when $V_A = N = V/2$ or, equivalently, when $f = n = \frac{1}{2}$ with $f = V_A/V$ and $n = N/V$. Then it always holds that $d_A(N_A) = d_B(N - N_A)$ for all $N_A = 0,\ldots,N$. In this case, we do not need an N_{crit} as the terms in both sums are the same.

We are unable to evaluate this sum exactly, but we can expand $\langle S_A\rangle_N$ in powers of V and approximate the sum by an integral

$$\langle S_A\rangle_N = \sum_{N_A=0}^{N}\varrho_{N_A}\varphi_{N_A} = \int_{-\infty}^{\infty}\varrho(n_A)\varphi(n_A)dn_A + o(1), \tag{49}$$

where $\varrho(n_A)$ is the saddle point approximation of $V\varrho_{n_AV} = Vd_Ad_B/d_N$, which represents the probability distribution for the intensive variable $n_A = N_A/V$. This is enough for computing the leading orders without double scaling. We find the normal distribution

$$\varrho(n_A) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{n_A-\bar{n}_A}{\sigma}\right)^2\right] + o(1) \tag{50}$$

with mean $\bar{n}_A = fn$ and variance $\sigma^2 = f(1-f)n(1-n)/V$.

In Appendix A 1 b, we carefully analyze the difference $\delta n_{\text{crit}} = n_{\text{crit}} - \bar{n}_A$ for $n_{\text{crit}} = N_{\text{crit}}/V$ and find that, for fixed $f < \frac{1}{2}$, one always has $\delta n_{\text{crit}} = O(1)$ and $\delta n_{\text{crit}} > 0$. Thus, for $f \neq \frac{1}{2}$, the center of the Gaussian $\bar{n}_A$ is sufficiently separated from n_{crit}. This allows us to disregard the second sum in Eq. (45) as it is exponentially suppressed. In the case that $f > \frac{1}{2}$, we can disregard the first sum because of the symmetry between the two subsystems A and B.

To find the observable $\varphi(n_A)$ from Eq. (48), we use Stirling's approximation

$$\Psi[d_N+1]-\Psi[\max(d_A,d_B)+1] = \ln\min\left(\frac{d_N}{d_B},\frac{d_N}{d_A}\right) + o(1). \tag{51}$$

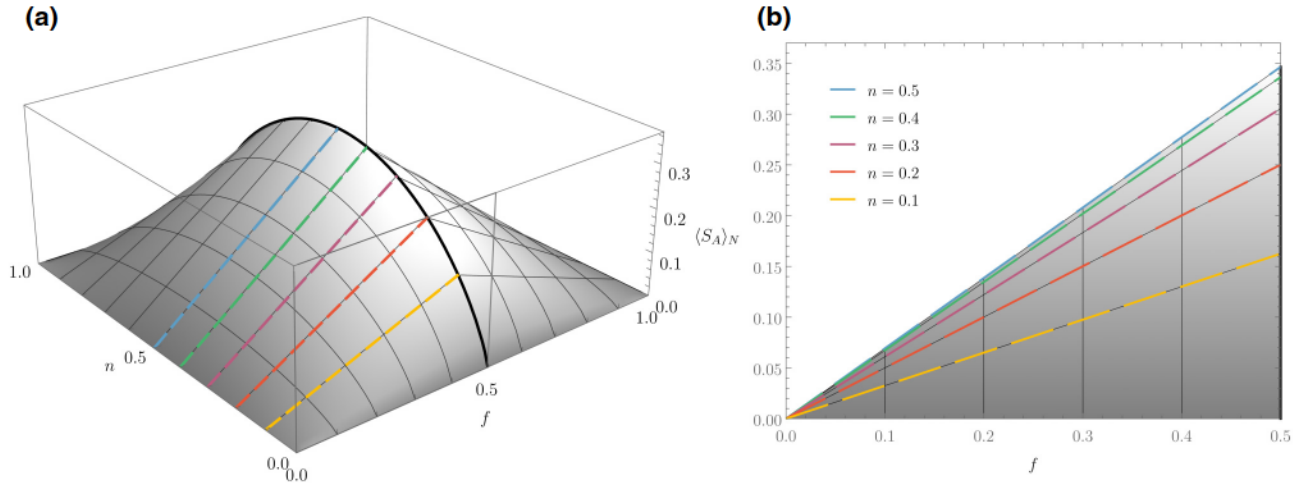


FIG. 5. The leading entanglement entropy $s_A(f,n) = \lim_{V\to\infty} \langle S_A\rangle_N / V$ from Eq. (54) [see Eq. (66)]. For $n = \frac{1}{2}$, $s_A(f,n)$ coincides with Page's result (maximal entanglement). (a) Three-dimensional plot as a function of the subsystem fraction $f = V_A/V$ and the filling ratio $n = N/V$. One can see the mirror symmetries $V_A \to V - V_A$ and $N \to V - N$. (b) Results at fixed n plotted as functions of f. The colored lines agree in both plots so that the right one can be seen as sections of the left one along the colored lines.

Moreover, it holds for $V \gg 1$ and fixed $f \in (0,1)$ that

$$\min\left(\frac{d_A - 1}{d_B}, \frac{d_B - 1}{d_A}\right) = \delta_{f,1/2}\delta_{n,1/2} + o(1). \tag{52}$$

The Kronecker delta is, in fact, a "relic" of a double scaling limit, see Figs. 3(b) and 3(c) for a similar result in the context of Page's setting without fixed particle number. It can be resolved by assuming that f is close to $\frac{1}{2}$ but not exactly at $\frac{1}{2}$; see Appendix A. When collecting all terms up to order $O(1)$, we obtain

$$\begin{aligned}\varphi(n_A) = [&n_A \ln(n_A) - f\,\ln(f) - n\ln[(1-n)/n] \\ &- \ln(1-n) + (f - n_A)\ln(f - n_A)]V \\ &+ \frac{1}{2}\ln\left[\frac{n_A(f-n_A)}{f(1-n)n}\right] - \frac{1}{2}\delta_{f,1/2}\delta_{n,1/2} + o(1)\end{aligned} \tag{53}$$

for $n_A \geq n_{\text{crit}}$. For $n_A \leq n_{\text{crit}}$, we need to apply the symmetries $n_A \to n - n_A$ and $f \to 1 - f$ in expansion (53).

In the limit $V \to \infty$, Gaussian (50) narrows because the standard deviation scales like $\sigma \sim 1/\sqrt{V}$. We can, therefore, expand $\varphi(n_A)$ in powers of $(n_A - \bar{n}_A)$ around the mean $\bar{n}_A$. In order to find the average up to a constant order, it suffices to expand up to the quadratic order and then calculate integral (49). Only for $f = \frac{1}{2}$, we have $\delta n_{\text{crit}} = o(1)$, so that we need to take into account the nonanalyticity in $\varphi(n_A)$ introduced by the symmetry when exchanging the two subsystems. In this case, we integrate two different Taylor expansions for $n_A \leq n/2$ and $n_A \geq n/2$, which will introduce a term of order $\sqrt{V}$, as discussed below.

Combining these results, we arrive at the main result of this subsection,

$$\begin{aligned}\langle S_A\rangle_N = &[(n-1)\ln(1-n) - n\ln(n)]fV \\ &- \sqrt{\frac{n(1-n)}{2\pi}}\left|\ln\left(\frac{1-n}{n}\right)\right|\delta_{f,1/2}\sqrt{V} \\ &+ \frac{f + \ln(1-f)}{2} - \frac{1}{2}\delta_{f,1/2}\delta_{n,1/2} + o(1),\end{aligned} \tag{54}$$

valid for $f \leq \frac{1}{2}$. The leading, volume-law, term in Eq. (54) is the same as that obtained in Refs. [43,46] using random matrix theory, and the same as in Ref. [157] [see Eq. (27)], where it is interpreted as the typical entanglement entropy in the (highly degenerate) eigenspace of a Hamiltonian of the form $\hat{H} = \hat{N} = \hat{N}_A + \hat{N}_B$. The subleading $\sqrt{V}$ term was first discussed in Ref. [43], specifically, it coincides with the bound for such a term computed at $f = \frac{1}{2}$ [43]. It is remarkable that, for $n \neq \frac{1}{2}$, the constant term is nothing but that obtained in Ref. [43] within a "mean field" calculation, while at $n = f = \frac{1}{2}$ the extra $-\frac{1}{2}$ correction was found in Ref. [43] numerically, both for random states as well as for eigenstates of a nonintegrable Hamiltonian. We had all the ingredients to guess the general form in Eq. (54). Its actual derivation with all the details fills several pages, and can be found in Appendix A. A visualization of the leading term in Eq. (54) can be found in Fig. 5.

An important question concerns the resolution of the Kronecker deltas in Eq. (54), which indicate nontrivial scaling limits. The Kronecker deltas are only obtained along the critical line $f = \frac{1}{2}$, which contains a multicritical point at $n = \frac{1}{2}$ when $V \to \infty$. One needs to take

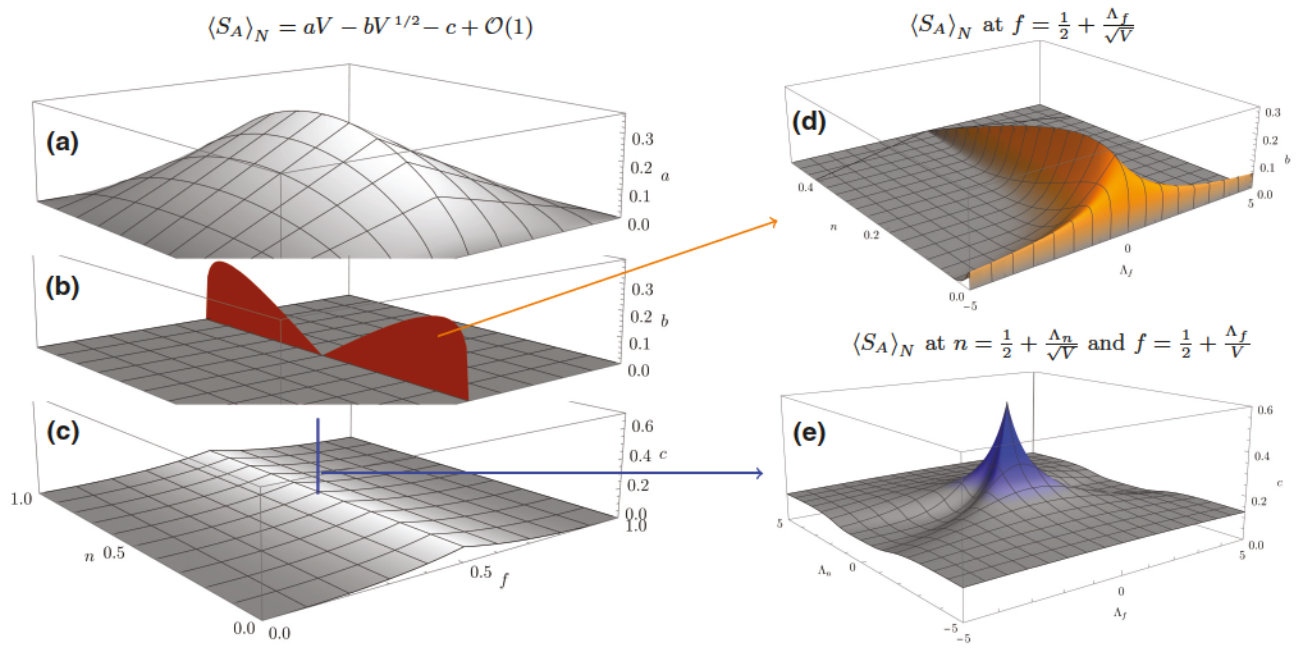


FIG. 6. The entanglement entropy $\langle S_A\rangle_N$ from Eq. (54) as viewed from the contributions of the first three terms in the expansion in V. (a)–(c) Three-dimensional plots as functions of the subsystem fraction $f = V_A/V$ and the filling ratio $n = N/V$. (d) Resolving the expansion coefficient b for $f = \frac{1}{2} + \Lambda_f/\sqrt{V}$ around $f = \frac{1}{2}$, as given by Eq. (A66), approaching zero for large $|\Lambda_f|$. (e) Resolving the expansion coefficient c for $n = \frac{1}{2} + \Lambda_n/\sqrt{V}$ and $f = \frac{1}{2} + \Lambda_f/V$ around $f = n = \frac{1}{2}$, as given by Eq. (A68), approaching $(2\ln 2 - 1)/4$ for large $|\Lambda_f|$ or $|\Lambda_n|$. We underline that the subleading contributions are multiplied by a minus sign.

the resolution into account because experiments are carried out in finite systems in which f and n can only be fixed within some experimental resolution. Consequently, it is important to understand within which margin of error one needs to choose f and n to observe the corresponding terms. This question can be answered by analyzing the limit $V \to \infty$ in the double scaling $f = \frac{1}{2} + V^{-\alpha}\Lambda_f$ and/or $n = \frac{1}{2} + V^{-\beta}\Lambda_n$. We find that the $\sqrt{V}$ correction in Eq. (54) (for fixed n) becomes visible for $\alpha = \frac{1}{2}$, i.e., whenever the difference between f and $\frac{1}{2}$ is of order $1/\sqrt{V}$ or smaller. The constant correction requires a more detailed analysis as it depends on the relative scaling of both f and n around $f = n = \frac{1}{2}$. Subtle cancelations have to be taken into account as not all sources of correction, such as N_{crit}, approximation (51), or the rewriting of the sum as an integral, are equally important; see Appendix A. The visualization of the terms in Eq. (54) that include Kronecker deltas, as well as their scaling, is presented in Fig. 6.

The variance $(\Delta S_A)^2_N = \langle S_A^2\rangle_N - \langle S_A\rangle^2_N$ of the entanglement entropy of pure quantum states in $\mathcal{H}^{(N)}$ can be found using the result in Eq. (24) of Ref. [157]. When expressed as a sum over the number of particles N_A it takes the form

$$(\Delta S_A)^2_N = \frac{1}{d_N+1}\left[\sum_{N_A=0}^{N}\varrho(\varphi^2_{N_A}+\chi_{N_A}) - \left(\sum_{N_A=0}^{N}\varrho_{N_A}\varphi_{N_A}\right)^2\right], \tag{55}$$

where ϱ_{N_A} and φ_{N_A} are given in Eq. (48) and χ_{N_A} is defined as

$$\chi = \begin{cases} (d_A+d_B)\Psi'(d_B+1) - (d_N+1)\Psi'(d_N+1) - \dfrac{(d_A-1)(d_A+2d_B-1)}{4d_B^2}, & d_A \le d_B, \\ (d_A+d_B)\Psi'(d_A+1) - (d_N+1)\Psi'(d_N+1) - \dfrac{(d_B-1)(d_B+2d_A-1)}{4d_A^2}, & d_A > d_B. \end{cases} \tag{56}$$

As earlier, d_N, $d_A(N_A)$, $d_B(N-N_A)$ are understood as functions of the particle number and are given by Eqs. (34) and (37). In the thermodynamic limit $V\to\infty$, at fixed subsystem fraction $f=V_A/V$ and fixed particle density $n=N/V$, the variance is exponentially small and its asymptotic scaling can be obtained via the saddle point methods of Appendix A. In particular, we have

$$\begin{aligned}\sum_{N_A=0}^{N}\varrho_{N_A}\varphi_{N_A}^2-\left(\sum_{N_A=0}^{N}\varrho_{N_A}\varphi_{N_A}\right)^2&=\int_{-\infty}^{\infty}\varrho(n_A)\varphi^2(n_A)dn_A\\&\quad-\left(\int_{-\infty}^{\infty}\varrho(n_A)\varphi(n_A)dn_A\right)^2+o(1)\\&=\left[f(1-f)-\frac{1}{2\pi}\delta_{f,1/2}\right]\\&\quad\times\left(\ln\frac{n}{1-n}\right)^2n(1-n)V+o(V)\end{aligned}\tag{57}$$

and

$$\sum_{N_A=0}^{N}\varrho_{N_A}\chi_{N_A}=\frac{1}{4}\delta_{f,1/2}\delta_{n,1/2}+o(1),\tag{58}$$

where we have used the fact that, for large dimensions, $d_A\gg 1$ and $d_B\gg 1$, χ scales as

$$\chi_{N_A}=\begin{cases}\dfrac{d_A}{2d_B}+O(1/d_B^2), & d_A<d_B,\\ \frac{1}{4}+o(1), & d_A=d_B,\\ \dfrac{d_B}{2d_A}+O(1/d_A^2), & d_A>d_B.\end{cases}\tag{59}$$

Therefore, the term in brackets in Eq. (55) is of order V, while the denominator d_N+1 is exponentially large. Using the Stirling approximation for d_N in Eq. (55), we find that

$$(\Delta S_A)_N^2=\alpha V^{3/2}e^{-\beta V}+o(e^{-\beta V}),\tag{60}$$

with

$$\begin{aligned}\alpha&=\sqrt{2\pi}\left[f(1-f)-\frac{1}{2\pi}\delta_{f,1/2}\right]\left(\ln\frac{n}{1-n}\right)^2\\&\quad\times[n(1-n)]^{3/2}+o(1),\\\beta&=-n\ln n-(1-n)\ln(1-n).\end{aligned}\tag{61}$$

This means that the average entanglement entropy in Eq. (54) is also the typical entanglement entropy of pure quantum states with N fermions, namely, the overwhelming majority of pure quantum states with N fermions have the entanglement entropy in Eq. (54).

3. Weighted average and variance

Having computed the average entanglement entropy of pure states with N particles, next we can compute the average over the entire Hilbert space. A subtlety is that the system is in either of the Hilbert spaces $\mathcal{H}_N$, but we do not know in which one. Therefore, while the distribution of the pure states with a fixed particle number is given quantum mechanically, meaning uniformly distributed over a unit sphere, we additionally have a classical probability for the particle number N.

With this in mind, we can average over $\langle S_A\rangle_N$ within each sector with N particles weighted by its Hilbert space dimension d_N from Eq. (34). More generally, we can introduce a weight parameter w and a probability P_N of finding N particles:

$$P_N=\frac{1}{Z}d_Ne^{-wN}.\tag{62}$$

Here $Z=\sum_{N=0}^{V}d_Ne^{-wN}=(1+e^{-w})^V$ normalizes the distribution. The average filling fraction $\bar{n}$ can be expressed in terms of the weight parameter w as

$$\bar{n}=\sum_{N=0}^{V}P_N\frac{N}{V}=\frac{1}{1+e^w}\tag{63}$$

with half-filling $\bar{n}=\frac{1}{2}$ corresponding to equiweighted sectors, i.e., $w=0$. The variance of the filling fraction,

$$(\Delta n)^2=\sum_{N=0}^{V}P_N\left(\frac{N}{V}-\bar{n}\right)^2=\frac{\bar{n}(1-\bar{n})}{V}\tag{64}$$

can be obtained easily by noting that P_N is a binomial distribution.

We calculate the average entanglement entropy at fixed weight parameter w,

$$\langle S_A\rangle_w=\sum_{N=0}^{V}P_N\langle S_A\rangle_N,\tag{65}$$

up to constant order in V by expanding $\langle S_A\rangle_N$ around $\bar{n}$ and then using the known variance $(\Delta n)^2$. Since $\langle S_A\rangle_N$ is analytic as a function of N (for $f<\frac{1}{2}$) and does not have any discontinuities in its derivatives, it suffices to expand its leading order (linear in V) around $\bar{n}$ as

$$\begin{aligned}s_A(n,f)&=[(n-1)\ln(1-n)-n\ln n]f\\&=[(\bar{n}-1)\ln(1-\bar{n})-\bar{n}\ln\bar{n}]f\\&\quad+f\ln[(1-\bar{n})/\bar{n}]-\frac{f(n-\bar{n})^2}{2(1-\bar{n})\bar{n}}\\&\quad+o(n-\bar{n})^3,\end{aligned}\tag{66}$$

and calculate its expectation value with respect to the binomial distribution. Using $\langle(n-\bar{n})^2\rangle=\sigma^2$, we find the

constant correction $-f/2$, which cancels the identical term in Eq. (54). Terms of order $V^{1/2}$ and V^0 can be directly evaluated at $n=\bar{n}$, where the binomial distribution is centered, because its finite width on those terms will only contribute corrections of subleading order $O(1)$. Hence, the resulting average is equal to

$$\langle S_A\rangle_w = \left[(\bar{n}-1)\ln(1-\bar{n})-\bar{n}\ln(\bar{n})\right]fV - \sqrt{\frac{\bar{n}(1-\bar{n})}{2\pi}}\left|\ln\left(\frac{1-\bar{n}}{\bar{n}}\right)\right|\delta_{f,1/2}\sqrt{V} + \frac{\ln(1-f)}{2} - \frac{2}{\pi}\delta_{f,1/2}\delta_{\bar{n},1/2} + o(1), \tag{67}$$

where $\bar{n}=1/(1+e^w)$ was computed in Eq. (63). A pedagogical derivation of Eq. (67) can be found in Appendix A 2. Interestingly, Eq. (67) can be summarized by the simple relation $\langle S_A\rangle_w = \langle S_A\rangle_{N=\bar{N}} - f/2 + o(1)$ except at $f=\bar{n}=\frac{1}{2}$, where the Kronecker delta from Eq. (54) leads to additional integrals, as explained in Appendix A 2.

For $w=0$ with $\bar{n}=\frac{1}{2}$, Eq. (67) describes the average entanglement entropy of uniformly weighted eigenstates of the number operator (with respect to the Haar measure). This average was computed in Ref. [49] as $\langle S_A\rangle_{w=0} = fV\ln 2 + \ln(1-f)/2 - 2\delta_{f,1/2}/\pi$, which coincides with Eq. (67) for $\bar{n}=\frac{1}{2}$.

Similarly, one can compute the variance of the weighted entanglement entropy

$$(\Delta S_A)^2_w = \sum_{N=0}^{V} P_N\langle S_A^2\rangle_N - \left(\sum_{N=0}^{V} P_N\langle S_A\rangle_N\right)^2 = \bar{n}(1-\bar{n})\left(\ln\frac{\bar{n}}{1-\bar{n}}\right)^2 fV + o(V). \tag{68}$$

Note that, while the variance $(\Delta S)^2_N$ at a fixed number of particles is exponentially small at large V, the weighted variance $(\Delta S)^2_w$ scales linearly in V because of the $O(V^{-1})$ variance $(\Delta n)^2$ in the filling fraction. For $f\neq 0$ and $\bar{n}\neq 0$, the leading-order term only vanishes at $\bar{n}=\frac{1}{2}$. However, we always have $\lim_{V\to\infty}(\Delta S_A)_w/\langle S_A\rangle_w = 0$, i.e., the *relative standard deviation* vanishes in the thermodynamic limit, so that the average entanglement entropy $\langle S_A\rangle_w$ and the *typical* eigenstate entanglement entropy always coincide.

III. PURE FERMIONIC GAUSSIAN STATES

In this section, we define fermionic Gaussian states and calculate the average and variance of the entanglement entropy for this family of states. Following Ref. [159], we do this first for pure fermionic Gaussian states, for which the number of particles is not fixed. Next, we derive new results for fermionic Gaussian states with a fixed number of particles. In both cases we mimic the idea of a uniformly distributed state. This works because in both cases there is a natural action of a compact group and the set is given by a single orbit of this group action. Thus, one can choose the unique Haar measure to generate an ensemble of fermionic Gaussian states.

It may be natural to ask whether the same analysis could also be carried out for bosonic Gaussian states. Unfortunately, the answer is in the negative. The ensemble of bosonic Gaussian states is noncompact with unbounded entanglement entropy since the corresponding invariance group is a noncompact one. So any group invariant average would diverge. Moreover, the only bosonic Gaussian state that has a fixed particle number is the vacuum with zero particles and zero entanglement. To circumvent the problem, one could fix the *average* number of particles. Then, the corresponding manifold would be again compact and one can average over all those Gaussian states (in a similar spirit as in Refs. [168,169]), but the resulting analysis would be rather different from our approach here. It may be possible to use a duality between bosonic and fermionic entanglement entropy of Gaussian states [170] for this, but we will not carry out this analysis here.

A. Definition of fermionic Gaussian states

Instead of starting with pure fermionic Gaussian states, it is easier to begin with mixed Gaussian states because the pure ones can be understood as limits of this definition. We choose a Majorana basis $\{\gamma_j\}_{j=1,\dots,2V}$ in the 2^V-dimensional Hilbert space $\mathcal{H}$ since the corresponding ensemble is easier to describe. This Majorana basis satisfies the anticommutation relation $\{\gamma_j,\gamma_k\}=\delta_{jk}$, meaning that they create a Clifford algebra and can be chosen to be Hermitian, $\gamma_j^\dagger=\gamma_j$. Moreover, it holds that $\mathrm{Tr}(\prod_{l=1}^m \gamma_{j_l})=0$ with $j_l\in\{1,\dots,V\}$ and any positive integer m whenever there is a γ_j that does not appear in this product with an even order. Otherwise, it holds that $\mathrm{Tr}(\prod_{l=1}^m \gamma_{j_l})=\pm 2^{V-m/2}$, which is up to a factor $2^{-m/2}$ the dimension of the representation of the Clifford algebra as well as the dimension of the Hilbert space $\mathcal{H}$.

A Gaussian state is then any density operator of the form

$$\hat{\rho}(\gamma) = \frac{\exp(-\sum_{j,k=1}^{2V} q_{jk}\gamma_j\gamma_k)}{\mathrm{Tr}\exp(-\sum_{j,k=1}^{2V} q_{jk}\gamma_j\gamma_k)} = \frac{\exp(-\gamma^\dagger Q\gamma)}{\mathrm{Tr}\exp(-\gamma^\dagger Q\gamma)} \tag{69}$$

with the Majorana operator-valued column vector $\gamma=(\gamma_1,\dots,\gamma_{2V})^\dagger$. This form gives the Gaussian states their name. The Hermiticity of $\hat{\rho}(\gamma)$ implies that the coefficient matrix $Q=\{q_{jk}\}_{j,k=1,\dots,2V_A}$ needs to be Hermitian, while the anticommutation relations of the Majorana basis allows

us to set the real symmetric part to zero. Indeed, due to

$$\sum_{j,k=1}^{2V} q_{jk}\gamma_j\gamma_k = \sum_{j=1}^{2V} q_{jj} + \sum_{1\le j<k\le 2V} q_{jk}(\gamma_j\gamma_k - \gamma_k\gamma_j) = \sum_{j=1}^{2V} q_{jj} + \sum_{1\le j<k\le 2V} (q_{jk} - q_{kj})\gamma_j\gamma_k, \quad (70)$$

we see that the diagonal part of Q only yields a constant while the symmetric one drops out. Thence, the coefficient matrix is a $2V\times 2V$ imaginary antisymmetric matrix $Q = -Q^* = -Q^T$. Such a matrix can be block diagonalized by an orthogonal matrix $M \in O(2V)$. In particular, it holds that $Q = M^T \mathrm{diag}(\lambda_1\tau_2,\ldots,\lambda_V\tau_2)M$, with the second Pauli matrix τ_2 and $\lambda_j \ge 0$. Introducing $\eta = (\eta_1,\ldots,\eta_{2V})^\dagger = M\gamma$, whose entries create another Majorana basis, the expression simplifies because of $\gamma^\dagger Q\gamma = -2i\sum_{j=1}^{V}\lambda_j\eta_{2j-1}\eta_{2j}$. We can readily compute the exponent

$$\exp[-\gamma^\dagger Q\gamma] = \prod_{j=1}^{V}(\cosh(\lambda_j) + 2i\sinh(\lambda_j)\eta_{2j-1}\eta_{2j}), \quad (71)$$

and the normalization

$$\mathrm{Tr}\exp[-\gamma^\dagger Q\gamma] = 2^V\prod_{j=1}^{V}\cosh(\lambda_j). \quad (72)$$

Summarizing, any Gaussian state has the compact form

$$\hat\rho(\gamma) = 2^{-V}\prod_{j=1}^{V}[1 + 2i\tanh(\lambda_j)\eta_{2j-1}\eta_{2j}], \quad (73)$$

where the λ_j are the singular values and the η_j are the Majorana basis in the corresponding eigenbasis of the matrix Q.

Gaussian states satisfy the Wick-Iserlis theorem, meaning that all moments can be expressed in terms of the first and second moments. Since the first moments vanish for fermions, all the information of a fermionic Gaussian state is encoded in the covariance matrix. When subtracting the identity and multiplying by the imaginary unit, we obtain the symplectic form

$$-i\tilde J = \mathrm{Tr}_{\mathcal H}\left[\hat\rho(\gamma)\left(\gamma\gamma^\dagger - \tfrac{1}{2}\mathbb{1}_{2V}\right)\right] = M^T\mathrm{Tr}_{\mathcal H}\left[\prod_{j=1}^{V}\left(\frac{1}{2} + i\tanh(\lambda_j)\eta_{2j-1}\eta_{2j}\right)(\eta\eta^\dagger - \mathbb{1}_{2V})\right]M. \quad (74)$$

We have emphasized that the trace is only over the Hilbert space $\mathcal H$ and not over the indices of the Majorana basis, which explains why we could take the orthogonal matrix M out the trace. The shift by half of the identity matrix $\frac{1}{2}\mathbb{1}_{2V}$ only subtracts the diagonal terms $\gamma_j^2 = \frac{1}{2}$, which do not contain any information. The symmetries of the Majorana basis tell us that the symplectic form $\tilde J$ is real antisymmetric. In a straightforward computation one can show that

$$\mathrm{Tr}_{\mathcal H}\left[\prod_{j=1}^{V}\left(\frac{1}{2} + i\tanh(\lambda_j)\eta_{2j-1}\eta_{2j}\right)\left(\eta\eta^\dagger - \frac{1}{2}\mathbb{1}_{2V}\right)\right] = \mathrm{diag}[\tanh(\lambda_1)\tau_2,\ldots,\tanh(\lambda_V)\tau_2]. \quad (75)$$

Therefore, the eigenvalues of $\tilde J$ are equal to $\pm ix_j = \pm i\tanh(\lambda_j)$, and the link between Q and $\tilde J$ is given by the bijective relation

$$\tilde J = i\tanh(Q). \quad (76)$$

As we can go back and forth between these two matrices, $\hat\rho(\gamma)$ is fully determined by $\tilde J$ so that it is suitable to adopt the notation $\hat\rho(\tilde J)$. For instance, we can express the von Neumann entropy in terms of $\tilde J$ because of

$$-\mathrm{Tr}[\hat\rho(\tilde J)\ln\hat\rho(\tilde J)] = \mathrm{Tr}\left[\prod_{j=1}^{V}\left(\frac{1}{2} + ix_j\eta_{2j-1}\eta_{2j}\right)\sum_{k=1}^{V}\ln\left(\frac{1}{2} + ix_j\eta_{2k-1}\eta_{2k}\right)\right] = \sum_{k=1}^{V}s(x_k) = \mathrm{Tr}\,s(i\tilde J) \quad (77)$$

with [171–173]

$$s(x) = -\left(\frac{1+x}{2}\right)\ln\left(\frac{1+x}{2}\right) - \left(\frac{1-x}{2}\right)\ln\left(\frac{1-x}{2}\right). \quad (78)$$

With the help of the von Neumann entropy, it is straightforward to identify the pure fermionic Gaussian states. Those are given when all eigenvalues are equal to $x_j = \pm 1$. Indeed, a density operator of the form $\hat\rho(\tilde J) = 2^{-V}\prod_{j=1}^{V}(1 \pm 2i\eta_{2j-1}\eta_{2j})$ satisfies the necessary and sufficient condition for pure states (in combination with positive semidefiniteness and the normalized trace), i.e.,

$$\hat\rho^2(\tilde J) = 2^{-2V}\prod_{j=1}^{V}(1 \pm 2i\eta_{2j-1}\eta_{2j})^2 = 2^{-V}\prod_{j=1}^{V}(1 \pm 2i\eta_{2j-1}\eta_{2j}) = \hat\rho(\tilde J). \quad (79)$$

The corresponding normalized state vector of $\hat\rho(\tilde J) = |\tilde J\rangle\langle\tilde J|$ is denoted by $|\tilde J\rangle$ and it is only determined up

to a complex phase. The set of all real antisymmetric matrices $\tilde{J}$ with eigenvalues $\pm i$ is described by $\{\tilde{J} = M^T J_0 M |\, M \in O(2V)$ and $J_0 = i\tau_2 \otimes \mathbb{1}_V\}$. This gives a natural parametrization of pure fermionic Gaussian states, which will be our starting point in Sec. III B.

B. Arbitrary number of particles

We first focus on the family of pure fermionic Gaussian states in which one does not fix the particle number, i.e., we include Gaussian states that consist of a superposition of states with different total particle number.

1. Statistical ensemble of states

As we have seen in the previews subsection, all pure fermionic Gaussian states can be described by their symplectic form $\tilde{J} = i\,\langle\tilde{J}|\gamma\gamma^\dagger - \mathbb{1}_{2V}|\tilde{J}\rangle = M^T J_0 M$, with an orthogonal matrix $M \in O(2V)$ and the symplectic unit $J_0 = i\tau_2 \otimes \mathbb{1}_V$, which is essentially a canonical embedding of the second Pauli matrix τ_2 in the $2V$-dimensional space and defines a complex structure; see Refs. [159,172,173]. One can render the relation between pure states and real antisymmetric matrices $\tilde{J}$ with eigenvalues $\pm i$, i.e., $\tilde{J}^2 = -\mathbb{1}_{2V}$, to a one-to-one correspondence when dividing all orthogonal matrices out of $O(2V)$ that commute with J_0. These matrices build a subgroup that is the real representation of the unitary group $U(V)$ (the direct sum of the fundamental and antifundamental representations). Thence, the set of all pure fermionic Gaussian states can be identified with the coset $O(2V)/U(V)$, which is $V(V-1)$ dimensional.

There is a natural $O(2V)$ group action on the pure states by $\tilde{J} \to M^T\tilde{J}M$ that corresponds to the change of an orthonormal Majorana basis. Therefore, adopting Page's idea of a uniform distribution that is given by a group action, we assume that the ensemble of random pure fermionic Gaussian states is created by the normalized Haar distribution on $O(2V)/U(V)$. Practically, this can be realized by drawing a Haar-distributed orthogonal matrix $M \in O(2V)$, and considering the pure state corresponding to the real antisymmetric matrix $\tilde{J} = M^T J_0 M$; see Ref. [159].

When restricting a pure fermionic Gaussian state $\hat{\rho}(\tilde{J})$ to a subsystem A with a $(d_A = 2^{V_A})$-dimensional Hilbert space $\mathcal{H}_A$, one obtains a mixed Gaussian state. The corresponding symplectic form can be obtained by a projection of the matrix $\tilde{J}$. Without loss of generality, we assume that $\tilde{\gamma} = (\gamma_1, \ldots, \gamma_{2V_A})^\dagger$ is an orthonormal Majorana basis only acting nontrivially on $\mathcal{H}_A$ but has a trivial action on the other Hilbert space $\mathcal{H}_B$. Defining Π_A as the projection of a $2V$ vector onto the first $2V_A$ components, it holds that $\tilde{\gamma} = \Pi_A\gamma$ and the new symplectic form corresponding to the state $\hat{\rho}_A(\tilde{J}) = \mathrm{Tr}_{\mathcal{H}_B}\,\hat{\rho}(\tilde{J})$ is

$$\begin{aligned}\tilde{J}_A &= \mathrm{Tr}_{\mathcal{H}_A}\left[\hat{\rho}_A(\tilde{J})\left(\tilde{\gamma}\tilde{\gamma}^\dagger - \tfrac{1}{2}\mathbb{1}_{2V_A}\right)\right] \\ &= \Pi_A\,\mathrm{Tr}_{\mathcal{H}}\left[\hat{\rho}(\tilde{J})\left(\gamma\gamma^\dagger - \tfrac{1}{2}\mathbb{1}_{2V}\right)\right]\Pi_A^T \\ &= \Pi_A\tilde{J}\Pi_A^T. \end{aligned} \tag{80}$$

Hence, the new covariance matrix is only an orthogonal projection of the old one onto its upper left $2V_A \times 2V_A$ block. Surely, it can be any diagonal block or even a more complicated embedding of this $2V_A \times 2V_A$ matrix $\tilde{J}_A$ in $\tilde{J}$. However, the group invariant generation of the pure fermionic Gaussian states tells us that all these embeddings are equivalent. Physically, this means that all these subsystems of $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ are essentially the same once the dimension d_A is fixed. We have already seen this picture in Page's setting.

In Ref. [159], it was shown that the random matrix $\tilde{J}_A = \Pi_A M^T J_0 M \Pi_A^T$ with a Haar distributed $M \in O(2V)$ has a joint probability density of its eigenvalues $i\mathrm{diag}(x_1\tau_2, \ldots, x_{V_A}\tau_2)$ of the form

$$P(x) = \mathcal{N}\prod_{j<k}(x_j - x_k)^2 \prod_{l=1}^{V_A}(1 - x_l^2)^{V-2V_A} \tag{81}$$

with $\mathcal{N}$ the normalization. Here, we already see that it is crucial to assume that $V_A \le V/2$; otherwise, we need to consider the density operator $\hat{\rho}_B(\tilde{J}) = \mathrm{Tr}_{\mathcal{H}_A}\,\hat{\rho}(\tilde{J})$. Indeed, it is again the same symmetry between subsystems A and B that is still true here, and the breakdown of analyticity is due to some eigenvalues, either of $\hat{\rho}_A(\tilde{J})$ or $\hat{\rho}_B(\tilde{J})$ being exactly zero. Those eigenvalues are related to the eigenvalues of $\tilde{J}_A$ or, equivalently, $\tilde{J}_B$ that are exactly $\pm i$.

The main idea that enters the computation of the joint probability distribution (81) is Proposition A.2 of Ref. [174], which shows what the eigenvalues of a corank-2 projection of a real antisymmetric matrix are. Then, one needs only repetitively apply this proposition, leading to the distribution above.

One important ingredient in the computation is that all k-point correlation functions can be expressed in terms of a single kernel function $K(x_l, x_j)$ as [175]

$$\begin{aligned}R_k(x_1, \ldots, x_k) &= \frac{V_A!}{(V_A - k)!}\int_{-1}^{1} dx_{k+1}\cdots\int_{-1}^{1} dx_{V_A} P(x) \\ &= \det[K(x_l, x_j)]_{l,j=1,\ldots,k}. \end{aligned} \tag{82}$$

In the mathematical branch of random matrix theory, this structure is known as a determinantal point process [176]. The average and variance of the entanglement entropy can then be traced back to an integral of the one-point function $R_1(x)$ and the two-point correlation function $R_2(x_1, x_2)$,

respectively. In general, higher cumulants of the entanglement entropy are averages of specific k-point correlation functions.

Distribution (81) is shared with the unitary Jacobi ensemble [136,163], which is a truncation of a Haar-distributed unitary matrix. Despite the fact that the eigenvalue statistics is the same with a unitary Jacobi ensemble, the eigenvector statistics is different. This can be readily seen in that the eigenvectors of the unitary Jacobi ensemble must be $U(V_A)$ invariant, while in the present case they are only $O(2V_A)$ invariant.

2. Average and variance

Our goal is again to compute the mean and variance of the entanglement entropy of subsystem A,

$$\langle S_A\rangle_G = \int S_A(|MJ_0M^{-1}\rangle)d\mu(M), \tag{83}$$

$$(\Delta S_A)^2_G = \int (S_A(|MJ_0M^{-1}\rangle) - \langle S_A\rangle_G)^2 d\mu(M), \tag{84}$$

where $S_A(|\tilde{J}\rangle) = \mathrm{Tr}\, s(i\tilde{J}_A)$ [see Eq. (78)] and $d\mu(M)$ now represents the Haar measure over the orthogonal group. Computing the average entanglement entropy over all Gaussian states was the main technical achievement of Ref. [159]. It was facilitated by recent results in random matrix theory [174], from which one can deduce what the joint probability distribution of the singular values $\tilde{J}_A$ are. For this, we make repetitive use of Proposition A.2 of Ref. [174] by always projecting away two rows of matrix J; first to $[J]_{N-1}$, then $[J]_{N-2}$, until we arrive at $[J]_{N_A}$. This yields for $x = (x_1,\ldots,x_{N_A})$ the distribution

$$P(x) = \frac{(\det X)^2}{N_A!}\left(\prod_{j=0}^{N_A-1} c_j^{-1}(1-x_{j+1}^2)^{\Delta}\right), \tag{85}$$

where we have the $N_A \times N_A$ matrix X and c_j given by

$$X_{ij} = p_{j-1}(x_i) = \mathcal{P}^{(\Delta,\Delta)}_{2j-2}(x_i), \tag{86}$$

$$c_j = \frac{2^{2\Delta}[(2j+\Delta)!]^2}{(2j)!\,(2j+2\Delta)!\,(4j+2\Delta+1)}, \tag{87}$$

with $\Delta = N_B - N_A \geq 0$. The k-point correlation functions from Eq. (82) are fully encoded by $K(x_a,x_b)$, which is given by the $k\times k$ matrix (with $a,b = 1,\ldots,k$) given by [136]

$$K(x,y) = \sum_{j=0}^{N_A-1} \psi_j(x)\psi_j(y), \quad \psi_j(x) = \frac{(1-x^2)^{\Delta/2}}{\sqrt{c_j}}p_j(x), \tag{88}$$

with $\int_0^1 \psi_j(x)\psi_k(x)dx = \delta_{jk}$. The average entanglement entropy can be written in terms of the one-point function,

$$\begin{aligned}\langle S_A\rangle_G &= \int_0^1 R_1(x)s(x)dx \\ &= \left(V-\tfrac{1}{2}\right)\Psi(2V) + \left(\tfrac{1}{2}+V_A-V\right)\Psi(2V-2V_A) \\ &\quad + \left(\tfrac{1}{4}-V_A\right)\Psi(V) - \tfrac{1}{4}\Psi(V-V_A) - V_A,\end{aligned} \tag{89}$$

where the details can be found in Appendix C 1. Anew, the symmetry $V_A \leftrightarrow V - V_A$ is not reflected in this result and needs to be introduced by hand. The origin of this breaking in the analytical result is, as in Page's setting, one of the two density operators $\hat{\rho}_A$ and $\hat{\rho}_B$ has an exact number of zero modes. The result corresponds to this system without these generic zero modes. This selection is a nonanalytical step. Indeed, there is a nonanalytic kink at $V_A = V/2$ (first derivative jumps there). However, it is difficult to see when plotting the result even for moderately small V (say of the order 10). The reason is that this kink vanishes like $1/V^2$, so that it is only of the order of 1% when $V = 10$.

For $f = V_A/V \leq \frac{1}{2}$, the thermodynamic limit reads

$$\begin{aligned}\langle S_A\rangle_G &= V[(\ln 2 - 1)f + (f-1)\ln(1-f)] \\ &\quad + \tfrac{1}{2}f + \tfrac{1}{4}\ln(1-f) + O(1/V),\end{aligned} \tag{90}$$

whose leading-order term was found earlier in Ref. [61] to give the average eigenstate entanglement entropy of number-preserving random quadratic Hamiltonians. This match is not a coincidence, as discussed in Sec. III C 3.

Interestingly, result (90) glued to its reflection $f \to 1-f$ at $f = \frac{1}{2}$ is 2 times differentiable at $f = \frac{1}{2}$. Thus, the nonanalyticity is hardly visible. Starting with the third derivative one can actually see the breaking of analyticity. We note that, in contrast to the case of general pure states considered by Page, in Eq. (90) there is no Kronecker delta contribution at $f = \frac{1}{2}$.

The variance $(\Delta S_A)^2$ can be computed from the matrix representation of the entanglement entropy function $s(x)$ with respect to the function $\psi_i(x)$:

$$s_{ij} = \int_{-1}^{1} s(x)\psi_i(x)\psi_j(x)\,dx. \tag{91}$$

In full analogy to the calculation in Ref. [159], we find that

$$\begin{aligned}(\Delta S_A)^2_G &= \int_{-1}^{1} s^2(x)K(x,x)dx \\ &\quad - \int_{-1}^{1} s(x_1)s(x_2)K^2(x_1,x_2)d^2x\end{aligned}$$

$$
\begin{aligned}
&= \int_{-1}^{1} s(x_1)s(x_2)K(x_1,x_2)[\delta(x_1-x_2) \\
&\quad - K(x_1,x_2)]d^2x \\
&= \int_{-1}^{1} s(x_1)s(x_2)\Big[\sum_{i=0}^{V_A-1}\psi_i(x_1)\psi_i(x_2)\Big] \\
&\quad \times \Big[\sum_{j=V_A}^{\infty}\psi_j(x_1)\psi_j(x_2)\Big]d^2x \\
&= \sum_{i=0}^{V_A-1}\sum_{j=V_A}^{\infty} s_{ij}^2 . \qquad (92)
\end{aligned}
$$

The variance of the entanglement entropy for a pure fermionic Gaussian state was computed to leading order in Ref. [159] and is given in

$$
(\Delta S_A)_G^2 = \frac{f + f^2 + \ln(1-f)}{2} + o(1). \qquad (93)
$$

We present a systematic derivation in Appendix C 2 based on certain integrals of Jacobi polynomials. Using the techniques of Refs. [149,177], we expect that the variance can be calculated in a closed compact form for fixed V as an expression in terms of digamma functions, analogous to Eq. (89). In fact, Huang and Wei [178] conjectured such an analytical formula.

C. Fixed number of particles

We consider fermionic Gaussian states with a fixed particle number N, i.e., the intersection of the family of fermionic Gaussian states and the Hilbert space $\mathcal{H}^{(N)}$.

For this, it is useful to change from the basis of Majorana operators γ_i to the basis of creation and annihilation operators. This basis, $\{\hat{f}_j = (1/\sqrt{2})(\gamma_{2j-1} + i\gamma_{2j}),\ \hat{f}_j^\dagger = (1/\sqrt{2})(\gamma_{2j-1} - i\gamma_{2j})\}_{j=1,\ldots,V}$, will be helpful when studying Gaussian states with fixed particle numbers. Those are given by three $\pi/4$ rotations of the form

$$
(\hat{f}_1,\ldots,\hat{f}_V,\hat{f}_1^\dagger,\ldots,\hat{f}_V^\dagger) = (\gamma_1,\ldots,\gamma_{2V})T^\dagger = \gamma^\dagger T^\dagger \qquad (94)
$$

with

$$
T = e^{i\pi/4}\exp\Big[-i\frac{\pi}{4}\tau_3\otimes\mathbb{1}_V\Big]\exp\Big[-i\frac{\pi}{4}\tau_1\otimes\mathbb{1}_V\Big], \qquad (95)
$$

and τ_1 and τ_3 being the first and third Pauli matrices. Hence, the symplectic form becomes a complex structure in this basis that is given by

$$
\begin{aligned}
J &= iT\,\mathrm{Tr}_{\mathcal{H}}\big[\hat{\rho}(\gamma)\big(\gamma\gamma^\dagger - \tfrac{1}{2}\mathbb{1}_{2V}\big)\big]T^\dagger \\
&= i\begin{pmatrix} \langle\tilde{J}|\hat{f}_i\hat{f}_j^\dagger - \hat{f}_j^\dagger\hat{f}_i|\tilde{J}\rangle & \langle\tilde{J}|\hat{f}_i^\dagger\hat{f}_j^\dagger - \hat{f}_j^\dagger\hat{f}_i^\dagger|\tilde{J}\rangle \\ \langle\tilde{J}|\hat{f}_i\hat{f}_j - \hat{f}_j\hat{f}_i|\tilde{J}\rangle & \langle\tilde{J}|\hat{f}_i^\dagger\hat{f}_j - \hat{f}_j\hat{f}_i^\dagger|\tilde{J}\rangle \end{pmatrix}_{i,j=1,\ldots,V},
\end{aligned} \qquad (96)
$$

where we have used the anticommutation relation $\{\hat{f}_k,\hat{f}_l^\dagger\} = \delta_{kl}$ and $\{\hat{f}_k,\hat{f}_l\} = \{\hat{f}_k^\dagger,\hat{f}_l^\dagger\} = 0$. The transformation of $\tilde{J} \to J$ is a unitary transformation.

For fermionic Gaussian states with a fixed number of particles, the off-diagonal blocks in Eq. (96) vanish because those contain expectation values of operators that change the particle number. Thus, we are left with the eigenvalues of the two V-by-V matrices $F_{ij} = -i\langle J|\hat{f}_i\hat{f}_j^\dagger - \hat{f}_j^\dagger\hat{f}_i|J\rangle$ and $G_{ij} = -i\langle J|\hat{f}_i^\dagger\hat{f}_j - \hat{f}_j\hat{f}_i^\dagger|J\rangle$. These two matrices are intimately related via $F = -G^{\mathsf{T}}$ due to the anticommutation relations. Actually, it is also a direct consequence of the anti-Hermiticity and the $\tau_1\otimes\mathbb{1}_V$ antisymmetry of $J = -J^\dagger = -(\tau_1\otimes\mathbb{1}_V)J^T(\tau_1\otimes\mathbb{1}_V)$. Therefore, when ix_j is an eigenvalue of the anti-Hermitian $V\times V$ matrix F, then $-ix_j$ has to be an eigenvalue of G. There are no additional symmetries of F and G, meaning that they can be arbitrary Hermitian matrices. Only their singular values are bounded to be inside the interval $[0,1]$ because this is already the case for the complex structure J that is inherited from the positive semidefiniteness of state $\hat{\rho}$.

Using the canonical commutation relations, it holds that

$$
F_{ij} = 2i\langle J|\hat{f}_j^\dagger\hat{f}_i|J\rangle - i\delta_{ij} = -G_{ji}. \qquad (97)
$$

This equation relates F and G to the one-body reduced density matrix that is defined as $C_{ij} = \langle J|\hat{f}_j^\dagger\hat{f}_i|J\rangle$. Indeed, the matrix C is Hermitian,

$$
\{C^\dagger\}_{ij} = C_{ji}^* = (\langle J|\hat{f}_i^\dagger\hat{f}_j|J\rangle)^* = \langle J|\hat{f}_j^\dagger\hat{f}_i|J\rangle = C_{ji}, \qquad (98)
$$

and positive semidefinite,

$$
v^\dagger C v = \sum_{i,j=1}^{V} C_{ij}v_i^* v_j = \Big\|\sum_{j=1}^{V} v_j\hat{f}_j\,|J\rangle\Big\|^2 \geq 0. \qquad (99)
$$

Moreover, its trace is fixed, $\mathrm{Tr}\,C = \langle J|\sum_{j=1}^{V}\hat{f}_j^\dagger\hat{f}_j|J\rangle = \langle J|\hat{N}|J\rangle = N$, in an eigenspace of the total number operator $\hat{N}$. Hence, after a proper normalization one can interpret C as a density matrix.

1. Statistical ensemble of states

We have seen that, for a pure fermionic Gaussian state, the eigenvalues of J must be $\pm i$ or invariantly written

$J^2 = -\mathbb{1}_{2V}$. Therefore, the eigenvalues of sub-blocks F and G are also $\pm i$ when we assume a fixed particle number; see Ref. [173]. Any basis $\{\hat{f}_j'\}_{j=1,\ldots,V}$, which admits the same canonical anticommutation relations and spanning the same space of creation operators as the original basis $\{\hat{f}_j\}_{j=1,\ldots,V}$, can be chosen in the construction of F and G. The set of these bases of creation operators is given by the action of the unitary group $U(V)$, i.e., $(\hat{f}'^{\dagger}_1, \ldots, \hat{f}'^{\dagger}_V) = (\hat{f}_1^{\dagger}, \ldots, \hat{f}_V^{\dagger})U$ with $U \in U(V)$. As each basis is bijectively related to a pure state $\hat{\rho} = |J\rangle\langle J|$, the set of all pure fermionic Gaussian states with a fixed particle number N can be generated by $F = U^{\dagger}F_0U$ and $G = U^T G_0 U^*$, where F_0 and G_0 are diagonal matrices with $\pm i$ on their diagonal. One can bring the number of eigenvalues with $+i$ and $-i$ in connection with the fixed number of particles N when tracing the matrix F, yielding

$$\mathrm{Tr}\, F = 2i\left\langle J \middle| \sum_{j=1}^{V} \hat{f}_j^{\dagger}\hat{f}_j \middle| J \right\rangle - iV = -i(V-2N). \quad (100)$$

Thus, F_0 can be chosen as $F_0 = i\mathrm{diag}(\mathbb{1}_N, -\mathbb{1}_{V-N})$, and, equivalently, $G_0 = i\mathrm{diag}(-\mathbb{1}_N, \mathbb{1}_{V-N})$. Similarly, we can write the one-body reduced density matrix $C = U^{\dagger}C_0U$ with $C_N = \mathrm{diag}(\mathbb{1}_N, 0, \ldots, 0)$ that comprises $V-N$ zeros, and where the subscript highlights the number of particles.

The group action of $U(V)$ on pure fermionic Gaussian states with a fixed particle number again suggests the notion of a uniform distribution. Hence, we generate the state ensemble by parameterizing the associated complex structure as $J = i\mathrm{diag}(U^{\dagger}[2C_N - \mathbb{1}_V]U, U^T[\mathbb{1}_V - 2C_N]U^*)$ with a Haar distributed $U \in U(V)$.

When we consider a subsystem consisting of the first V_A sites, we need to restrict J to the $2V_A \times 2V_A$ matrix J_A, in which both F and G are restricted to the left upper $V_A \times V_A$ blocks. This choice, as before, results in no loss of generality since the Haar-distributed matrix U covers any other kind of orthogonal projection. That the restriction to a sub-block is indeed directly related to the restriction of a subsystem follows along the same lines as in the case without a fixed particle number. One needs to compute the covariance matrix that is given by the annihilation and creation operators that only act nontrivially in the Hilbert space $\mathcal{H}_A$, which is again equivalent with a projection $J_A = \mathrm{diag}(F_A, G_A)$ and, thus, $F_A = \hat{\Pi}_A F \hat{\Pi}_A^T$ and $G_A = \hat{\Pi}_A G \hat{\Pi}_A^T$, where $\hat{\Pi}_A$ projects onto the first V_A rows.

Instead of using the spectrum $\pm ix$ of J_A, it is sometimes convenient to use the eigenvalues y of the $V_A \times V_A$ restricted one-body reduced density matrix $C_A = \hat{\Pi}_A C \hat{\Pi}_A^T$. It still holds that $J_A = i\mathrm{diag}(2C_A - \mathbb{1}_{V_A}, \mathbb{1}_{V_A} - 2C_A)$. This implies that, for the entanglement entropy, we have $\mathrm{Tr}\, s(iJ) = 2\,\mathrm{Tr}\, s(2C_A - \mathbb{1}_{V_A})$ based on Eq. (97). In terms of eigenvalues this reads $s(x) = s(2y-1)$. The entanglement entropy per volume $s(2y-1)$ vanishes for $y = 0$ and $y = 1$, due to $s(\pm 1) = 0$. Therefore, the entanglement entropy S_A is invariant under changing the number of eigenvalues 0 or 1 in C_A.

The generation of C_A is then given by a Haar-random unitary $V \times V$ matrix U and the matrix product

$$C_A = [U]_{V_A \times N}[U]^{\dagger}_{N \times V_A}, \quad (101)$$

where $[U]_{V_A \times N}$ is the $V_A \times N$ upper left sub-block of the matrix U. The matrix $[U]_{V_A \times N}$ is also known as the truncated unitary ensemble or simply the unitary Jacobi ensemble in random matrix theory [136,163]. It appears in several contexts such as quantum transport [179] and quantum scattering [180], as it can be seen as a sub-block of an S matrix.

Let us summarize the symmetries of the above setting.

(i) The particle-hole symmetry, which is given by $N \leftrightarrow V - N$, is reflected when replacing $C = U^{\dagger}C_N U \to \mathbb{1}_V - C = U^{\dagger}(\mathbb{1}_V - C_N)U$. Exploiting the symmetry $s(x) = s(-x)$, it holds that

$$\mathrm{Tr}\, s(\hat{\Pi}_A[2C - \mathbb{1}_V]\hat{\Pi}_A) = \mathrm{Tr}\, s(\hat{\Pi}_A[2(\mathbb{1}_V - C) - \mathbb{1}_V]\hat{\Pi}_A), \quad (102)$$

which underlines this symmetry.

(ii) There is again a symmetry between subsystems A and B. Anew, it is not immediate as the selection is always given by the smallest of the two complementary diagonal blocks [one of size $V_A \times V_A$ and of size $(V - V_A) \times (V - V_A)$] of C. These are anew given by having a density matrix without zero modes. Thus, we actually expect to put this symmetry in by hand as before.

(iii) Surprisingly, this manual implementation of the exchange of subsystems is not really needed for fermionic Gaussian states as there is another, more subtle symmetry that relates to the number of exact eigenvalues at $+i$ of the symplectic form $\tilde{J}_A$. This is manifested in an additional particle-subsystem symmetry $V_A \leftrightarrow N$. Its mathematical origin is that the spectra of $[U]_{V_A \times N}[U^{\dagger}]_{N \times V_A}$ and $[U^{\dagger}]_{N \times V_A}[U]_{V_A \times N}$ only differ by the number of zero eigenvalues, which correspond to exact eigenvalues $\pm i$ of $\tilde{J}_A$. Physically, this means that there are fermionic modes in the eigenbasis of $\tilde{J}_A$ that only act on $\mathcal{H}_A$ and do not act on the sub-Hilbert space $\mathcal{H}_B$. The particle-subsystem symmetry also needs to be introduced by hand as the calculation requires that C_A has no generic zero eigenvalues. Therefore, we can expect a breaking of analyticity at the symmetry axes $N = V_A$ and $N = V - V_A$ because of the particle-hole symmetry $N \leftrightarrow V - N$. One consequence of the particle-subsystem symmetry is that the symmetry axis defined by $V_A = V/2$ must have the same

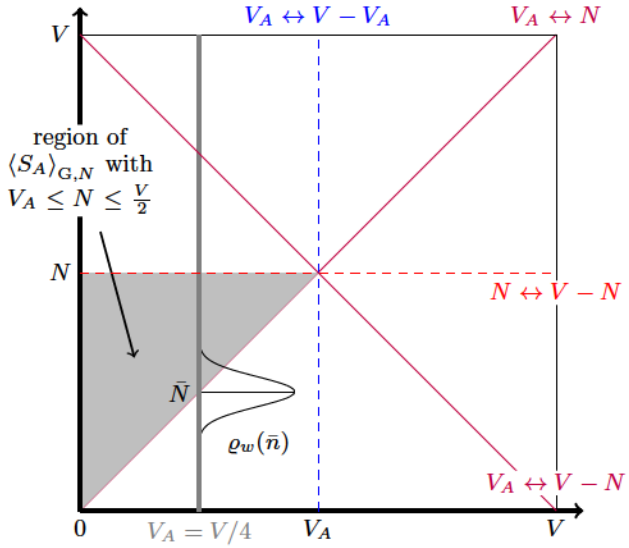


FIG. 7. Illustration of the symmetries of the average entanglement entropy $\langle S_A\rangle_{G,N}$ as a function of V_A and N. When we compute the average $\langle S_A\rangle_{G,w}$ as a function of (w, V_A, V), we are effectively integrating against the density function $\varrho_w(N)$ at fixed V_A, which is approximately a Gaussian (plus corrections) centered at the expectation value $\bar{N} = Ve^{-w}/(1+e^{-w})$. Transitions are characterized by an enhanced correction of order $1/\sqrt{V}$ to $\langle S_A\rangle_{G,w}$, and occur whenever $\bar{N} = V_A$, as this is the point where the increasingly narrow Gaussian is integrated against two different analytic functions on either side of its peak (due to the particle-subsystem symmetry). The function $\langle S_A\rangle_{G,N}$ is analytical across the subsystem and particle-hole symmetries (dashed lines), but *not* across particle-subsystem symmetries (solid lines).

analyticity properties as the symmetry axis $N = V/2$. This is the reason why the implementation of the symmetry between the two subsystems is not needed.

The three symmetries create the overall symmetry group $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \simeq \mathbb{Z}_2 \otimes \mathbb{Z}_4$, which can be visualized by the respective mirror axes. The latter product $\mathbb{Z}_2 \otimes \mathbb{Z}_4$ reflects the fact that there is a finite rotation group $\mathbb{Z}_4$ and a point reflection group $\mathbb{Z}_2$, which commute. When we compute $\langle S_A\rangle_{G,N}$ in Eq. (113) for $V_A \leq N \leq N/2$, we only compute it on one eighth of the available parameter space. Using the above symmetries, one can easily deduce $\langle S_A\rangle_{G,N}$ for any other values of V_A and N. We illustrate the symmetries and the respective transitions in Fig. 7.

The eigenvalue distribution of iF is given by the unitary Jacobi ensemble, as discussed in Refs. [136,181]. This distribution has the form

$$P(x) = \mathcal{N} \prod_{j<k} (x_j - x_k)^2 \prod_{i=1}^{V_A} (1+x_i)^{N-V_A} (1-x_i)^{V-V_A-N}. \tag{103}$$

We can rewrite this probability distribution as

$$P(x) = \frac{(\det X)^2}{V_A!} \left[\prod_{j=0}^{V_A-1} c_j^{-1} (1-x_{j+1})^\alpha (1+x_{j+1})^\beta \right], \tag{104}$$

where we have the $V_A \times V_A$ matrix X and c_j given by

$$X_{ij} = p_{j-1}(x_i) = \mathcal{P}_{j-1}^{(\alpha,\beta)}(x_i), \tag{105}$$

$$c_j = \frac{2^{\alpha+\beta+1}}{2j+\alpha+\beta+1} \frac{(j+\alpha)!\,(j+\beta)!}{(j+\alpha+\beta)!\,j!}, \tag{106}$$

with

$$\alpha = V - N \geq 0, \tag{107}$$

$$\beta = V - N - V_A \geq 0, \tag{108}$$

and $\mathcal{P}_n^{(\alpha,\beta)}(z)$ the Jacobi polynomials. We can define the function

$$\psi_j(x) = \frac{1}{\sqrt{c_j}} \mathcal{P}_j^{(\alpha,\beta)}(x), \tag{109}$$

which allows us to express the level density and the two-point kernel as

$$R_1(x) = \sum_{i=0}^{V_A-1} \psi_i^2(x), \tag{110}$$

$$K(x,y) = \sum_{i=0}^{V_A-1} \psi_i(x)\psi_i(y). \tag{111}$$

Equation (82) underlines that the kernel is a centerpiece in the general spectral statistics of determinantal point processes, as it is here.

The above analytical preparations are our starting point for the computations that are performed in Appendix D and whose results are summarized in Sec. III C 2.

2. Average and variance

The average entanglement entropy over all pure fermionic Gaussian states with total particle number N and a subsystem volume V_A out of a total volume V (with $V_A \leq N \leq V/2$) is

$$\langle S_A\rangle_{G,N} = V_A \int_{-1}^{1} R_1(x) s(x)\, dx \tag{112}$$

with $s(x)$ from Eq. (78) and $R_1(x)$ is the one point function. The evaluation of this integral is explained in detail

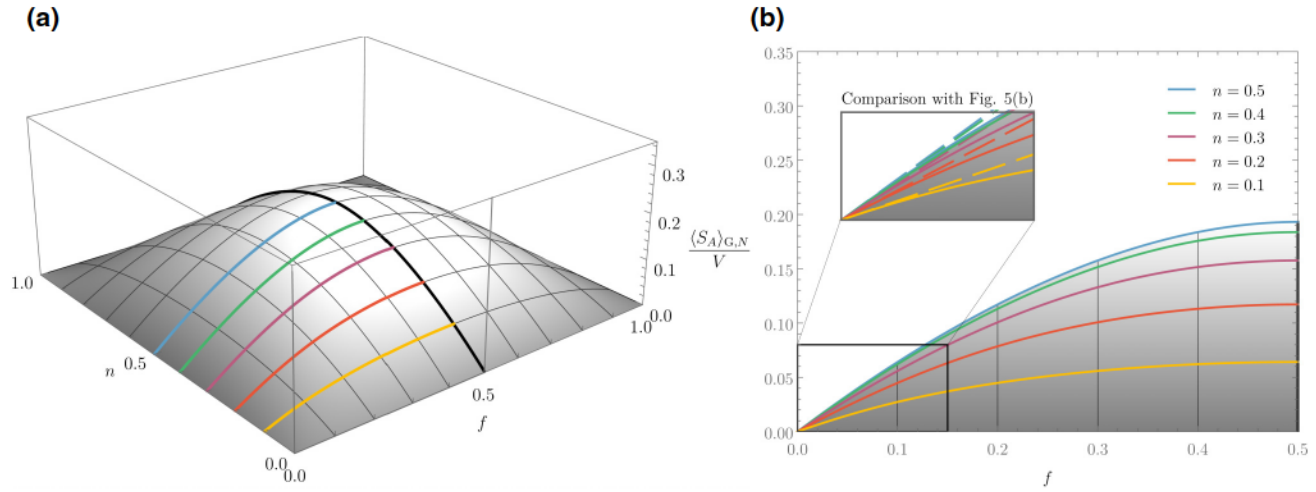


FIG. 8. The leading order of the entanglement entropy $s_A^G(f,n) = \lim_{V\to\infty} \langle S_A\rangle_{G,N}/V$ from Eq. (114) [see also Eq. (D28)]. One can see the mirror symmetries $V_A \to V - V_A$, $N \to V - N$, and $V_A \to N$. For $n = \frac{1}{2}$, $s_A^G(f,n)$ coincides with the formula derived in Refs. [61,159]. (a) Three-dimensional plot as a function of the subsystem fraction $f = V_A/V$ and the filling ratio $n = N/V$. One can see the mirror symmetries $V_A \to V - V_A$ and $N \to V - N$. (b) Results at fixed n plotted as functions of f. The inset shows that $s_A^G(f,n) \sim s_A(f,n)$ with $s_A(f,n) = \lim_{V\to\infty} \langle S_A\rangle_N/V$ [meaning Eq. (66) in Page's setting] as $f \to 0$. As before, the colored curves are the same in the left and right plots.

in Appendix D 1. We obtain

$$\langle S_A\rangle_{G,N} = 1 - \frac{V_A}{V}(1+V) - \frac{NV_A}{V}\Psi(N) + V\Psi(V) + \frac{V_A(N-V)}{V}\Psi(V-N) + (V_A - V)\Psi \times (V - V_A + 1) \quad (113)$$

for $V_A \le N \le V/2$, where $\Psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function. All other values of N and V_A can be computed by using the fact that the entanglement entropy is symmetric under $N \to V - N$, $V_A \to V - V_A$, and $N \leftrightarrow V_A$. Let us emphasize that Eq. (113) is already symmetric under $N \to V - N$. This will play an important role when identifying $\sqrt{V}$ contributions for averages at fixed weight parameter w.

If we define $n = N/V$ and $f = V_A/V$, we can expand this formula in V to find the thermodynamic limit

$$\langle S_A\rangle_{G,N} = [(f-1)\ln(1-f) + f[(n-1)\ln(1-n) - n\ln n - 1]]V + \frac{f[1-f+n(1-n)]}{12(1-f)(1-n)n}\frac{1}{V} + O(V^{-3}), \quad (114)$$

where we assume that $f \le n \le \frac{1}{2}$. One can use the symmetries discussed in Fig. 7 to find $\langle S_A\rangle_{G,N}$ for the other parameters. We note that the leading orders for $f \le \frac{1}{2}$ and $n = \frac{1}{2}$ read

$$\langle S_A\rangle_{G,N=V/2} = [(\ln 2 - 1)f + (f-1)\ln(1-f)]V + \frac{(5/4 - f)f}{3(1-f)}\frac{1}{V} + O(V^{-3}). \quad (115)$$

Remarkably, the leading-order in V is the same as that found in Ref. [61] for the average over eigenstates of number-conserving random Hamiltonians, and in Ref. [159] for the ensemble of all fermionic Gaussian states. Why this is no coincidence will become apparent in Sec. III C 3. In Fig. 8, we visualize our analytical results for the leading order of $\langle S_A\rangle_{G,N}$.

We compute the variance $(\Delta S_A)_{G,N}$ using the same strategy as Eq. (93) based on s_{ij} from Eq. (C12), where $\psi_i(x)$ now comes from Eq. (109). In Appendix D 2, we study the asymptotics of s_{ij} around the leading contribution s_{V_A-1,V_A} in the limit $V \to \infty$ at a fixed subsystem fraction $f = V_A/V$ and particle number $n = N/V$. We find that

$$(\Delta S_A)^2_{G,N} = \ln(1-f) + f + f^2 + f^2(2n-1)\ln\left(\frac{1}{n} - 1\right) + f(f-1)(n-1)n\ln^2\left(\frac{1}{n} - 1\right) + o(1) \quad (116)$$

for $0 < f \le n \le \frac{1}{2}$. We visualize this result in Fig. 9. At $n = \frac{1}{2}$, the expression above simplifies to $\lim_{V\to\infty}(\Delta S_A)^2_{G,N}(f,\frac{1}{2}) = f + f^2 + \ln(1-f)$, which is exactly twice the variance (93) found for the entanglement entropy of *all* fermionic Gaussian states [159].

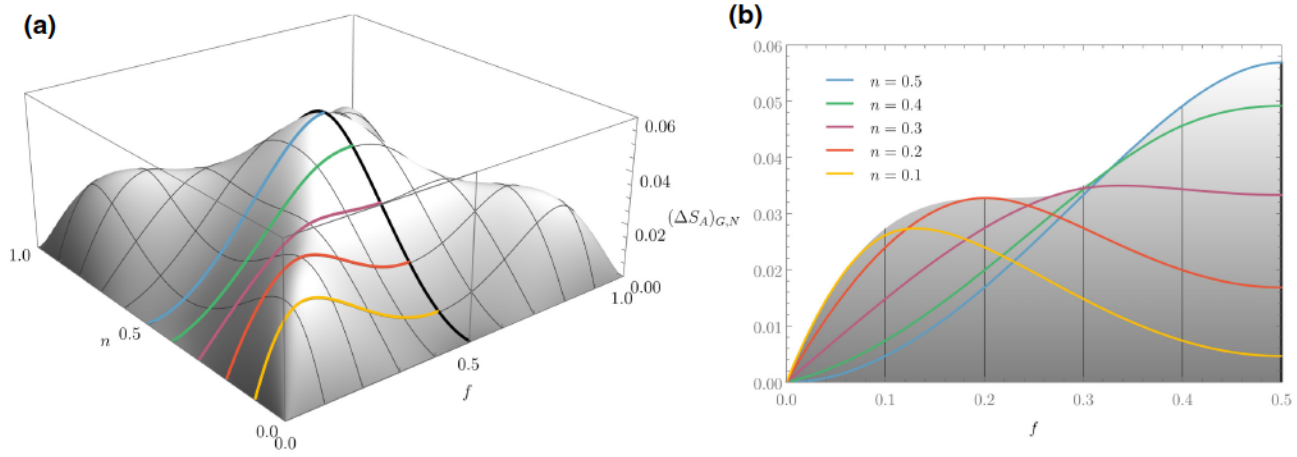


FIG. 9. The leading order of the standard deviation $(\Delta S_A)_{G,N}$ of the entanglement entropy S_A, from Eq. (116). (a) Three-dimensional plot as a function of the subsystem fraction $f = V_A/V$ and the filling ratio $n = N/V$. (b) Results at fixed n plotted as functions of f. The colored curves in the right plot are the sections with the same color in the left plot.

3. *Weighted average and variance*

Following the definitions in Sec. II B 3, we can define the weighted average entanglement entropy of Gaussian states

$$\langle S_A\rangle_{G,w} = \sum_{N=0}^{V} P_N \langle S_A\rangle_{G,N}, \tag{117}$$

where P_N is a function of w as defined in Eq. (62), and produces an average filling ratio $\bar{n} = \bar{N}/V = 1/(1+e^w)$ [see Eq. (63)]. In the thermodynamic limit, $V\to\infty$, we can approximate the binomial distribution by a continuous probability distribution $\varrho_w(n)$, which approaches a Gaussian plus corrections (see Appendix D 2 a) that becomes increasingly peaked at $n = \bar{n}$.

When we average over all Gaussian states with a fixed number of particles, a natural weight is given when $w = 0$, i.e., we weigh each particle-number sector by its Hilbert space dimension Ω_N. If we would draw a random eigenstate of a random quadratic number-conserving Hamiltonian (see Sec. IV A), the resulting *eigenstate entanglement entropy* will thus correspond to $w = 0$. The leading-order average for eigenstates of such Hamiltonians was derived in Ref. [61], and was later shown numerically [62] and analytically [159] to coincide with the leading-order average over eigenstates of random quadratic Hamiltonians without number conservation or, equivalently, over all fermionic Gaussian states. The present calculation explains these coincidences by showing explicitly how the average at $w = 0$ corresponds to the peak of the binomial distribution at $n = \bar{n} = \frac{1}{2}$, so at leading order $\langle S_A\rangle_{G,w=0} = \langle S_A\rangle_G + O(1)$.

We calculate the binomial average over N analytically up to order $1/V$ in Appendix D 2 a. The resulting leading-order behavior, as a function of $\bar{n}$ and f, is given by

$$\langle S_A\rangle_{G,w} = \begin{cases} [(f-1)\ln(1-f) + f[(\bar{n}-1)\ln(1-\bar{n}) - 1 - \bar{n}\ln\bar{n}]]V - \dfrac{f}{2} \\ \quad + \dfrac{(f-2)f}{12(f-1)}\dfrac{1}{V} + O(1/V^3), & f < \bar{n} \le \frac{1}{2}, \\ [(f-1)\bar{n}\ln(1-f) - \bar{n}(1+f\ln f) + (\bar{n}-1)\ln(1-\bar{n})]V - \dfrac{\bar{n}}{2} \\ \quad + \dfrac{\bar{n}(1-f+f^2)}{12f(1-f)}\dfrac{1}{V} + O(1/V^3), & \bar{n} < f \le \frac{1}{2}, \\ [(f^2-1)\ln(1-f) - f(1+f\ln f)]V - \dfrac{f}{2} + \sqrt{\dfrac{(1-f)f}{18\pi}}\dfrac{1}{\sqrt{V}} \\ \quad + \dfrac{1+f}{24(1-f)}\dfrac{1}{V} + O(1/V^{3/2}), & f = \bar{n} < \frac{1}{2}, \\ \left[\ln 2 - \dfrac{1}{2}\right]V - \dfrac{1}{4} + \dfrac{1}{3\sqrt{2\pi}}\dfrac{1}{\sqrt{V}} + \dfrac{1}{8}\dfrac{1}{V} + O(1/V^{3/2}), & f = \bar{n} = \frac{1}{2}, \end{cases} \tag{118}$$

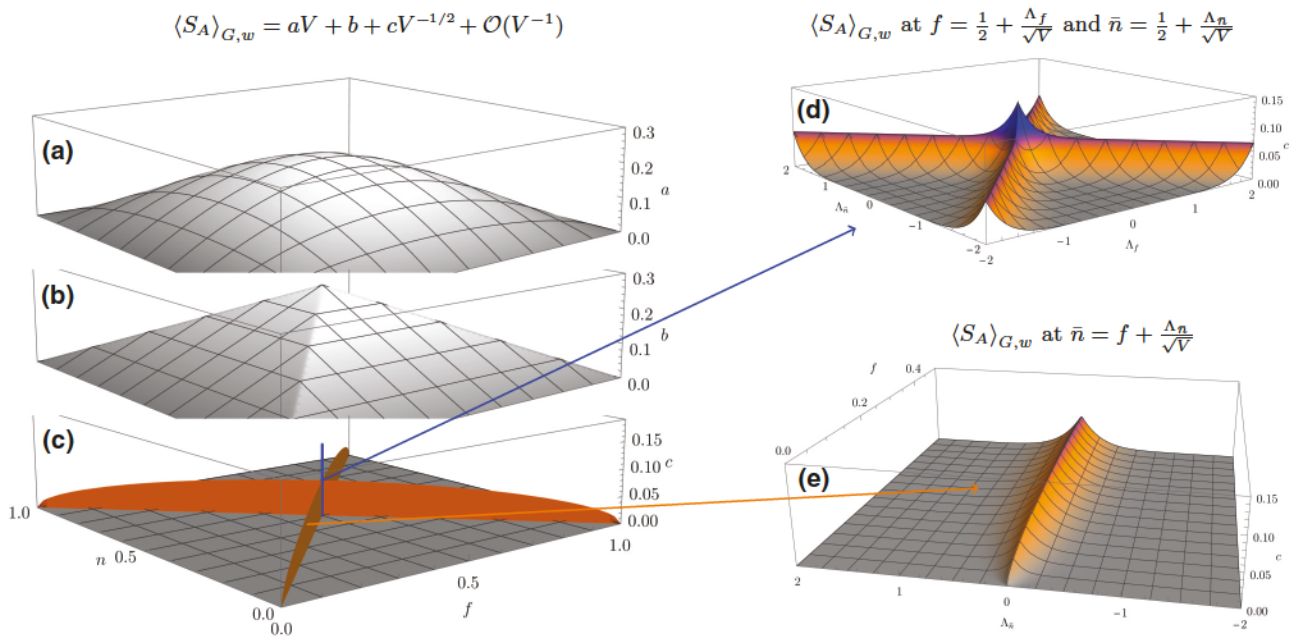


FIG. 10. The subsystem entanglement entropy $\langle S_A\rangle_{G,w}$ from Eq. (119) as viewed from the contributions of the first three terms in the expansion in V. (a)–(c) Three-dimensional plots as functions of the subsystem fraction $f = V_A/V$ and the filling ratio $n = N/V$. (d) Resolving b for $\bar{n} = \frac{1}{2} + \Lambda_{\bar{n}}/\sqrt{V}$ and $f = \frac{1}{2} + \Lambda_f/\sqrt{V}$ around $f = \bar{n} = \frac{1}{2}$, as given by Eq. (D34). (e) Resolving discontinuities of c for $\bar{n} = f + \Lambda_{\bar{n}}/\sqrt{V}$ around $f = \bar{n} < \frac{1}{2}$ given by Eq. (D44).

for $f, \bar{n} \leq \frac{1}{2}$. To $O(1/\sqrt{V})$, the above results can be recast in the following compact equation:

$$\langle S_A\rangle_{G,w} = [(\mu_- - 1)\ln(1-\mu_-) + \mu_-[(\mu_+ - 1)\ln(1-\mu_+) - 1 - \mu_+ \ln \mu_+]]V - \frac{\mu_-}{2} + \delta_{f,\bar{n}}\left(\sqrt{\frac{f\,(1-f\,)}{18\pi}} + \delta_{f,1/2}\frac{1}{6\sqrt{2\pi}}\right)\frac{1}{\sqrt{V}} + O\left(\frac{1}{V}\right) \tag{119}$$

with $\mu_+ = \max(f\,, \bar{n})$ and $\mu_- = \min(f\,, \bar{n})$. Equation (119) is visualized in Figs. 10(a)–10(c). Note that $\langle S_A\rangle_{G,w}$ satisfies the particle-subsystem symmetry $\bar{n} \leftrightarrow f$ only up to $1/\sqrt{V}$, which is not surprising considering that this symmetry is only exact for averages at *fixed* N. We compared our analytical results with those of numerical calculations. In Fig. 11, we show some of the finite-size scaling analyses that we carried out.

We note that the asymptotic behavior of the average entanglement entropy $\langle S_A\rangle_{G,w}(f)$ is characterized by a nonanalytic behavior along the symmetry axes $f = \bar{n}$ and $f = 1 - \bar{n}$. This gives rise to distinct corrections in the thermodynamic limit. Most importantly, we have an enhancement of order $1/\sqrt{V}$ whenever $f = \bar{n}$ and, in particular, at $f = \bar{n} = \frac{1}{2}$. These regimes are due to the various symmetries of the average entanglement entropy $\langle S_A\rangle_{G,N}$ as a function of V_A and N. In particular, there is the aforementioned particle-subsystem symmetry, which states that $\langle S_A\rangle_{G,N}$ is invariant under interchanging $N \leftrightarrow V_A$. This symmetry has to be put in by hand and results in Kronecker deltas as in Page's setting, although there the Kronecker deltas resulted from a different symmetry. In Appendix D 2 c, we resolve the Kronecker deltas in Eq. (119) by studying the asymptotics for either $\bar{n} = f + \Lambda_f/\sqrt{V}$ or $\bar{n} = \frac{1}{2} + \Lambda_f/\sqrt{V}$ and $f = \frac{1}{2} + \Lambda_f/\sqrt{V}$. The resulting expressions (D35) and (D43) are visualized in Figs. 10(d) and 10(e).

When computing the average entanglement entropy $\langle S_A\rangle_{G,w}$ from Eq. (D26) for weight parameter w and subsystem size $V_A \leq V/2$, one expects transitions to occur whenever the expectation value $\bar{N} = V\bar{n} = Ve^{-w}/(1+e^{-w})$ crosses a symmetry axis, where there is a discontinuity in the third derivative of $\langle S_A\rangle_{G,N}$. Interestingly, we only have a transition at $\bar{n} = f$ (i.e., $\bar{N} = V_A$) and $\bar{n} = f = \frac{1}{2}$, but not at $\bar{n} = \frac{1}{2}$ (i.e., $\bar{N} = V/2$) due to the fact that Eq. (113) is symmetric in $N \leftrightarrow V - N$, so that our function $\langle S_A\rangle_{G,N}$ is analytic across the particle-hole symmetry. The reason for this is that $\langle S_A\rangle_{G,N}$ has continuous derivatives up to order

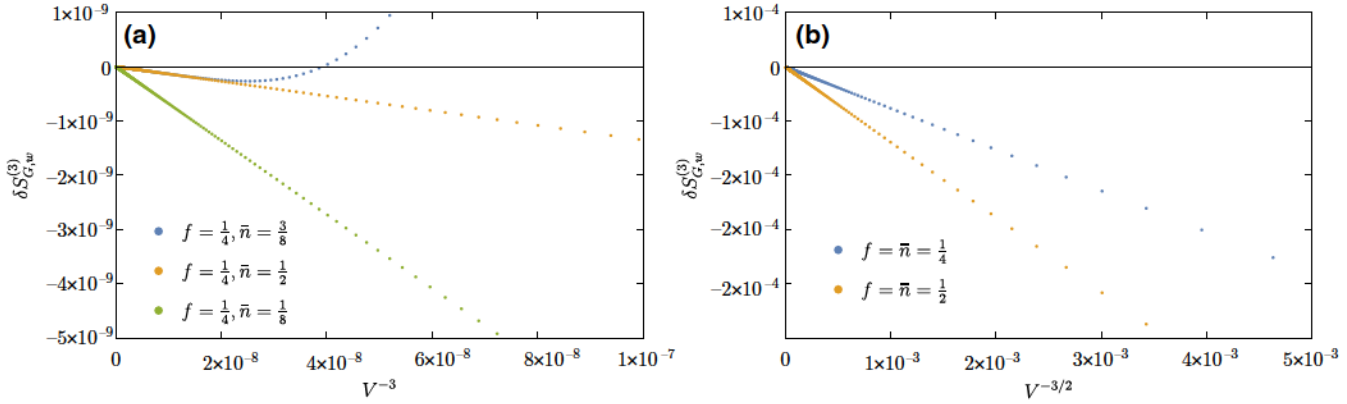


FIG. 11. Asymptotics of $\langle S_A\rangle_{G,w}$. We compare the exact values of the average entanglement entropy $\langle S_A\rangle_{G,w}$ computed numerically by evaluating the sum in Eq. (117) using Eq. (114) for $V \leq 2000$, with the asymptotic results, Eq. (118), i.e., we show the difference $\delta S^{(3)}_{G,w} = \langle S_A\rangle_{G,w} - \langle S_A\rangle^{(3)}_{G,w}$, where $\langle S_A\rangle^{(3)}_{G,w}$ corresponds to expansion (118) up to order $1/V$. In particular, one can see that the next order is $1/V^{3/2}$ if $f = \bar{n}$, and $1/V^3$ otherwise.

three at $\bar{n} = \frac{1}{2}$, so that we would not see a transition if we only expand up to order $1/V$. However, we expect that the fifth-order term of $s^G_A(f, n) = \lim_{V\to\infty} \langle S_A\rangle_{G,N}/V$ [see also Eq. (D28)] expanded in powers of $(\bar{n} - n)$ will contribute a square root enhancement of order $V^{-3/2}$ for $\bar{n} = \frac{1}{2}$.

Finally, let us also comment about the leading order of the variance at fixed weight parameter w. The leading-order contribution is due to the variance $(\Delta n)^2$ in the number of particles, Eq. (64). As a result, while the variance at fixed particle number is $O(1)$, Eq. (116), the variance at fixed w scales linearly with the volume

$$(\Delta S_A)^2_{G,w} = \begin{cases} \bar{n}(1-\bar{n})\left[\ln\left(\dfrac{\bar{n}}{1-\bar{n}}\right)\right]^2 f^2 V + o(V), & f \leq \bar{n}, \\ \bar{n}(1-\bar{n})[(1-f)\ln(1-f) + f\ln f + \ln(1-\bar{n})]^2 V + o(V), & f > \bar{n}, \end{cases} \tag{120}$$

with $f \leq \frac{1}{2}$. Note that, at $w = 0$ (corresponding to $\bar{n} = \frac{1}{2}$), the leading-order $O(V)$ term vanishes. In general, we have $\lim_{V\to\infty}(\Delta S_A)_{G,w}/\langle S_A\rangle_{G,w} = 0$, which shows that in the thermodynamic limit the average (118) also gives the typical value of the entanglement entropy.

IV. EXACT RELATION TO RANDOM HAMILTONIANS

So far, we have focused on ensembles of quantum states and computed statistical properties of the entanglement entropy with respect to the following six ensembles: (1a) random states, (2a) random states with fixed total particle number, (3a) weighted averages over random states with fixed total particle number, (1b) random fermionic Gaussian states, (2b) random fermionic Gaussian states with fixed total particle number, and (3b) weighted averages over random fermionic Gaussian states with fixed total particle number. In this section, we shift the focus from ensembles of quantum states to random Hamiltonians, their eigenstates, and their dynamics.

A. Random many-body Hamiltonians

Ensembles (1a), (2a), and (3a) can be realized using eigenstates (even only ground states) of random Hamiltonians that are traditional random matrices. The ensuing Hamiltonians give an *exact* correspondence to Page's setting, i.e., the averages and variances will agree at all orders (meaning even at finite V) when the respective random Hamiltonian satisfies the properties discussed next.

We first consider case (1a), for which the number of particles is not fixed. The state vector in this case explores the entire sphere of the Hilbert space $\mathcal{H}$. Thus, any random Hamiltonian that creates a Haar-distributed random state vector is suitable. For instance, let us study the random-matrix Hamiltonian

$$\hat{H}_{1a} = \sum_{\kappa,\lambda=1}^{2^V} C_{\kappa\lambda}\,|v_\kappa\rangle\langle v_\lambda|\,, \tag{121}$$

where $|v_\lambda\rangle$ is an orthonormal basis of the Hilbert space and $C_{\kappa\lambda}$ is a Haar-distributed random matrix. To get Haar-distributed eigenvectors, the diagonalization $C = U^\dagger E U$

must involve random matrices U drawn from the Haar measure of $U(2^V)$, while the distribution of the eigenvalues appearing in the diagonal matrix E can be arbitrary. A simple, and one of the most common examples of such a distribution for C is given by the GUE [117,118,136],

$$P(\hat{H}_{1a}) = 2^{-2^{V-1}}\pi^{-2^{2V-1}}\exp\left[-\frac{1}{2}\sum_{\kappa,\lambda=1}^{2^V}|C_{\kappa\lambda}|^2\right] = 2^{-2^{V-1}}\pi^{-2^{2V-1}}e^{-\mathrm{Tr}\,\hat{H}_{1a}^2/2}. \quad (122)$$

To relate the Hamiltonian $\hat{H}_{1a}$ to many-body Hamiltonians, we rewrite it as a polynomial in fermionic creation and annihilation operators,

$$\hat{H}_{1a} = \sum_{l=0}^{2V}\sum_{j_1,\ldots,j_l=1}^{2V} c^{(l)}_{j_1,\ldots,j_l}\hat{\xi}_{j_1}\cdots\hat{\xi}_{j_l} \quad (123)$$

with $\{\hat{\xi}_j\}_{j=1,\ldots,2V} = (\hat{f}_1,\ldots,\hat{f}_V,\hat{f}_1^\dagger,\ldots,\hat{f}_V^\dagger)$. The coefficients $c^{(l)}_{j_1,\ldots,j_l}$ satisfy symmetries that reflect the anticommutation relations, $\{\hat{f}_k,\hat{f}_l\} = \{\hat{f}_k^\dagger,\hat{f}_l^\dagger\} = 0$ and $\{\hat{f}_k,\hat{f}_l^\dagger\} = \delta_{kl}$, and the Hermiticity of $\hat{H}_{1a}$, and the fact that in each sum over $c^{(l)}_{j_1,\ldots,j_l}$ there are exactly l operators involved that cannot be reduced to a smaller order of a many-body interaction. Exploiting the unitary matrix T in Eq. (95), in particular going into a Majorana basis, shows that $\tilde{c}^{(l)}_{k_1,\ldots,k_l} = \sum_{j_1,\ldots,j_l=1}^{2V} c^{(l)}_{j_1,\ldots,j_l}\prod_{a=1}^{l} T_{j_a k_a}$ is totally skew symmetric in the indices and is real when $l(l-1)/2$ is even and imaginary when $l(l-1)/2$ is odd.

The statistical distribution of the coefficients $c^{(l)}_{j_1,\ldots,j_l}$ is determined by the distribution of matrix $C_{\mu\nu}$. The best way to see this is to go into the Majorana basis $\gamma_1,\ldots,\gamma_{2V}$ via relation (94). Then, one needs to take into account the normalization $\gamma_j^2 = \frac{1}{2}\mathbb{1}_{2^V}$ to determine this distribution, which leads to

$$P(\hat{H}_{1a}) = \prod_{l=1}^{2V}\prod_{1\le j_1<\cdots<j_l\le l}\sqrt{\frac{2^{V-l-1}l!}{\pi}} \times \exp[-2^{V-l-1}l!\,|\tilde{c}^{(l)}_{j_1,\ldots,j_l}|^2]. \quad (124)$$

To derive this result, one needs to use the fact that the trace of a product of γ_j is only nonvanishing when each γ_j appears with an even number in this product.

The statistical properties of the entanglement entropy in eigenstates of the random Hamiltonian $\hat{H}_{1a}$ are described exactly by the results of Sec. II A since the Hamiltonian is invariant under the conjugations of the unitary group $U(2^V)$. Hence, its eigenstates are uniformly distributed over the unit sphere in $\mathcal{H}$. We emphasize that this Hamiltonian is not parity preserving such that the superselection rule (either only even or odd powers in ξ_j) does not apply.

SYK models [127,128] are related to this construction. For the q-body SYK (or in short SYKq), one sets $c^{(l)}_{j_1,\ldots,j_l} = 0$ for all $l \neq q$, and chooses the Gaussian distribution for $c^{(q)}_{j_1,\ldots,j_q}$, as we have done here. For a fixed q, these models have more symmetries than we have by adding up all q. Thus, they reflect all kinds of random matrix symmetry classes (actually one can find all ten classes of the Altland-Zirnbauer classification [106,182]) and follow the Bott periodicity [183] in V and q; see Refs. [109,110]. When mixing SYKq with different q's, it is likely that in the large-V limit one ends up in the same class as our random Hamiltonian $\hat{H}_{1a}$. However, for a fixed q, it might happen that the subleading orders differ from our results.

We turn now to case (2a), in which we need to implement number conservation in the random Hamiltonian. Based on the direct sum decomposition (31) of the Hilbert space, we define an orthonormal basis $|v_\mu^{(N)}\rangle$ of the N-particle Hilbert space $\mathcal{H}^{(N)}$ of dimension d_N given in Eq. (34). To parallel (2a), we consider a random-matrix Hamiltonian in this particle sector given by

$$\hat{H}_{2a}^{(N)} = \sum_{\kappa,\lambda=1}^{d_N}\tilde{C}^{(N)}_{\kappa\lambda}\,|v_\kappa^{(N)}\rangle\,\langle v_\lambda^{(N)}|\,, \quad (125)$$

where the Hermitian matrix $\tilde{C}^{(N)} = \{\tilde{C}^{(N)}_{\kappa\lambda}\}_{\kappa,\lambda=1,\ldots,d_N}$ is assumed to be a $U(d_N)$ group invariant random matrix, i.e., it can be diagonalized, $\tilde{C}^{(N)} = U^\dagger E U$, via a Haar-distributed unitary matrix $U \in U(d_N)$. Anew, the GUE is a simple example of such a distribution [in particular, Eq. (122) with 2^V replaced by d_N], but the class is more general and does not constrain the diagonal matrix E that comprises the energies.

As before, we want to express the Hamiltonian $\hat{H}_{2a}^{(N)}$ in terms of a polynomial in fermionic creation and annihilation operators. This cannot be done without adding an orthogonal projection onto the Hilbert space with N particles. Instead of $\hat{H}_{2a}^{(N)}$, we consider the direct sum

$$\hat{H}_{2a/3a} = \bigoplus_{N=0}^{V}\hat{H}_{2a}^{(N)}, \quad (126)$$

which is now expressible in terms of $(\hat{f}_1,\ldots,\hat{f}_V,\hat{f}_1^\dagger,\ldots,\hat{f}_V^\dagger)$. While an arbitrary random Hamiltonian $\hat{H}_{2a}^{(N)}$ has d_N^2 degrees of freedom, the random Hamiltonian $\hat{H}_{2a/3a}$ has $\sum_{N=0}^{V} d_N^2 = \binom{2V}{V}$ degrees of freedom. The subscript already indicates that this random Hamiltonian also describes case (3a), where one averages over states with different particle numbers.

Another subtle point is that particle-number conservation does not allow any odd powers of these operators nor

does it allow an unbalanced number of creation and annihilation operators. Thus, the general form of $\hat{H}_{2a/3a}$ is given by

$$\hat{H}_{2a/3a} = \sum_{l=0}^{V} \sum_{j_1,\ldots,j_l,k_1,\ldots,k_l=1}^{V} c^{(l)}_{j_1\ldots j_l,k_1\ldots k_l} \hat{f}_{j_1}^{\dagger} \cdots \hat{f}_{j_l}^{\dagger} \hat{f}_{k_l} \cdots \hat{f}_{k_1}, \tag{127}$$

which, by construction, commutes with the particle-number operator. The coefficients $c^{(l)}_{j_1\ldots j_l,k_1\ldots k_l}$ are totally skew symmetric in the first l indices as well as skew symmetric in the last l ones. To ensure the Hermiticity of $\hat{H}_{2a/3a}$, it holds that the $\binom{V}{l} \times \binom{V}{l}$ matrix $C^{(l)}$ created by $c^{(l)}_{j_1\ldots j_l,k_1\ldots k_l}$ for a fixed l is Hermitian. It is simple to check that the degrees of freedom given by all $c^{(l)}_{j_1\ldots j_l,k_1\ldots k_l}$ for all l add up to the formerly obtained $\binom{2V}{V}$.

The statistical distribution of the coefficients $c^{(l)}_{j_1\ldots j_l,k_1\ldots k_l}$ is determined by the distribution of the matrices $\tilde{C}^{(N)}$ for all N. Since the coefficients $c^{(l)}_{j_1\ldots j_l,k_1\ldots k_l}$ are linear combinations of matrix elements of all $\tilde{C}^{(0)}, \ldots, \tilde{C}^{(V)}$, their distributions will in general be coupled despite the fact that one can choose the distributions of $\tilde{C}^{(N)}$ to be completely independent and not identical.

As mentioned before, the simplest choice are GUEs in $\tilde{C}^{(N)}$ with the same variance, which yields

$$P(\hat{H}_{2a/3a}) = \frac{\sqrt{\det \Sigma}}{2^{2^{V-1}} \pi^{\ell^2/2}} \exp\left[-\frac{1}{2} \operatorname{Tr} \hat{H}^2_{2a/3a} \right], \tag{128}$$

where $\ell = \binom{2V}{V}$. The matrix Σ is $\ell \times \ell$ dimensional, with matrix entries $\Sigma_{j_1\ldots j_l j'_1\ldots j'_l,k_1\ldots k_m k'_1\ldots k'_m}$ equal to $(l!)^2(m!)^2 \operatorname{Tr}[\hat{f}_{j'_1}^{\dagger} \cdots \hat{f}_{j'_l}^{\dagger} \hat{f}_{j_l} \cdots \hat{f}_{j_1} \hat{f}_{k_1}^{\dagger} \cdots \hat{f}_{k_m}^{\dagger} \hat{f}_{k'_m} \cdots \hat{f}_{k'_1}]$ with $l, m = 0, \ldots, V$ and $1 \le j_1 < \cdots < j_l \le V$, $1 \le j'_1 < \cdots < j'_l \le V$, $1 \le k_1 < \cdots < k_m \le V$ as well as $1 \le k'_1 < \cdots < k'_m \le V$. It is very sparse because the trace is only nonvanishing when each $\hat{f}_j$ appears as often as its Hermitian conjugate $\hat{f}_j^{\dagger}$, and it might happen that there are operators that annihilate each other, such as $\hat{f}_j^2 = (\hat{f}_j^{\dagger})^2 = 0$. Thence, Σ only contains entries equal to $\pm 2^L (l!)^2 (m!)^2$ with an $L \in \{0, \ldots, V\}$. The factor $1/\pi^{\ell^2/2}$ reflects the number of total degrees of freedom while $1/2^{2^{V-1}}$ results from the counting of how many real coefficients exist, namely, $\sum_{l=0}^{V} \binom{V}{l} = 2^V$. The simplest nontrivial example is obtained for $V = 2$ where Σ is equal to

$$\Sigma = \begin{pmatrix} 4 & 2 & 0 & 0 & 2 & 4 \\ 2 & 2 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 2 & 4 \\ 4 & 4 & 0 & 0 & 4 & 4 \end{pmatrix} \tag{129}$$

with the ordering of the coefficients $(c^{(0)}, c^{(1)}_{1,1}, c^{(1)}_{1,2}, c^{(1)}_{2,1}, c^{(1)}_{2,2}, c^{(2)}_{12,12})$. This example underscores how intertwined the correlations of the coefficients $c^{(l)}_{j_1\ldots j_l,k_1\ldots k_l}$ can become if one wants to write the exact realization of cases (2a) and (3a) in terms of annihilation and creation operators. Thence, from a practical point of view, it is simpler to generate these random Hamiltonians in terms of the independent GUE generated coefficients $\tilde{C}^{(N)}_{\kappa\lambda}$.

When focusing on a certain particle-number sector, the statistical properties of the entanglement entropy in eigenstates of $\hat{H}_{2a/3a}$ are described exactly by the results of Sec. II B, meaning case (2a). When considering all sectors, we have case (3a). For distribution (128), picking any eigenstate is equally likely, this yields the weight $\binom{V}{N}$ to find a state in the sector $\mathcal{H}^{(N)}$, which corresponds to a weighted average with $w = 0$. Implementing a weighted average with $w > 0$ is also possible, but must be largely done by hand, i.e., we would organize the eigenstates of a random Hamiltonian based on their particle number and then choose one at random using the statistical weight encoded by w.

Many-body interacting Hamiltonians studied in nuclear physics [113–115,120–122] are related to these kinds of Hamiltonian. They, as well as the SYK models, are called embedded random matrices [116,123–126]. For instance, for a q-body Hamiltonian, we set $c^{(l)}_{j_1\ldots j_l,k_1\ldots k_l} = 0$ for all $l \neq q$ and choose the above Gaussian distribution for $c^{(l)}_{j_1\ldots j_l,k_1\ldots k_l}$. As for case (1a) and SYKq for a fixed q, the many-body Hamiltonian may satisfy additional global symmetries so that subleading terms may deviate from our results. However, we expect that a mixture of q-body interactions should speed up the convergence to the leading-order result in the thermodynamic limit $V \to \infty$.

B. Random quadratic Hamiltonians

Case (1b) for random pure fermionic Gaussian states is obtained from $\hat{H}_{1a}$ by setting all coefficients $c^{(l)}_{i_1,\ldots,i_l} = 0$ whenever $l \neq 2$ in Eq. (123); the resulting random

quadratic Hamiltonian reads

$$\hat{H}_{1b} = \sum_{i,j=1}^{2V} c_{ij}^{(2)} \hat{\xi}_i \hat{\xi}_j \tag{130}$$

with coefficients $c_{ij}^{(2)}$ drawn from a probability distribution that depends only on matrix invariants of $TC_{(2)}T^T$ with $C_{(2)} = \{c_{ij}^{(2)}\}_{i,j=1,\ldots,2V}$, such as traces $\mathrm{Tr}(TC_{(2)}T^T)^{2k}$. Then the invariance under $O(2V)$ is guaranteed, which is needed for the uniformly distributed pure fermionic Gaussian states that are the eigenvectors of this Hamiltonian. The Gaussian choice as the distribution of the coefficients $c_{ij}^{(2)}$ is equal to

$$P(\hat{H}_{1b}) = \prod_{1\le j_1<j_2\le 2V} \sqrt{\frac{2^V}{\pi}} e^{-2^V |\tilde{c}_{j_1 j_2}^{(2)}|^2}, \tag{131}$$

when starting from the distribution of $\hat{H}_{1a}$. Since $C_{(2)}$ is unitarily equivalent to the real antisymmetric $2V \times 2V$ matrix $2TCT^T$, which can be seen in the Majorana basis, the Gaussian ensemble is also known as the Gaussian real antisymmetric ensemble [117] or GAOE in Ref. [104] (it also has other acronyms such as BdG-S in Ref. [109] or class BD in Refs. [106,182] as it is not as canonical as the GUE). We would like to emphasize that the real antisymmetry of TCT^T is the reason why there is no minus sign in the exponent of the distribution, i.e., $\mathrm{Tr}(TC_{(2)}T^T)^2 < 0$. When going over from the coefficients $\tilde{c}_{jk}^{(2)}$ in the Majorana basis to $c_{jk}^{(2)}$ in the creation-annihilation basis, we need to employ $TT^T = \tau_1 \otimes \mathbb{1}_V/2$ and properly normalize the distribution (multiplying it by the Jacobian), which gives rise to a factor $2^{-V(V-1)}$. Then, the distribution is

$$P(\hat{H}_{1b}) = \frac{2^{V(V-1)(V-3)/2}}{\pi^{V(V-1)/2}} \exp[2^{V-3}\,\mathrm{Tr}(C_{(2)}[\tau_1 \otimes \mathbb{1}_V])^2]. \tag{132}$$

Certainly, the factors of 2 can be absorbed when choosing a general standard deviation for the Gaussian ensemble.

The eigenvectors of $\hat{H}_{1b}$ are Haar distributed with respect to the group $O(2V)$, as discussed in Ref. [173], though they are slightly rotated by the unitary matrix $\sqrt{2}T$; see Eq. (95). The eigenstates of this Hamiltonian are fermionic Gaussian states, and the statistical properties of the entanglement entropy in these eigenstates are described exactly by the results of Sec. III B.

For cases (2b) and (3b), we can repeat the fixed-particle-number steps for random quadratic Hamiltonians, specifically, we set $c_{j_1\ldots j_l,k_1\ldots k_l}^{(l)} = 0$ for all $l \neq 1$ in Eq. (127). The most general quadratic fermionic Hamiltonian that commutes with the total number operator $\hat{N}$ defined in Eq. (31) can be written as

$$\hat{H}_{2b/3b} = \sum_{i,j=1}^{V} \tilde{c}_{i,j}^{(2)} \hat{f}_i^\dagger \hat{f}_j \tag{133}$$

with the coefficients $\tilde{c}_{i,j}^{(2)}$ drawn from an ensemble that is invariant under the conjugate action of $U(V)$; in particular, the Hermitian random matrix $\tilde{C} = \{\tilde{c}_{i,j}^{(2)}\}_{i,j=1,\ldots,V}$ and $U\tilde{C}U^\dagger$ with an arbitrary fixed $U \in U(V)$ are equally distributed. The simplest example for such a distribution is a $V \times V$ GUE. That one agrees with a quadratic particle-number-conserving SYK model in its Dirac fermion formulation (the Dirac SYK2 model for short), which is a free random Hamiltonian. The volume-law coefficient $s_A^G(f,n) = \lim_{V\to\infty} \langle S_A\rangle_{G,N}/V$ [see Eq. (115)], was first conjectured in the context of quadratic Hamiltonians whose single-particle eigenstates can be well approximated by eigenstates of random matrices [61].

In general, all eigenstates of nondegenerate quadratic random Hamiltonians of this $U(V)$ invariance are Gaussian states. Note that degenerate eigenspaces will contain superpositions of Gaussian states, which are not Gaussian themselves. For most random Hamiltonians, the subset of degenerate Hamiltonians is a set of measure zero, so it can be ignored.

Anew, we have treated cases (2b) and (3b) with a single Hamiltonian. For the former case, one needs to restrict the Hamiltonian eigenstates to the sector with N particles, while for the latter case, one needs to compute the weighted average over all eigenstates. For the average over all sectors, we again pick an arbitrary eigenstate of the Hamiltonian $H_{2b/3b}$. A random many-body eigenstate of this Hamiltonian will be in the sector with N particles with probability $2^{-V}\binom{V}{N}$, so a *typical* eigenstate will have the entanglement entropy $\langle S_A\rangle_{G,w=0}$ found in Eq. (118). If one restricts the analysis to eigenstates with N particles then the statistical properties of the entanglement entropy are described exactly by the results of Sec. III C 2.

In Fig. 12, we show numerical results for the average entanglement entropy of the particle-number-conserving SYK2 model (133). The numerical results in Fig. 12(a) behave as expected from the analytical predictions for the volume-law contribution (118) as a function of V, and the ones in Fig. 12(b) behave as expected from the analytical predictions for the subleading terms given in Eq. (118). In the simulations, the average is carried out over all sectors with a fixed particle number, where the weight is given by the dimension of the sector. This corresponds to $w = 0$ or, equivalently, $\bar{n} = \frac{1}{2}$.

C. Dynamical averages

Another important application of our results concerns the study of quantum many-body stochastic dynamics

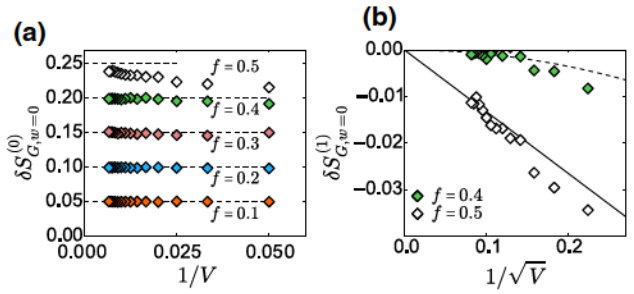


FIG. 12. Finite-size scaling of the average entanglement entropy difference for random quadratic fermionic Hamiltonians. (a) Plot of $\delta S^{(0)}_{G,w=0} = \langle S_A\rangle^{(0)}_{G,w=0} - \bar{S}$, where $\langle S_A\rangle^{(0)}_{G,w=0}$ is the volume-law term in Eq. (118) and $\bar{S}$ denotes the numerical results for eigenstates of the particle-number conserving SYK2 model at $\bar{n} = \frac{1}{2}$, versus $1/V$ at subsystem fractions (from bottom to top) $f = 0.1, 0.2, 0.3, 0.4, 0.5$. Horizontal lines are constants $f/2$. (b) Plot of $\delta S^{(1)}_{G,w=0} = \langle S_A\rangle^{(1)}_{G,w=0} - \bar{S}$ versus $1/\sqrt{V}$ at subsystem fractions $f = 0.4$ and 0.5, where $\langle S_A\rangle^{(1)}_{G,w=0} = \langle S_A\rangle^{(0)}_{G,w=0} - f/2$. The lines are the second subleading terms from Eq. (118). Specifically, the solid line (corresponding to $f = 0.5$) is the function $-1/(3\sqrt{2\pi})(1/\sqrt{V})$, and the dashed line (corresponding to $f = 0.4$) is the function $-f(f-2)/[12(f-1)](1/V)$. The numerical results for $\bar{S}$ are from Ref. [61].

[184]. We consider a time-dependent random Hamiltonian (with or without particle conservation) for which the time derivative $\dot{C}_{\mu\nu}$ or $\dot{c}^{(2)}_{ij}$ of the coefficient matrices are delta correlated in time, i.e., for which we have

$$\left\langle \dot{C}_{\mu\nu}(t)\, \dot{C}_{\tau\sigma}(t')\right\rangle \propto \gamma\delta(t-t'), \tag{134}$$

$$\left\langle \dot{c}^{(2)}_{ij}(t)\, \dot{c}^{(2)}_{kl}(t')\right\rangle \propto \gamma\delta(t-t'). \tag{135}$$

If we evolve an initial quantum state (Gaussian for a quadratic Hamiltonian and with fixed particle number when it is a particle-conserving Hamiltonian), the time evolution leads to an *ergodic* exploration of the respective space of states considered in (1a), (1b), (2a), and (2b), provided that the strength γ of the fluctuations is sufficiently large in the thermodynamic limit. Indeed, it is known from the Brownian motion on the unitary group $U(d)$ (see Ref. [185] for the random matrix version and Ref. [186] for the corresponding eigenvalues) that there is a phase transition at a critical value $t_{\text{crit}} = t_{\text{crit}}(d) \propto d$ when the matrix size d goes to infinity. In Ref. [186] it was found that the level density of such a unitary random matrix does not have the entire complex unit circle as a support. We presume that this has a direct consequence for the states, too, since they have been constructed via the natural group action on the states.

This implies that the time evolution will only uniformly sample the full ensemble of states with respect to the Haar measure when the time is sufficiently large compared to the underlying group dimension, which is $U(2^V)$ or $\bigoplus_{N=0}^{V} U(d_N)$ for the Page setting and $O(2V)$ or $U(V)$ for pure fermionic Gaussian states without and with particle-number conservation, respectively. Then, the asymptotic time average of the entanglement entropy will coincide with the respective averages computed in the previous sections. Moreover, we also expect that the standard deviation gives a good approximation of the expected fluctuations about this average over time.

Again, this analysis applies to both general and quadratic Hamiltonians. The latter corresponds to the quantum simple symmetric exclusion process introduced in Refs. [187,188], for which the statistical properties of the Renyi entropies were studied in Ref. [181].

V. RELATION TO PHYSICAL HAMILTONIANS

The goal of this section is to contrast the results for the entanglement entropies from Secs. II and III to those in eigenstates of physical Hamiltonians on a lattice. For the latter, we mostly have in mind local Hamiltonians with short-range hoppings and interactions (involving only a few neighboring lattice sites). Results for interacting Hamiltonians are discussed in Sec. V A, while results for quadratic (noninteracting) Hamiltonians are discussed in Secs. V B and V C.

A. Quantum-chaotic interacting model

We focus on a model of interacting hard-core bosons in a one-dimensional (1D) lattice with V sites, as described by the Hamiltonian

$$\hat{H}_{\text{HCB}} = -t_1 \sum_{l=1}^{V} (\hat{b}^\dagger_{l+1}\hat{b}_l + \hat{b}^\dagger_l \hat{b}_{l+1}) - t_2 \sum_{l=1}^{V} (\hat{b}^\dagger_{l+2}\hat{b}_l + \hat{b}^\dagger_l \hat{b}_{l+2}) + V_1 \sum_{l=1}^{V} \hat{n}_l\hat{n}_{l+1} + V_2 \sum_{l=1}^{V} \hat{n}_l\hat{n}_{l+2}, \tag{136}$$

where $\hat{b}^\dagger_l$ ($\hat{b}_l$) creates (annihilates) a boson at site l and $\hat{n}_l = \hat{b}^\dagger_l\hat{b}_l$ is the site occupation operator. The operators $\hat{b}^\dagger_l$ and $\hat{b}_l$ satisfy the commutation relations $[\hat{b}_j, \hat{b}_k] = [\hat{b}^\dagger_j, \hat{b}^\dagger_k] = 0$ and $[\hat{b}_j, \hat{b}^\dagger_k] = \delta_{jk}$, supplemented by a hard-core constraint $(\hat{b}_l)^2 = (\hat{b}^\dagger_l)^2 = 0$ on physical states, which tells us that in physical states there can be at most one boson in a lattice site.

Implementing this constraint is subtle; see Refs. [189, 190]. We cannot assume the relation $(\hat{b}_l)^2 = (\hat{b}^\dagger_l)^2 = 0$ in an operator way as otherwise one finds that the algebra is zero. One needs to interpret this constraint as follows: $\langle\tilde{n}|(\hat{b}_l)^2|\tilde{m}\rangle = \langle\tilde{n}|(\hat{b}^\dagger_l)^2|\tilde{m}\rangle = 0$ for physical states $|\tilde{n}\rangle$ and $|\tilde{m}\rangle$, which are only given by the occupation numbers $\tilde{n} = (n_1, \ldots, n_V), \tilde{m} = (m_1, \ldots, m_V) \in \{0,1\}^V$. The crucial

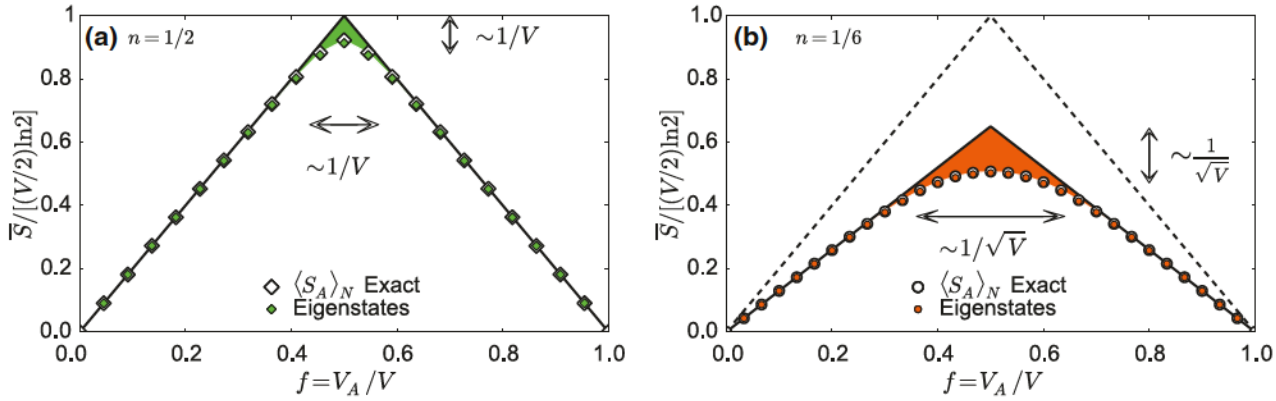


FIG. 13. Average entanglement entropy density $\bar{S}/[(V/2)\ln 2]$ (filled symbols) of the eigenstates of the quantum-chaotic interacting Hamiltonian in Eq. (136), which is particle-number conserving, as a function of the subsystem fraction $f = V_A/V$. The average is carried out over the central 20% of the eigenstates. Open symbols (overlapping with the filled ones) depict the corresponding exact result for general pure states, given by $\langle S_A\rangle_N$ from Eq. (45) for the same filling and system size, while the lines are the thermodynamic limit results from Eq. (54). The particle filling n and the number of lattice sites V are (a) $n = \frac{1}{2}$ and $V = 22$, and (b) $n = \frac{1}{6}$ and $V = 30$.

point is that virtual states are allowed to have more than one boson on a site. For instance, for the expectation value in a single site ($V = 1$), it holds that

$$\langle 0|\hat{b}^2(\hat{b}^\dagger)^2|0\rangle = \langle 0|2\hat{b}\hat{b}^\dagger|0\rangle = 2, \tag{137}$$

where we have exploited $[\hat{b}, (\hat{b}^\dagger)^2] = 2\hat{b}^\dagger$ and $\hat{b}\,|0\rangle = 0$. If one wants to replace the creation and annihilation operators by the regular spin operators in $\mathrm{sl}_{\mathbb{R}}(2)$, one first needs to normal order the operators via the standard bosonic commutation relations, meaning that the creation operators $\hat{b}_j^\dagger$ need to be placed to the left and the annihilation operators $\hat{b}_j$ to the right, and then replace them by spin operators. Mathematically, this means that each operator has to be dealt with in the infinite-dimensional Fock space, but at the end this space is projected onto the subspace with zero or one boson per site. This recipe naturally follows from how expectation values need to be calculated to describe the results of measurements in bosonic systems with very large on-site interactions [189,190].

Model (136) contains both nearest-neighbor and the next-nearest-neighbor hoppings and interactions, it is translationally invariant, and conserves the total number of hard-core bosons. When written in terms of $\frac{1}{2}$ spins, Hamiltonian (136) is known as the (extended) spin-$\frac{1}{2}$ XXZ chain. It has been of much interest to the condensed matter and mathematical physics communities, and been used to describe the behavior of solid-state materials [67].

Hamiltonian (136) is integrable when $t_2 = V_2 = 0$, independently of the values of t_1 and V_1. Then, it can be solved exactly using a Bethe ansatz [67]. When $t_2 \neq 0$ and/or $V_2 \neq 0$ (for $t_1 \neq 0$ and $V_1 \neq 0$), Hamiltonian (136) is nonintegrable. In the thermodynamic limit, one expects such a nonintegrable Hamiltonian to exhibit *many-body* quantum chaos, namely, one expects the distribution of the spacings of nearest-neighbor levels of the many-body energy spectrum to be described by the Wigner-Dyson statistics of random matrix theory [29,191–193]. This is why in the nonintegrable regime we refer to Hamiltonian (136) as a quantum-chaotic interacting Hamiltonian.

In finite (and relatively small, tens of sites) systems, i.e., those that can be solved using full exact diagonalization (as we do here), there is a crossover regime between integrability and quantum chaos as the magnitude of t_2 and/or V_2 depart from zero (for $t_1 \sim V_1 \neq 0$) [191–195]. In such a crossover regime, for small values of t_2 and/or V_2 relative to $t_1 \sim V_1$, the statistical properties of the many-body energy spectrum cannot be described by classical random matrix ensembles. When t_1, t_2, V_1, and V_2 are all similar in magnitude, full exact diagonalization calculations in Refs. [191,193] showed that Hamiltonian (136) exhibits many-body quantum chaos. To be in this quantum chaotic regime, in the calculations reported here we set $t_1 = t_2 = 1$ and $V_1 = V_2 = 1.1$.

The entanglement entropy of eigenstates of $\hat{H}_{\mathrm{HCB}}$ in Eq. (136) was calculated using full exact diagonalization after resolving all the symmetries of the model. The average entanglement entropy $\bar{S}$ was computed over the central 20% of the energy eigenstates (from all symmetry sectors; see Ref. [43] for details). Figure 13 shows the behavior of the average entanglement entropy density $\bar{S}/[(V/2)\ln 2]$ as a function of the subsystem fraction f. Two remarkable features of those numerical results are as follows. (i) They are nearly identical for the Hamiltonian eigenstates (filled symbols) and the result from Eq. (45). The small differences between them are quantified in Fig. 14. (ii) The deviations of $\bar{S}$ from the maximal entanglement entropy (shaded region) due to subleading terms may be substantial in finite systems, and they depend strongly on the particle

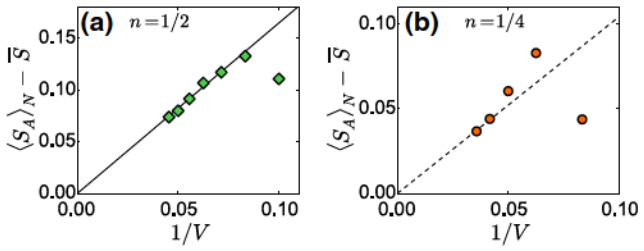


FIG. 14. Finite-size scaling of the average entanglement entropy difference $\langle S_A\rangle_N - \bar{S}$ versus $1/V$, at the subsystem fraction $f = \frac{1}{2}$. Here $\langle S_A\rangle_N$ is given by Eq. (45) and $\bar{S}$ denotes the numerical results for the eigenstates of the quantum-chaotic interacting Hamiltonian in Eq. (136), averaged over the central 20% of the energy spectrum. The particle fillings n are (a) $n = \frac{1}{2}$ and (b) $n = \frac{1}{4}$. The solid line in (a) shows a fit of a function a_1/V to the numerical results at $V \geq 14$, with $a_1 = 1.65$. The dashed line in (b) depicts $1/V$ behavior and is plotted as a guide to the eye.

filling n [compare the results for $n = \frac{1}{2}$ in Fig. 13(a) and for $n = \frac{1}{6}$ in Fig. 13(b)].

Figure 13 also mirrors the finite-size effects observed in the asymptotic expansion (54), illustrated in Fig. 6. The vertical (double arrow) difference at $f = \frac{1}{2}$ refers to the next-to-leading-order finite-size correction of the entanglement entropy density, which is $O(1/V)$ for $n = \frac{1}{2}$ and $O(1/\sqrt{V})$ for $n \neq \frac{1}{2}$. The horizontal (double arrow) difference refers to the spread of the Kronecker delta $\delta_{f,1/2}$ for finite systems, which is $O(1/V)$ for $n = \frac{1}{2}$ and $O(1/\sqrt{V})$ for $n \neq \frac{1}{2}$. We resolved them in Figs. 6(e) and 6(d), respectively.

Next, we resolve to which degree the average entanglement entropy $\bar{S}$ of Hamiltonian eigenstates agrees with the analytical predictions from Sec. II, in particular with $\langle S_A\rangle_N$ in Eq. (45). In Fig. 14, we plot the finite-size scaling of $\langle S_A\rangle_N - \bar{S}$. In all cases under consideration, the differences appear to vanish in the thermodynamic limit. This suggests that not only the volume-law contribution of Eq. (45), but also subleading terms including the $O(1)$ terms correctly predict the eigenstate entanglement entropies of quantum-chaotic interacting Hamiltonians. The differences between the latter and Eq. (45) appear to scale algebraically as $1/V^{\zeta}$, with $\zeta = O(1)$. The numerical results suggest that $\zeta = 1$ at $n = \frac{1}{2}$ [see the solid line in Fig. 14(a)], while they are not conclusive at $n = \frac{1}{4}$ [the dashed line in Fig. 14(b) depicts $1/V$ behavior and is plotted as a guide to the eye]. We note that, in Fig. 14, $\langle S_A\rangle_N - \bar{S} > 0$, implying that the asymptotic entanglement entropy is approached from below as the system size increases.

B. Quantum-chaotic quadratic model

Next, we focus on a quadratic model, namely, a model whose Hamiltonian is bilinear in fermionic creation and annihilation operators. We explore how well the results for fermionic Gaussian states from Sec. III predict the behavior of the entanglement entropy in eigenstates of a particle-number-conserving quadratic model that exhibits *single-particle* quantum chaos. By single-particle quantum chaos we mean that the statistical properties of the single-particle energy spectrum are described by the Wigner-Dyson statistics of random matrix theory. Hence, we refer to this model as a quantum-chaotic quadratic model [62]. This is to be contrasted to the model in Sec. V A, which exhibits *many-body* quantum chaos, and to which we referred to as a quantum-chaotic interacting model.

A well-known quadratic model that exhibits single-particle quantum chaos is the 3D Anderson model below the localization transition. The Hamiltonian of this model reads

$$\hat{H}_{\text{And}} = -t \sum_{\langle i,j\rangle} (\hat{f}_i^{\dagger}\hat{f}_j + \hat{f}_j^{\dagger}\hat{f}_i) + \frac{W}{2}\sum_i \varepsilon_i \hat{n}_i, \qquad (138)$$

where the first sum runs over nearest-neighbor sites on a cubic lattice. The operator $\hat{f}_j^{\dagger}$ ($\hat{f}_j$) creates (annihilates) a spinless fermion at site j, and $\hat{n}_j = \hat{f}_j^{\dagger}\hat{f}_j$ is the site occupation operator. The operators $\hat{f}_j^{\dagger}$ and $\hat{f}_j$ satisfy the standard anticommutation relations $\{\hat{f}_l, \hat{f}_k\} = \{\hat{f}_l^{\dagger}, \hat{f}_k^{\dagger}\} = 0$ and $\{\hat{f}_l, \hat{f}_k^{\dagger}\} = \delta_{lk}$. The single-site occupation energies $\varepsilon_i \in [-1, 1]$ are independently and identically distributed random numbers drawn from a box distribution. The 3D Anderson model exhibits a delocalization-localization transition at the critical disorder $W_c \approx 16.5$ (see, e.g., Refs. [196–199] for reviews). Our focus here is on disorder strengths well below this transition, $W \ll W_c$. We stress that, when referring to single-particle quantum chaos in the context of the 3D Anderson model (138), we have in mind the fixed Hilbert space $\mathcal{H}_1$ as the model of a single particle.

Even though it has been known for decades that the single-particle spectral properties of the 3D Anderson model in the delocalized regime are well described by the Wigner-Dyson statistics [200–202], the entanglement entropy of energy eigenstates was studied only recently [62]. The latter study showed that the volume-law contribution of typical many-body eigenstates is accurately described by the volume-law term of the asymptotic expression in Eq. (90) for $n = \frac{1}{2}$, which is the same as that in Eq. (118) for $\bar{n} = \frac{1}{2}$. This result suggests that the leading (volume-law) term in the eigenstate entanglement entropy of the 3D Anderson model deep in the delocalized regime is universal. In the main panel of Fig. 15, we plot the average eigenstate entanglement entropy density $\bar{S}/[(V/2)\ln 2]$ of randomly selected eigenstates as a function of the subsystem fraction f. The results show

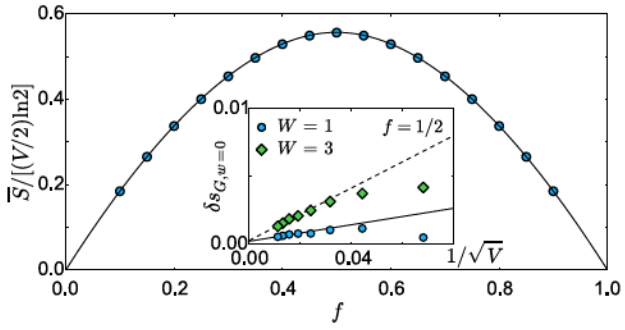


FIG. 15. Average entanglement entropy density $\bar{S}/[(V/2)\ln 2]$ in the 3D Anderson model (138) at $\bar{n} = \frac{1}{2}$. Main panel: plot of $\bar{S}/[(V/2)\ln 2]$ versus f at disorder strength $W = 1$, in a cubic lattice with $V = 8000$ sites (symbols). The results are obtained by averaging over 100 randomly selected many-body eigenstates and ten Hamiltonian realizations. The solid line is the corresponding thermodynamic limit result for fermionic Gaussian states given by $\langle S_A\rangle_{G,w=0}$ in Eq. (118). Inset: plot of $\delta s_{G,w=0} = (\langle S_A\rangle_{G,w=0} - \bar{S})/[(V/2)\ln 2]$ versus $1/\sqrt{V}$ at $f = \frac{1}{2}$, for $W = 1$ and 3, where $\langle S_A\rangle_{G,w=0}$ corresponds to the fermionic Gaussian states [Eq. (117)] at $w = 0$ and the same V as $\bar{S}$. The results for $\bar{S}$ are obtained by averaging over 10^2 to 10^4 randomly selected many-body eigenstates and over 5 to 500 Hamiltonian realizations. Lines are linear fits $a_0 + a_1/\sqrt{V}$ to the results for $V \geq 2000$. We get $a_0 = 2.4 \times 10^{-4}$ and $a_1 = 0.03$ for $W = 1$ (solid line), and $a_0 = 3.0 \times 10^{-4}$ and $a_1 = 0.10$ for $W = 3$ (dashed line). The numerical results for $\bar{S}$ are from Ref. [62].

remarkable agreement with the corresponding thermodynamic limit expression for the weighted average entanglement entropy over fermionic Gaussian states $\langle S_A\rangle_{G,w=0}$ in Eq. (118).

In spite of the latter agreement, we note that the average entanglement entropy over fermionic Gaussian states does not describe the first subleading term of the average entanglement entropy in the 3D Anderson model. As shown in the inset of Fig. 15, the first subleading term in the latter model scales $\propto \sqrt{V}$ at $f = \frac{1}{2}$. No such term appears in $\langle S_A\rangle_{G,w=0}$ in Eq. (118). The fact that, for the 3D Anderson model, the subleading $O(\sqrt{V})$ term is not described by Eq. (118) is in stark contrast to what we found in Sec. V A for a quantum-chaotic *interacting* model. In the latter case, subleading terms that are $O(1)$ or greater in the physical model are properly described by the average $\langle S_A\rangle_N$ in Eq. (45). Hence, the origin of the $O(\sqrt{V})$ contribution to the entanglement entropy of eigenstates in the 3D Anderson model remains an open question. Such a contribution is not present in our analytical calculations of the averages over Gaussian states.

C. Translationally invariant noninteracting fermions

Next, we consider a paradigmatic quadratic model that does not exhibit quantum chaos at the single-particle level. Namely, translationally invariant noninteracting fermions, for which the Hamiltonian is a sum of hopping terms over nearest-neighbor sites [the first term in Eq. (138)]. For simplicity, we focus on the 1D case

$$\hat{H}_T^{1D} = -\sum_{i=1}^{V} (\hat{f}_i^\dagger \hat{f}_{i+1} + \hat{f}_{i+1}^\dagger \hat{f}_i) \tag{139}$$

with periodic boundary conditions, $\hat{f}_{V+1} \equiv \hat{f}_1$. The single-particle eigenenergies of the model in Eq. (139) are given by the well-known expression $\epsilon_n = -2\cos(2\pi n/V)$ with $n = 0, 1, \ldots, V-1$, which makes apparent that the statistical properties of the single-particle spectrum are not described by the Wigner-Dyson statistics.

The average eigenstate entanglement entropy of the model in Eq. (139) was studied in Ref. [42] (before the universal predictions for the quantum-chaotic quadratic models and the fermionic Gaussian states were derived). The numerical calculations in Ref. [42] were carried out by averaging the entanglement entropy over the full set of 2^V many-body eigenstates. Remarkably, the numerical results were found to converge rapidly to the thermodynamic limit result, as shown for the case of $f = \frac{1}{2}$ in the inset of Fig. 16. Thanks to that scaling, we find the volume-law coefficient s_T^∞ of the average entanglement entropy $\bar{S}_T = s_T^\infty V_A \ln 2$ at $f = \frac{1}{2}$ to high numerical accuracy, $s_T^\infty = 0.5378(1)$, which is consistent with the result reported in Ref. [42]. This is to be contrasted to the volume-law coefficient $s_{G,w=0}^\infty$ of fermionic Gaussian states $\langle S_A\rangle_{G,w=0} = s_{G,w=0}^\infty V_A \ln 2$ from Eq. (118), which yields $s_{G,w=0}^\infty = 0.5573$. We then see that s_T^∞ and $s_{G,w=0}^\infty$ are close

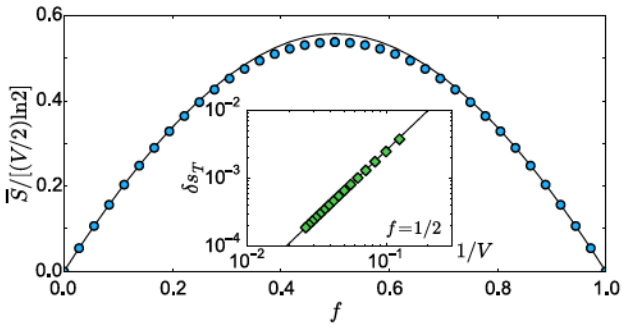


FIG. 16. Average entanglement entropy density $\bar{S}/[(V/2)\ln 2]$ of translationally invariant noninteracting fermions in a one-dimensional lattice, described by the Hamiltonian in Eq. (139). Main panel: plot of $\bar{S}/[(V/2)\ln 2]$ versus f in the lattice with $V = 36$ sites. The results are obtained by averaging over all 2^V many-body eigenstates. The solid line is the corresponding thermodynamic limit result for fermionic Gaussian states given by $\langle S_A\rangle_{G,w=0}$ in Eq. (118). Inset: plot of $\delta s_T = (\bar{S}_T - \bar{S})/([V/2]\ln 2)$ versus $1/V$ at $f = \frac{1}{2}$, where $\bar{S}_T/([V/2]\ln 2) = 0.5378$. The solid line shows the function a/V^ζ, with $a = 0.23$ and $\zeta = 1.96$. The numerical results for $\bar{S}$ are from Ref. [42].

TABLE I. Overview of the results discussed in this tutorial. We list the main results, indicate up to which order in V we derived the respective expressions (and if there exists an exact formula), and where the respective formulas can be found (equations, figures, references). Most results for fixed particle number are new, but if special cases or the leading-order term were already known before, we cite the relevant works after the equation in the main text.

	(a) General pure states		(b) Pure fermionic Gaussian states	
(1) No particle number	$\langle S_A\rangle = aV - b + O(2^{-V})$ **& exact**	→ (25), Fig. 3, [152]	$\langle S_A\rangle_G = aV + b + O(1/V)$ **& exact**	→ (89), Fig. 8, [159]
	$(\Delta S_A)^2 = \alpha e^{-\beta V} + o(e^{-\beta V})$	→ (29), [165]	$(\Delta S_A)^2_G = a + o(1)$	→ (93), [159]
(2) Fixed particle number	$\langle S_A\rangle_N = aV - b\sqrt{V} - c + o(1)$	→ (54), Fig. 6	$\langle S_A\rangle_{G,N} = aV - b/V + O(1/V^2)$ **& exact**	→ (90)
	$(\Delta S_A)^2_N = \alpha V^{3/2} e^{-\beta V}$	→ (60)	$(\Delta S_A)^2_{G,N} = a + o(1)$	→ (116), Fig. 9
(3) Fixed weight	$\langle S_A\rangle_w = aV + b + c\sqrt{V} + o(1)$	→ (67)	$\langle S_A\rangle_{G,w} = aV + b + c/\sqrt{V} + d/V + o(1/V)$	→ (118), Fig. 10
	$(\Delta S_A)^2_w = aV + o(V)$	→ (68)	$(\Delta S_A)^2_{G,w} = aV + o(V)$	→ (120)

but different. The full curve for S_T as a function of f, for $V = 36$, is shown in Fig. 16 together with the full curve for $\langle S_A\rangle_{G,w=0}$ from Eq. (118). They are clearly different and, given the abovementioned fast convergence of the numerical results with V, we expect the differences to remain in the thermodynamic limit. The exact analytical form of the $\bar{S}_T(f)$ curve for translationally invariant free fermions remains elusive, but tight bounds have already been calculated [50].

We conclude by noting that, for the translationally invariant quantum-chaotic interacting model studied in Sec. V A, the average eigenstate entanglement entropy is accurately described by the corresponding entanglement entropy of general pure states. The role of Hamiltonian symmetries in the average entanglement entropy of energy eigenstates in quantum-chaotic interacting and quantum-chaotic quadratic models remains an important question to be explored in future studies.

VI. SUMMARY AND OUTLOOK

In this section, we briefly summarize the key results discussed in this tutorial, and give an outlook of where we envision the methods introduced to be applicable. We also mention some open questions in the context of the entanglement entropy of typical pure states.

A. Summary

We provided a pedagogical introduction to the current understanding of the behavior of the entanglement entropy of pure quantum states. We derived analytical expressions for the average entanglement entropy of general and Gaussian states, and considered states with and without a fixed number of particles. A comprehensive summary of the results discussed can be found in Table I, where we contrast results for: (1) arbitrary particle number, (2) fixed particle number N, and (3) fixed weight parameter w for both (a) general pure states and (b) Gaussian states. This yields the six state ensembles (1a) through (3b).

For both Gaussian and general pure states, the leading-order behavior $\langle S_A\rangle$ at half-filling $N = V/2$ coincides with the full average without fixing the total particle number, while the next-to-leading-order terms differ. For general pure states, we confirmed an additional contribution proportional to $\sqrt{V}$ at $f = \frac{1}{2}$ in Eq. (54), previously found in Ref. [43]. For Gaussian states, we derived the exact formula, which does not contain such a term and has a next-to-leading-order term of order $1/V$ [Eq. (114)]. However, we did find a contribution of order $1/\sqrt{V}$ in the asymptotic average $\langle S_A\rangle_{G,w}$ at fixed w with $f = \bar{n}$, i.e., whenever the subsystem fraction f equals the average filling ratio $\bar{n} = \langle N/V\rangle = 1/(1 + e^w)$.

We traced back these contributions to the nonanalytic behavior of the average entanglement entropy as a function of the subsystem fraction f and the filling ratio n. In the case of Gaussian states, we identified the additional particle-subsystem symmetry $n \leftrightarrow f$, which is responsible for the $1/\sqrt{V}$ term. From a mathematical perspective, the origin of the $\sqrt{V}$ term in $\langle S_A\rangle_N$ is therefore the same as that of the $1/\sqrt{V}$ term in $\langle S_A\rangle_{G,w}$, namely, both calculations involve the average of a nonanalytic function with respect to an approximately Gaussian statistical distribution. Square root powers of V appear exactly when the mean of the Gaussian lies in a neighborhood of the nonanalyticity, i.e., there is a jump in one of the function's derivatives.

Finally, we connected the results obtained for the average entanglement entropy in the six ensembles of states mentioned before to the average entanglement entropy in eigenstates of specific random matrices and of physical Hamiltonians. Maybe the most surprising result in the context of quantum-chaotic interacting Hamiltonians is that not only does the leading term in the average agree with the corresponding ensemble average, but also subleading terms that are $O(1)$ or larger in the volume, e.g., $O(\sqrt{V})$. Why this is so is a question that deserves to be

further explored. Equally intriguing is to understand why the same is not true in the case of quantum-chaotic quadratic Hamiltonians.

B. Outlook

Looking forward, an important question is how general are the methods and results discussed here. We focused on fermionic systems, for which we can compare general pure states with Gaussian pure states, and unveiled the effect of fixing the total particle number. Our results for general pure states apply equally to hard-core bosons and spin-$\frac{1}{2}$ systems. In the latter, the total magnetization plays the role that the total particle number plays in fermionic and hard-core boson models.

1. Typical eigenstate entanglement entropy as a diagnostic of quantum chaos and integrability

As mentioned in the Introduction, a novel picture that the recent numerical studies such as those discussed in Sec. V have started to consolidate is that typical many-body eigenstates of quantum-chaotic interacting Hamiltonians have similar entanglement properties as typical pure states in the Hilbert space. In parallel, typical many-body eigenstates of quantum-chaotic quadratic Hamiltonians have similar entanglement properties as typical Gaussian pure states. We quantified how similar they are by showing that typical eigenstates of a specific quantum-chaotic interacting Hamiltonian exhibit $O(1)$ and greater terms in the entanglement entropy that are the same than in typical pure states in the Hilbert space. For typical many-body eigenstates of quantum-chaotic quadratic Hamiltonians, we showed that the $O(V_A)$ term is the same as in typical Gaussian pure states. These statements (for $V_A = fV \leq V/2$) are true independently of whether one deals with states in which the number of particles is fixed or not.

In the context of Hamiltonians that do not exhibit many-body quantum chaos, namely, in which the many-body level spacing distributions are not described by the Wigner surmise [29], we showed that typical many-body energy eigenstates of translationally invariant non-interacting fermions exhibit an $O(V_A)$ term that behaves qualitatively similar (but is not equal) to that obtained for typical Gaussian pure states, namely, the prefactor of such a term is a function of the subsystem fraction f. The same behavior was found in Ref. [55] for the typical entanglement entropy of many-body eigenstates of the integrable spin-$\frac{1}{2}$ XXZ chain. This is fundamentally different from what happens in typical many-body eigenstates of quantum-chaotic interacting Hamiltonians, in which the prefactor is maximal (it depends only on the filling n) as in typical pure states.

Hence, as conjectured in Ref. [55], the entanglement entropy of typical many-body energy eigenstates can be used to distinguish models that exhibit many-body quantum chaos (whose level spacing distributions are described by the Wigner surmise, and are expected to thermalize when taken far from equilibrium [29]) from those that do not. This is a welcome addition to the toolbox for identifying quantum chaos as it relies on the properties of the eigenstates as opposed to the properties of the eigenenergies. Other entanglement-based diagnostics of quantum chaos and integrability have been proposed in recent years, among them are the operator entanglement growth [203–205]; the diagonal entropy [206,207], the mutual information scrambling [204], and entanglement revivals [208] after quantum quenches; the tripartite operator mutual information [209,210]; and the entanglement negativity between two subsystems in a tripartition of many-body energy eigenstates [211].

It is important to emphasize that an advantage of using the entanglement properties of energy eigenstates, instead of the properties of the eigenenergies, is that one does not need to resolve all the symmetries of the model nor does one need to do an unfolding of the spectrum, which are of paramount importance to identify quantum chaos using the eigenenergies, as discussed in Sec. I D. In addition, in comparison to some of the entanglement diagnostics that were mentioned above, one does not need to study dynamics. Further works are needed on interacting integrable models to establish whether the leading term of the entanglement entropy of typical many-body energy eigenstates is universal or not, and to understand the nature of the subleading terms. So far, results are available only for the integrable spin-$\frac{1}{2}$ XXZ chain [55].

2. Beyond qubit-based systems

The analytical tools introduced and explained in this tutorial can be used beyond the fermionic systems we studied (and beyond the spin-$\frac{1}{2}$ and hard-core boson systems we mentioned), and facilitate the study of bosonic systems with a fixed particle number. To be concrete, a bosonic subsystem with V_A out of V bosonic modes and total particle number N can be treated analogously to Eq. (45), but with dimensions respecting the bosonic commutation statistics, i.e.,

$$d_A(N_A) = \frac{(N_A + V_A - 1)!}{N_A!\,(V_A - 1)!}, \tag{140}$$

$$d_B(N - N_A) = \frac{(N - N_A + V - V_A - 1)}{(N - N_A)!\,(V - V_A - 1)!}, \tag{141}$$

$$d_N = \frac{(N + V - 1)!}{N!\,(V - 1)!}, \tag{142}$$

which follows from the combinatorics of sampling with replacement without caring about the order, e.g., for d_A, we ask how many ways there are to distribute N_A indistinguishable particles over V_A sites (where each site can hold arbitrarily many particles). Anew, it holds that $\sum_{N_A=0}^{N} d_A(N_A) d_B(N-N_A) = d_N$.

Following Page's approach, we again choose an arbitrary uniformly distributed random vector state in the Hilbert space $\mathcal{H}_N$. Thus, the invariance of the state under the unitary group $U(d_N)$, now with a different dimension d_N, still applies. Therefore, we can follow the same strategy as in Sec. II B, in particular we can exploit Eq. (45) with dimensions (140). This yields, in the thermodynamic limit with fixed $f \in (0, \frac{1}{2})$ and $n \in (0, \infty)$,

$$\begin{aligned}\langle S_A \rangle_{\mathrm{bos},N} = {} & fV[n \ln(1+n^{-1}) + \ln(1+n)] \\ & + \sqrt{V}\sqrt{\frac{n+n^2}{8\pi}} \ln(1+n^{-1}) \delta_{f,1/2} \\ & + \frac{f + \ln(1-f)}{2} + o(1), \end{aligned} \tag{143}$$

$$\langle S_A \rangle_{\mathrm{bos},w} = \langle S_A \rangle_{\mathrm{bos},N=\bar{n}V} - \frac{f}{2} + o(1), \tag{144}$$

where the weighted average is only meaningful for $w > 0$, for which $\bar{n} = 1/(e^w - 1)$. Note that there is no particle-hole symmetry for bosons, and that $n = N/V$ can be arbitrarily large.

Other natural generalizations are spin-s systems with $s > \frac{1}{2}$ and systems consisting of distinguishable particles. These cases can also be studied using the methods discussed in this tutorial after carrying out the respective combinatorics of the Hilbert space dimensions d_A and d_B. Also, systems with global symmetries such as time-reversal invariance or chirality can be considered, which have an impact on the respective symmetry group so that the Hilbert space is not invariant anymore under $U(d_N)$ but only under $O(d_N)$ or $U(d_{N_1}) \times U(d_{N_2})$. The leading terms are expected to be the same, as the respective random matrix ensembles share the same level densities. Deviations are expected to occur in subleading terms.

3. Other ensembles and entanglement measures

We focused on ensembles of states, general and Gaussian pure states for arbitrary and fixed particle numbers, which mirror the entanglement properties of typical ("infinite-temperature") eigenstates of physical lattice models. It is also possible to construct ensembles of pure states in which one fixes the energy, which mirror the entanglement properties of "finite-temperature" eigenstates of physical lattice models. Steps in this direction have already been taken using different tools; see, e.g., Refs. [33,47,51,52,157,213]. In the context the scaling of the eigenstate entanglement entropy at different energy densities ("temperatures"), let us also emphasize that all the average entanglement entropies computed in this tutorial exhibited a leading volume-law term, namely, the leading term in the average entropies scales with the number of modes V and is thus agnostic to the individual shape or area of the subsystem. In contrast, as discussed in the Introduction, it is well known that low-energy states of many physical systems of interest exhibit a leading area law term. An important open question is whether one can define ensembles of pure states that exhibit leading terms in the entanglement entropy that are area law.

Instead of considering the von Neumann entanglement entropy, one can also consider other quantities that are defined with respect to the invariant spectrum of the reduced density operator $\hat{\rho}_A = \mathrm{Tr}_{\mathcal{H}_B} |\psi\rangle\langle\psi|$ of a pure state $|\psi\rangle$. Such quantities include the well-known Renyi entropies $S_A^{(n)}(|\psi\rangle)$, and the eigenstate capacity [214]. We focused on the von Neumann entropy, as it is arguably the most prominent measure of bipartite entanglement. Nonetheless, we expect that our findings can also be extended to the aforementioned quantities; see, e.g., Refs. [62,86,88,215] for studies of Renyi entropies and Refs. [178,216] for studies of the eigenstate capacity.

It would also be interesting to explore multipartite entanglement measures for different ensembles of pure states. This will likely require new techniques, and it is not clear what the most suitable measure is. The latter question is the subject of ongoing research.

ACKNOWLEDGMENTS

We would like to thank Pietro Donà, Peter Forrester, Patrycja Łydżba, Lorenzo Piroli, and Nicholas Witte for inspiring discussions. E.B. acknowledges support from the National Science Foundation, Grant No. PHY-1806428, and from the John Templeton Foundation via the ID 61466 grant, as part of the "Quantum Information Structure of Spacetime (QISS)" project (qiss.fr). L.H. gratefully acknowledges support from the Alexander von Humboldt Foundation. M.K. acknowledges support from the Australian Research Council (ARC) under Grant No. DP210102887. M.R. acknowledges support from the National Science Foundation under Grant No. 2012145. L.V. acknowledges support from the Slovenian Research Agency (ARRS), Research core fundings Grants No. P1-0044 and No. J1-1696. L.H. and M.K. are also grateful to the MATRIX Institute in Creswick for hosting the online research programme and workshop "Structured Random Matrices Downunder" (26 July–13 August 2021).

APPENDIX A: GENERAL PURE STATES WITH A FIXED NUMBER OF PARTICLES

1. Average entanglement entropy

In sum (45) we have different contributions. Since our goal is to asymptotically expand the sum as a whole up to $O(1)$, which includes three different orders, namely V, $\sqrt{V}$, and 1, we need to approach it systematically. The first step is to identify the various contributions of such an expansion. Let us list them as follows.

1. The difference of the digamma functions in Eq. (45) can always be asymptotically expanded because the dimensions d_N, $d_B(N-N_A)$, and $d_A(N_A)$ grow with V. The only exceptions are when either the volumes of subsystem A are $V_A = 0, V$ or the occupation numbers are $N = 0, V$. These situations correspond to the trivial cases that we do not split the system into two subsystems or that all sites are empty or occupied, which we always exclude. When assuming $n = N/V$ and $f = V_A/V$ being of order one and having a finite distance from 1, the dimensions grow exponentially in V.
   The dimension does not grow or only grows polynomially in V when $N_A \approx 0, V_A$ for $d_A(N_A)$ or $N - N_A = 0, V - V_A$ for $d_B(N-N_A)$. However, the prefactor $\varrho_{N_A} = d_A(N_A)d_B(N-N_A)/d_N$ will be exponentially small for these cases, as we will see below.

Thus, we can expand the difference of the digamma functions as

$$\begin{aligned}&\Psi[d_N+1]-\Psi[d_B(N-N_A)+1]\\&\quad=\ln\left(\frac{d_N}{d_B(N-N_A)}\right)+O(e^{-\gamma_1 V}),\\&\Psi[d_N+1]-\Psi[d_A(N_A)+1]\\&\quad=\ln\left(\frac{d_N}{d_A(N_A)}\right)+O(e^{-\gamma_2 V}),\end{aligned}\tag{A1}$$

where $\gamma_1, \gamma_2 > 0$ are two nonvanishing rates that only state how fast the exponential correction of this approximation vanishes. These logarithms of the ratios are expanded in Appendix A 1 a.

2. A second contribution results from N_{crit}, which determines the fraction of the contribution of the two sums. It is defined by the largest positive integer such that it still holds that $d_A(N_A) \le d_B(N-N_A)$. We need a $1/V$ expansion of this integer as slight deviations may yield order-one terms. We derive such an expansion in Appendix A 1 b.
3. Once we know the $1/V$ expansion of N_{crit}, we can combine it with approximation (A1) and consider the first part of the sum, which is

$$\begin{aligned}\Sigma_1 &= \sum_{N_A=0}^{N_{\text{crit}}}\frac{d_A(N_A)d_B(N-N_A)}{d_N}(\Psi[d_N+1]-\Psi[d_B(N-N_A)+1])\\&\quad+\sum_{N_A=N_{\text{crit}}+1}^{N}\frac{d_A(N_A)d_B(N-N_A)}{d_N}(\Psi[d_N+1]-\Psi[d_A(N_A)+1])\\&=\sum_{N_A=0}^{N_{\text{crit}}}\frac{d_A(N_A)d_B(N-N_A)}{d_N}\ln\left(\frac{d_N}{d_B(N-N_A)}\right)+\sum_{N_A=N_{\text{crit}}+1}^{N}\frac{d_A(N_A)d_B(N-N_A)}{d_N}\ln\left(\frac{d_N}{d_A(N_A)}\right)+O(e^{-\tilde{\gamma}V})\end{aligned}\tag{A2}$$

with a fixed $\tilde{\gamma} > 0$. We note that we can extend the second sum to the upper terminal N as the binomial factor is equal to 0 whenever $N_A > V_A$.
What is still needed is to expand the dimensions. The factor $\varrho_{N_A} = d_A(N_A)d_B(N-N_A)/d_N$ is a crucial ingredient for this purpose, as already pointed out. For large V, it will have approximately a Gaussian shape in N_A with a center $\bar{n}_A V$ and a width σ_A. The problematic point is that the two sums create a kink at N_{crit}. Thence, one needs to make a case discussion when $|N_{\text{crit}} - \bar{n}_A V|$ is maximally of order σV or not. We have already pointed out above Eq. (51) that $\bar{n}_A = nf$ and $\sigma\sqrt{f(1-f)n(1-n)}/\sqrt{V}$.

The logarithm of the ratio of the dimensions will be Taylor expanded in $(N_A/V - nf)$ and will be of order $1/\sqrt{V}$. Therefore, we only need to understand the asymptotic expansion of the moments

$$\begin{aligned}\mathcal{M}_j &= c_-\sum_{N_A=0}^{N_{\text{crit}}}\frac{d_A(N_A)d_B(N-N_A)}{d_N}\left(\frac{N_A}{V}-nf\right)^j\\&\quad+c_+\sum_{N_A=N_{\text{crit}}+1}^{N}\frac{d_A(N_A)d_B(N-N_A)}{d_N}\\&\quad\times\left(\frac{N_A}{V}-nf\right)^j,\end{aligned}\tag{A3}$$

where c_- and c_+ are some N_A-independent coefficients, which in principle depend on the order j of the moment, though.We show in Appendix A 1 b that $\bar{n}_A V$ is never larger than $N_{\rm crit}$ so that the maximum of ϱ_{N_A} is always in the first sum while the second sum only contains an exponential tail; see Fig. 17. Thus, it is suitable to consider the splitting

$$\begin{aligned}\mathcal{M}_j &= c_- \sum_{N_A=0}^{N} \frac{d_A(N_A) d_B(N-N_A)}{d_N} \left(\frac{N_A}{V} - nf\right)^j \\ &\quad + (c_+ - c_-)\Delta\mathcal{M}_j, \\ \Delta\mathcal{M}_j &= \sum_{N_A=N_{\rm crit}+1}^{N} \frac{d_A(N_A) d_B(N-N_A)}{d_N} \left(\frac{N_A}{V} - nf\right)^j .\end{aligned} \tag{A4}$$

The advantage is that the first sum with N_A running from 0 to N can be seen as a contribution that is always present. Only when $|N_{\rm crit} - \bar{n}_A V|$ is at most of order $\sigma V = \sqrt{f(1-f)n(1-n)V}$ will the second sum $\Delta\mathcal{M}_j$ be of importance. We carry out the expansion of M_j in Appendix A 1 c.

4. The remaining part of sum (45) is

$$\begin{aligned}\Sigma_2 &= \sum_{N_A=0}^{N_{\rm crit}} \frac{d_A(N_A) d_B(N-N_A)}{d_N} \frac{d_A(N_A)-1}{2d_B(N-N_A)} \\ &\quad + \sum_{N_A=N_{\rm crit}+1}^{N} \\ &\quad \times \frac{d_A(N_A) d_B(N-N_A)}{d_N} \frac{d_B(N-N_A)-1}{2d_A(N_A)}.\end{aligned} \tag{A5}$$

As mentioned, the dimensions $d_A(N_A)$ and $d_B(N-N_A)$ usually grow exponentially in V, and where they do not grow like that they will be suppressed by the exponential tails of ϱ_{N_A}. Thus, we can omit the terms $1/[2d_B(N-N_A)]$ and $1/[2d_A(N_A)]$, resulting in an exponentially suppressed correction. Secondly, the two sums can be combined into the form

$$\begin{aligned}\Sigma_2 &= \frac{1}{2} \sum_{N_A=0}^{\min(N,V_A)} \frac{d_A(N_A) d_B(N-N_A)}{d_N} \\ &\quad \times \min\left\{\frac{d_A(N_A)}{d_B(N-N_A)}, \frac{d_B(N-N_A)}{d_A(N_A)}\right\} \\ &\quad + O(e^{-\tilde{\gamma}V})\end{aligned} \tag{A6}$$

with another fixed $\tilde{\gamma} > 0$.

One may ask why we do not deal with this sum in the same way as we have done for the difference of the digamma functions. First, we cannot extend the upper terminal of the sum to N as the ratio $d_A(N_A)/d_B(N-N_A)$ can cancel the zero of the weight ϱ_{N_A}. Second, the ratio $d_A(N_A)/d_B(N-N_A)$ can exponentially grow as well as shrink in V which, in principle, can shift the maximum of the weight ϱ_{N_A}. We analyze this behavior in Appendix A 1 d.

Once all these contributions have been analyzed and expanded in $1/\sqrt{V}$, we combine the intermediate results in Appendix A 1 e.

a. Asymptotic expansion of the logarithm of the ratio of dimensions

As we have seen in approximation (A1) of the differences of the digamma functions, it is suitable to expand the logarithm of the ratio

$$\begin{aligned}\ln\left(\frac{d_N}{d_A(N_A)}\right) &= \ln\left(\frac{V!\,N_A!\,(V_A-N_A)!}{N!\,(V-N)!\,V_A!}\right) \\ &= \ln\left(\frac{V!\,(n_A V)!\,([f-n_A]V)!}{(nV)!\,([1-n]V)!\,(fV)!}\right).\end{aligned} \tag{A7}$$

The approximation of the other logarithm follows from the symmetry relation $(N_A, V_A) \to (N-N_A, V-V_A)$ or, equivalently, $(n_A, f) \to (n-n_A, 1-f)$.

We carry out the series expansion of Eq. (A7) in two steps. First, we take the asymptotic expansion in V. For this purpose, we take into account the facts that f, $1-f$, n, $1-n > 0$ are of order one so that one can use of the Stirling formula (A8) for four of the six factorials, namely $V!/(nV)!\,([1-n]V)!\,(fV)!$. For the other two terms $(n_A V)!\,([f-n_A]V)!$, we exploit the knowledge that the maximal contribution of the binomial distribution is given at $\bar{n}_A = nf$. Hence, it also follows that the argument in the remaining factorials grows linearly with V. Therefore, we exploit the Stirling approximation

$$\begin{aligned}(kV)! &= \sqrt{2\pi kV}(kV+1)^{kV} e^{-kV-1}[1+O(V^{-1})] \\ &= \sqrt{2\pi kV}(kV)^{kV} e^{-kV}[1+O(V^{-1})]'\end{aligned} \tag{A8}$$

with k being of order one, for any of the six factorials. Collecting these approximations, the large-V expansion yields

$$\begin{aligned}\ln\left(\frac{d_N}{d_A(N_A)}\right) &= V[n_A \ln(n_A) \\ &\quad + (f-n_A)\ln(f-n_A) - n\ln(n) \\ &\quad - (1-n)\ln(1-n) - f\ln(f)] \\ &\quad + \frac{1}{2}\ln\left(\frac{n_A(f-n_A)}{n(1-n)f}\right) + O(V^{-1}).\end{aligned} \tag{A9}$$

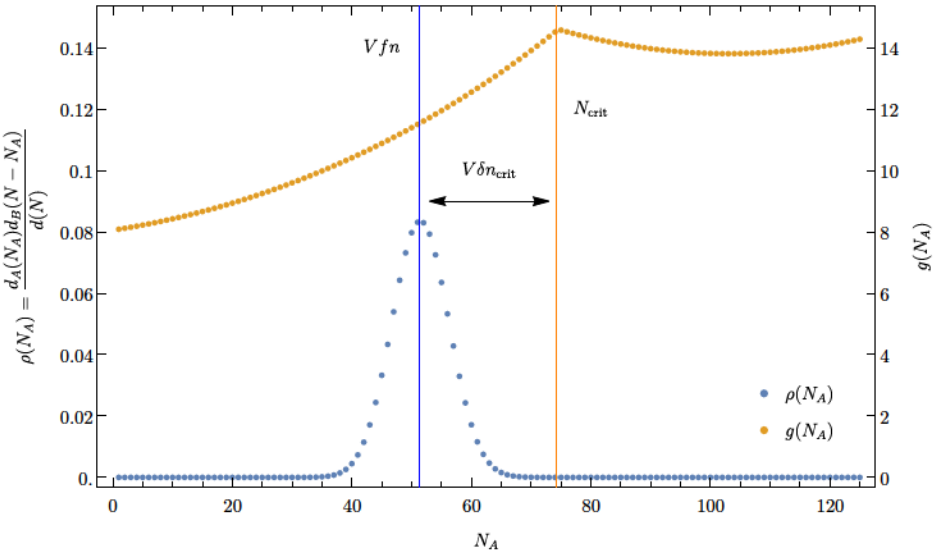


FIG. 17. Illustration of the contributions of sum (45) for $V = 500, f = \frac{1}{2}$, and $n = \frac{1}{4} - 4/\sqrt{L}$, where we plot the density $\varrho_{N_A} = d_A(N_A)d_B(N-N_A)/d_N$ (blue circles) and the approximate observable $g(N_A)$ (yellow circles), which is given by $\ln d_N/d_A(N_A)$ for $d_A(N_A) \leq d_B(N-N_A)$ and $\ln[d_N/d_B(N-N_A)]$ otherwise. Note that we use different scales for ϱ_{N_A} and $g(N_A)$. As reference points we have added the center of the Gaussian approximation of ϱ_{N_A} (blue vertical line) and the point N_{crit} of the kink of the summands (yellow vertical line).

In the second step, we replace $n_A = nf + \delta n_A$ with δn_A of order $1/\sqrt{V}$ to obtain

$$\ln\left(\frac{d_N}{d_A(N_A)}\right) = V(f-1)[n\ln(n) + (1-n)\ln(1-n)] + V\delta n_A \ln\left(\frac{n}{1-n}\right) + \frac{V\delta n_A^2}{2n(1-n)f} + \frac{1}{2}\ln(f) + O(V^{-1/2}). \tag{A10}$$

As aforementioned, the asymptotic for the second logarithm $\ln(d_N/d_B(N-N_A))$ can also be obtained using Eq. (A10) when exploiting the symmetry $(n_A, f) \leftrightarrow (n - n_A, 1-f)$. We can indeed apply this symmetry to Eq. (A10) as the weight ϱ_{N_A} also shares this symmetry. Here, we note that the maximum of the weight $\bar{n}_A = nf$ indeed implies that $n - \bar{n}_A = n(1-f)$, reflecting the symmetry. Thence, we have

$$\ln\left(\frac{d_N}{d_B(N-N_A)}\right) = -Vf\,[n\ln(n) + (1-n)\ln(1-n)] + V\delta n_A \ln\left(\frac{1-n}{n}\right) + \frac{V\delta n_A^2}{2n(1-n)(1-f)} + \frac{1}{2}\ln(1-f) + O(V^{-1/2}). \tag{A11}$$

For computational purposes in the ensuing section, we need to take the difference of Eqs. (A10) and (A11); recall the third step of the outline of computing the asymptotic expansion. In particular, when $1-2f$ is much smaller than one, this is important because the expectation value of the two terms tells us how many terms we need to take into account in the computation. The difference is

$$\ln\left(\frac{d_B(N-N_A)}{d_A(N_A)}\right) = -V(1-2f)[n\ln(n) + (1-n)\ln(1-n)] + 2V\delta n_A \ln\left(\frac{n}{1-n}\right) + o(1), \tag{A12}$$

where we have set $1-2f \ll 1$. Hence, this approximation is only true in this case. For the other case that f is not close enough to $\frac{1}{2}$, we do not need the respective expression, as we will see.

b. Computation of N_{crit}

As already mentioned, N_{crit} is given by the condition $d_A(N_A)/d_B(N-N_A) = 1$. This ratio can be approximated via Stirling's formula (A8),

$$\frac{d_A(N_A)}{d_B(N-N_A)} = \frac{V_A!\,(N-N_A)!\,(V-V_A-N+N_A)!}{N_A!\,(V_A-N_A)!\,(V-V_A)!}$$
$$= \frac{(fV)!\,([n-n_A]V)!\,([1-f\ -n+n_A]V)!}{(n_AV)!\,([f\ -n_A]V)!\,([1-f]V)!}$$
$$= \sqrt{\frac{f\,(n-n_A)(1-f\ -n+n_A)}{n_A(f\ -n_A)(1-f)}}\left[\frac{f^f\,(n-n_A)^{n-n_A}(1-f\ -n+n_A)^{1-f\ -n+n_A}}{n_A^{n_A}(f\ -n_A)^{f\ -n_A}(1-f)^{1-f}}\right]^V[1+O(V^{-1})]. \quad \text{(A13)}$$

As in the previous subsection, we need to go even further and take into account the facts that n_A is concentrated at $\bar{n}_A = nf$ and the small deviation $\delta n_A = n_A - \bar{n}_A$ is of order $1/\sqrt{V}$. Thus, the Taylor expansion in δn_A about this point leads to

$$\frac{d_A(N_A)}{d_B(N-N_A)} = \sqrt{\frac{1-f}{f}}[n^n(1-n)^{1-n}]^{V(1-2f)}\exp\left[2V\delta n_A\ln\left(\frac{1-n}{n}\right)+\frac{V(2f\ -1)\delta n_A^2}{2nf\,(1-f\,)(1-n)}\right][1+O(V^{-1/2})]. \quad \text{(A14)}$$

We would like to recall that we choose $0 < f \leq \frac{1}{2}$ due to the symmetries of the setting. Thence, the leading prefactor shows that the ratio vanishes exponentially when $1-2f$ is larger than order $1/V$ because of $\frac{1}{2} \leq n^n(1-n)^{(1-n)} < 1$ for any $0 < n \leq \frac{1}{2}$. There is, however, a transition regime like in the Page setting without particle-number conservation, namely, when $1-2f \propto 1/V$. Then the two dimensions become comparable. From the equation above, we can also read off that only for $f \approx \frac{1}{2}$ is the maximum $\bar{n}_AV = nfV$ or, equivalently, $\delta n_A \approx 0$ of the weight $\varrho_{N_A} = d_A(N_A)d_B(N-N_A)/d_N$ close to N_{crit}. Although it does not say, yet, when N_{crit} is in a distance of order $\sqrt{V}$ away from the maximum $\bar{n}_AV = nfV$. Indeed, when δn_A and $1-2f$ are of order $1/\sqrt{V}$, there is a possibility that the two growing terms cancel each other so that it is not necessary that $1-2f$ needs to be of order $1/V$ as $1/\sqrt{V}$ is already suitable. This is exactly what happens in a particular regime, as we will see below.

One additional comment, when $1-2f\ > 0$ is much larger than $1/V$, we have $d_A(N_A) \ll d_B(N-N_A)$. Thus, the maximum $\bar{n}_AV$ lies in the first part of sum (A2). Thus, in the case that the terminal N_{crit} is in the exponentially suppressed part of the tail of ϱ_{N_A}, we can concentrate only on the first part of sum (A2) and disregard the second part.

Thence, we only need to understand what $N_{\text{crit}} - nfV$ is when $1-2f \ll 1$ in the limit $V \to \infty$ for the average of Eq. (A11). It is crucial to know when to cut the sum of N_A in the lower terminal and exploit the other branch of the Page curve for the remaining sum. A subtle point is that we need to find $\delta n_{\text{crit}} = N_{\text{crit}}/V - nf$ up to order $1/V$, since $N_{\text{crit}} = V(nf\ + \delta n_{\text{crit}})$ is multiplied by V and the spacing between the step size of the summation index is 1. Therefore, we have to refine the expansion in Eq. (A14).

The Taylor expansion of the Stirling approximation one order higher is given by Eq. (5.11.1) of Ref. [164]:

$$\ln[(kV)!] = \frac{1}{2}\ln[2\pi kV] + kV\ln(kV) - kV - \frac{1}{12kV} + O(V^{-2}). \quad \text{(A15)}$$

We exploit this expansion in Eq. (A13) and use the notation $N_{\text{crit}} = nfV + \delta n_{\text{crit}}V$. As N_{crit} is defined by the condition $d_A(N_{\text{crit}})/d_B(N-N_{\text{crit}}) = 1$, we take its logarithm and evaluate the resulting equation up to order $1/V$,

$$0 = \ln\left(\frac{d_A(N_{\text{crit}})}{d_B(N-N_{\text{crit}})}\right)$$
$$= \frac{1}{2}\ln\left(\frac{f\,[n(1-f)-\delta n_{\text{crit}}][(1-f)(1-n)+\delta n_{\text{crit}}]}{[nf\ +\delta n_{\text{crit}}][f\,(1-n)-\delta n_{\text{crit}}](1-f)}\right)$$
$$+ V\ln\left[\frac{f^f\,[n(1-f\,)-\delta n_{\text{crit}}]^{n(1-f\,)-\delta n_{\text{crit}}}[(1-f)(1-n)+\delta n_{\text{crit}}]^{(1-f)(1-n)+\delta n_{\text{crit}}}}{(nf\ +\delta n_{\text{crit}})^{nf\ +\delta n_{\text{crit}}}[f\,(1-n)-\delta n_{\text{crit}}]^{f\,(1-n)-\delta n_{\text{crit}}}(1-f)^{1-f}}\right]$$
$$- \frac{1}{12V}\left(\frac{1}{f}+\frac{1}{n(1-f)-\delta n_{\text{crit}}}+\frac{1}{1-f\,(1-n)+\delta n_{\text{crit}}}-\frac{1}{nf\ +\delta n_{\text{crit}}}\right.$$
$$\left.-\frac{1}{f\,(1-n)-\delta n_{\text{crit}}}-\frac{1}{1-f}\right)+O(V^{-2}). \quad \text{(A16)}$$

The very last term before the correction is at least of order $(1-2f)/V \ll 1/V$ when δn_{crit} is of order $1/\sqrt{V}$ because

$$\begin{aligned}
&\frac{1}{12V}\left(\frac{1}{f}+\frac{1}{n(1-f)-\delta n_{\mathrm{crit}}}+\frac{1}{1-f(1-n)+\delta n_{\mathrm{crit}}}-\frac{1}{nf+\delta n_{\mathrm{crit}}}-\frac{1}{f(1-n)-\delta n_{\mathrm{crit}}}-\frac{1}{1-f}\right)\\
&\quad=\frac{1-2f}{12Vf(1-f)}\left(1-\frac{1}{n(1-n)}\right)+O(V^{-3/2})\\
&\quad=O\left(\frac{1-2f}{V}\right)\\
&\quad=o(V^{-1})
\end{aligned}\tag{A17}$$

and $1-2f \ll 1$. Thus, we can neglect it.

The other terms still need to be Taylor expanded in δn_{crit} as they are of order $1/\sqrt{V}$ or smaller. Expanding this expression in δn_{crit} up to order $1/V$, which is equivalent with the fourth order in δn_{crit} as we multiply one term with V, we find that

$$\begin{aligned}
0=&\frac{1}{2}\ln\left(\frac{1-f}{f}\right)+\frac{(2n-1)\delta n_{\mathrm{crit}}}{2n(1-n)f(1-f)}+\frac{(1-2f)\delta n_{\mathrm{crit}}^2}{4f^2(1-f)^2}\left(\frac{1}{n^2}+\frac{1}{(1-n)^2}\right)\\
&+V(1-2f)\ln[n^n(1-n)^{1-n}]+2V\delta n_{\mathrm{crit}}\ln\left[\frac{1-n}{n}\right]+\frac{V(2f-1)\delta n_{\mathrm{crit}}^2}{2nf(1-f)(1-n)}\\
&+\frac{V(1-2n)\delta n_{\mathrm{crit}}^3}{6n^2(1-n)^2}\left(\frac{1}{(1-f)^2}+\frac{1}{f^2}\right)+\frac{V(2f-1)(f^2-f+1)\delta n_{\mathrm{crit}}^4}{12f^3(1-f)^3}\left(\frac{1}{n^3}+\frac{1}{(1-n)^3}\right)\\
&+o(V^{-1})\\
=&\frac{1}{2}\ln\left(\frac{1-f}{f}\right)+\frac{(2n-1)\delta n_{\mathrm{crit}}}{2n(1-n)f(1-f)}+V(1-2f)\ln[n^n(1-n)^{1-n}]+2V\delta n_{\mathrm{crit}}\ln\left[\frac{1-n}{n}\right]\\
&+\frac{V(2f-1)\delta n_{\mathrm{crit}}^2}{2nf(1-f)(1-n)}+\frac{V(1-2n)\delta n_{\mathrm{crit}}^3}{6n^2(1-n)^2}\left(\frac{1}{(1-f)^2}+\frac{1}{f^2}\right)+o(V^{-1}).
\end{aligned}\tag{A18}$$

In the second equality, we have omitted all terms that are of lower order than $1/V$, making use of the fact that $1-2f \ll 1$.

The equation above can be recast into an implicit expression for δn_{crit}. One first puts all δn_{crit}-dependent terms on one side, then pulls out the common factor δn_{crit}, and finally divides the entire expression by the prefactor of δn_{crit}, yielding

$$\begin{aligned}
\delta n_{\mathrm{crit}}=&-\left[\frac{(2n-1)}{2n(1-n)f(1-f)}+2V\ln\left[\frac{1-n}{n}\right]+\frac{V(2f-1)\delta n_{\mathrm{crit}}}{2nf(1-f)(1-n)}\right.\\
&\left.+\frac{V(1-2n)\delta n_{\mathrm{crit}}^2}{6n^2(1-n)^2}\left(\frac{1}{(1-f)^2}+\frac{1}{f^2}\right)\right]^{-1}\left[\frac{\ln[(1-f)/f]}{2}+V(1-2f)\ln[n^n(1-n)^{1-n}]\right]+O(V^{-3/2}).
\end{aligned}\tag{A19}$$

In the next step, we need to find out the leading contribution of the denominator. We pull out the factor $\ln[(1-n)/n]$ in the denominator. Then it is immediate that there are only two possibilities since the ratio $(1-2n)/\ln[(1-n)/n]$ is of order one whenever $n, 1-n>0$ are of order one. Either $(1-2f)\delta n_{\mathrm{crit}}/(1-2n)$ is of order one or larger or it is much smaller than 1. In the former case, n needs to be very close to $\frac{1}{2}$ as $(1-2f)\delta n_{\mathrm{crit}}=o(V^{-1/2})$. This means that $V(1-2f)\ln[n^n(1-n)^{1-n}]\approx -2\ln(2)V(1-2f)$ is the dominant term in the numerator. The latter, however, implies that δn_{crit} and $V(1-2f)/[V(1-2f)\delta n_{\mathrm{crit}}]=1/\delta n_{\mathrm{crit}}$ are of the same order. Hence, δn_{crit} must be of order one, which is in contradiction with the fact that it is of at most of order $1/\sqrt{V}$. In conclusion, the ratio $(1-2f)\delta n_{\mathrm{crit}}/(1-2n)$ is never of order one or larger, and the denominator is always dominated by $2V\ln[(1-n)/n]$ about which we need to expand.

The question is how far we need to go with this expansion. Here, we have to discuss different regimes.

1. The leading-order term of δn_{crit}, which is given in Eq. (A20), tells us that, for $1-2n \ll \sqrt{V}(1-2f)$, the terminal $V\delta n_{\mathrm{crit}}$ is of an order larger than $\sqrt{V}$, meaning that the maximum of the weight $\varrho_{N_A}=d_A(N_A)d_B(N-N_A)/d_N$ plus the standard deviation is out of reach of N_{crit}. Thus, we can disregard the second sum in Eq. (A2) and can extend the first sum to the upper terminal N.

2. When $V(1-2f)/(1-2n)$ ranges between order $\sqrt{V}$ and one, then it holds that

$$\delta n_{\rm crit} = -\frac{(1-2f)\ln[n^n(1-n)^{1-n}]}{2\ln[(1-n)/n]} + o(V^{-1}). \tag{A20}$$

   This means that $N_{\rm crit}$ is the integer smaller than or equal to $nfV - V(1-2f)\ln[n^n(1-n)^{1-n}]/2\ln[(1-n)/n]$.
3. In the case $(1-2n) \gg V(1-2f)$, $\delta n_{\rm crit}$ is so close to 0 that we can set $N_{\rm crit}$ as the integer smaller than or equal to nfV. This case can also be seen as a particular limit of the second case.

Note that, for all three cases, we have $1-2f \ll 1$. However, case 1 shows that, for $1-2f \gg 1/\sqrt{V} \gtrsim (1-2n)/\sqrt{V}$, we will never find a contribution from $N_{\rm crit}$ as it will always be exponentially suppressed by the weight ϱ_{N_A}. We recall that the cases $1-2f \gg 1/\sqrt{V}$ are always suppressed due to the exponentially small ratio $d_A(N_A)/d_B(N-N_A)$ regardless what n is. Thus, the earliest occurrence of order $\sqrt{V}$ and order-one terms in the entanglement entropy will be when f enters the regime about $f \approx \frac{1}{2}$ at a distance of order $1/\sqrt{V}$.

Let us highlight that this approximation of $N_{\rm crit}$ only works for the average over Eq. (A11). For the average over the ratio $\min\{d_A(N_A)/d_B(N-N_A), d_B(N-N_A)/d_A(N_A)\}$, we need to be more careful as there are additional exponential terms in δn_A that may shift the maximum of the summands in the average over ϱ_{N_A}.

c. Moments of ϱ_{N_A}

To compute the entanglement entropy, we need the moments of the weight $\varrho_{N_A} = d_A(N_A)d_B(N-N_A)/d_N$ either for the full sum with terminals at 0 and N or for that with a cutoff due to $N_{\rm crit}$. In general, we need to compute $\mathcal{M}_j$ defined in Eq. (A4). It can be rewritten in terms of a derivative acting on a sum, i.e.,

$$\begin{aligned}\mathcal{M}_j = {}& (\partial_\lambda - nf)^j \\ & \times \Bigg[c_- \sum_{N_A=0}^{N} \frac{V_A!\,(V-V_A)!\,N!\,(V-N)!}{V!\,N_A!\,(V_A-N_A)!\,(N-N_A)!\,(V-V_A-N+N_A)!} e^{N_A\lambda/V} \\ & + (c_+ - c_-) \\ & \times \sum_{N_A=N_{\rm crit}+1}^{N} \frac{V_A!\,(V-V_A)!\,N!\,(V-N)!}{V!\,N_A!\,(V_A-N_A)!\,(N-N_A)!\,(V-V_A-N+N_A)!} e^{N_A\lambda/V} \Bigg]_{\lambda=0}\end{aligned} \tag{A21}$$

with $j \in \mathbb{N}$. The two coefficients $c_\pm$ grow maximally with V, and $\Delta c = c_+ - c_-$ grows at most like $\sqrt{V}$ for $j=0$ and like V for $j=1$; cf. Eqs. (A11) and (A12). The second sum without Δc has been denoted by $\Delta\mathcal{M}_j$ [see (A4)], while the first sum will be labeled as

$$\widetilde{\mathcal{M}}_j = (\partial_\lambda - nf)^j \sum_{N_A=0}^{N} \frac{V_A!\,(V-V_A)!\,N!\,(V-N)!}{V!\,N_A!\,(V_A-N_A)!\,(N-N_A)!\,(V-V_A-N+N_A)!} e^{N_A\lambda/V}\Big|_{\lambda=0}. \tag{A22}$$

Evaluation of the first sum (A22). As we have seen, the maximum of ϱ_{N_A} is close to $\bar{n}_A = nfV$, which always lies below $N_{\rm crit}$. It can be approximated by a Gaussian with variance $\sqrt{V}$. Hence, when $N_{\rm crit} - nfV$ is larger than order $\sqrt{V}$, only the first sum contributes, and we can approximate Eq. (A21) as $\mathcal{M}_j = \widetilde{\mathcal{M}}_j + o(1)$. The error is exponentially small due to the Gaussian tail. Otherwise, the sum $\Delta\mathcal{M}_j$ will contribute, too. Yet, the ensuing computation still applies for the first sum (A22) as it only takes into account the fact that the upper terminal in the sum is N.

To compute Eq. (A22), we make use of the contour integral

$$\frac{(V-N)!}{(V_A-N_A)!\,(V-V_A-N+N_A)!} = \oint_{|z|=1} \frac{dz}{2\pi i z} \frac{(1+z)^{V-N}}{z^{V_A-N_A}}, \tag{A23}$$

which is based on the binomial sum $(1+z)^{V-N} = \sum_{l=0}^{V-N} \binom{V-N}{l} z^l$. Then, we can carry out the sum exactly and find that

$$\widetilde{\mathcal{M}}_j = c_- \frac{V_A!\,(V-V_A)!}{V!} (\partial_\lambda - nf)^j \oint_{|z|=1} \frac{dz}{2\pi i z} \frac{(1+z)^{V-N}(1+ze^{\lambda/V})^N}{z^{V_A}}\bigg|_{\lambda=0}. \tag{A24}$$

Comparing with Eq. (A10), we only need the first and second centered moments. We note that, due to the proper normalization of ϱ_{N_A}, it holds that $\widetilde{\mathcal{M}}_0 = 1$. The first centered moment,

$$\begin{aligned}
\widetilde{\mathcal{M}}_1 &= \frac{V_A!\,(V-V_A)!}{V!} \oint_{|z|=1} \frac{dz}{2\pi i z} [nz - nf\,(1+z)] \frac{(1+z)^{V-1}}{z^{V_A}} + o(1) \\
&= \frac{V_A!\,(V-V_A)!}{V!} \left[n(1-f) \frac{(V-1)!}{(V_A-1)!\,(V-V_A)!} - nf \frac{(V-1)!}{V_A!\,(V-V_A-1)!} \right] \\
&= \frac{n(1-f)V_A}{V} - nf \frac{(V-V_A)}{V} \\
&= 0,
\end{aligned} \tag{A25}$$

vanishes exactly without any approximation; note that $V_A/V = f$. In the same manner we compute the second centered moment, where we have to go up to order $1/V$ as we multiply by V [see Eq. (A11)]:

$$\begin{aligned}
\widetilde{\mathcal{M}}_2 &= \frac{V_A!\,(V-V_A)!}{V!} \oint_{|z|=1} \frac{dz}{2\pi i z} \left[n^2(1-f)^2 z^2 + \left(\frac{n}{V} - 2n^2 f\,(1-f) \right) z + n^2 f^2 \right] \frac{(1+z)^{V-2}}{z^{V_A}} \\
&= \frac{V_A!\,(V-V_A)!}{V!} \left[n^2(1-f)^2 \frac{(V-2)!}{(V_A-2)!\,(V-V_A)!} \right. \\
&\quad + \left(\frac{n}{V} - 2n^2 f\,(1-f) \right) \frac{(V-2)!}{(V_A-1)!\,(V-V_A-1)!} \\
&\quad \left. + n^2 f^2 \frac{(V-2)!}{V_A!\,(V-V_A-2)!} \right] \\
&= n^2(1-f)^2 f \left(f + \frac{f-1}{V} \right) + \left(\frac{n}{V} - 2n^2 f\,(1-f) \right) f\,(1-f) \left(1 + \frac{1}{V} \right) \\
&\quad + n^2 f^2 (1-f) \left(1 - f - \frac{f}{V} \right) + O(V^{-2}) \\
&= \frac{n(1-n) f\,(1-f)}{V} + O(V^{-2}).
\end{aligned} \tag{A26}$$

The third equality follows by Taylor expanding in $1/V$, which amounts to an error term starting with the order $1/V^2$. Actually, the leading orders of the two centered moments are not very surprising as they are the mean and the variance of ϱ_{N_A}. This calculation only checks that there are no additional corrections that may become relevant.

Returning to the approximation $\mathcal{M}_j = \widetilde{\mathcal{M}}_j + o(1)$ for $N_{\text{crit}} - nfV \gg \sqrt{V}$, the first moment $\mathcal{M}_1$ is exponentially suppressed as it is only given by the second sum $\Delta\mathcal{M}_1$ that lies in the exponentially small tail of ϱ_{N_A}. In contrast, the correction to the leading order in the second centered moment $\mathcal{M}_2$ is of order $1/V$ when c_- is of order V, which is indeed the case; see Eq. (A11).

Finally, we combine the two moments $\widetilde{\mathcal{M}}_1$ and $\widetilde{\mathcal{M}}_2$ with expansion (A11) to compute the sum

$$\begin{aligned}
&\sum_{N_A=0}^{N} \frac{d_A(N_A) d_B(N-N_A)}{d_N} \ln\left(\frac{d_N}{d_B(N-N_A)} \right) \\
&\quad = -Vf\,[n\ln(n) + (1-n)\ln(1-n)] + \frac{f}{2} + \frac{\ln(1-f)}{2} + O\left(\frac{1}{\sqrt{V}} \right).
\end{aligned} \tag{A27}$$

This is our first intermediate result, which will be needed when combining it with the full entanglement entropy.

The case $N_{\rm crit} - nfV \leq O(\sqrt{V})$ *and the term* $\Delta\mathcal{M}_j$. When $N_{\rm crit}$ is in the vicinity of $\bar{n}_A V = nfV = nV/2 + n(2f\ -1)V/2$ (namely of order $\sqrt{V}$), we need to take into account the fact that both sums, $\widetilde{\mathcal{M}}_j$ and $\Delta\mathcal{M}_j$, contribute. Recall that $\delta f\ = 1 - 2f$ must be of order $1/\sqrt{V}$ or smaller and $\delta n = 1 - 2n$ cannot be less than order $\sqrt{V}(1 - 2f)$ to find this regime. This implies that the coefficient for the constant term ($j = 0$) of Eq. (A12) will be at most of order $\sqrt{V}$. Thus, we need to expand the sum for this case up to order $1/\sqrt{V}$. For $j = 1$, we still need to expand up to order $1/V$, though this expansion does not have a term of order one, as we will see.

We consider the sum

$$\Delta\mathcal{M}_j = \left(\partial_\lambda + \frac{N_{\rm crit}+1}{V} - nf\right)^j \times \left[\sum_{N_A=N_{\rm crit}+1}^{N} \frac{V_A!\,(V-V_A)!\,N!\,(V-N)!}{V!\,N_A!\,(V_A-N_A)!\,(N-N_A)!\,(V-V_A-N+N_A)!} e^{(N_A-N_{\rm crit}-1)\lambda/V}\right]_{\lambda=0} \tag{A28}$$

for $j = 0, 1$, where we have pushed a factor $\exp[-(N_{\rm crit}+1)\lambda/V]$ into the derivative as it will be beneficial for the ensuing calculation. The summands can be extended by $(N - N_{\rm crit} - 1)!\,N_{\rm crit}!\,(N_A - N_{\rm crit} - 1)!$ in the numerator and the denominator to exploit the identity

$$\frac{N_{\rm crit}!\,(N_A - N_{\rm crit} - 1)!\,(V-N)!}{N_A!\,(V_A-N_A)!\,(V-V_A-N+N_A)!} = \int_0^1 dx \oint_{|z|=1} \frac{dz}{2\pi i z} \frac{(1+z)^{V-N}}{z^{V-V_A-N+N_A}} x^{N_A-N_{\rm crit}-1}(1-x)^{N_{\rm crit}}. \tag{A29}$$

The integrand over the auxiliary variable x is also known as the beta distribution. Let us underline that it always holds that $N_{\rm crit} < N_A \leq N$ for $\Delta\mathcal{M}_j$. The sum can be carried out exactly anew, and we find that, for the zeroth moment,

$$\Delta\mathcal{M}_0 = \frac{V_A!\,(V-V_A)!\,N!}{N_{\rm crit}!(N-N_{\rm crit}-1)!\,V!} \int_0^1 dx \oint_{|z|=1} \frac{dz}{2\pi i z} \frac{(1+z)^{V-N}}{z^{V-V_A}} (1-x)^{N_{\rm crit}}(x+z)^{N-N_{\rm crit}-1} \tag{A30}$$

and, for the first moment,

$$\Delta\mathcal{M}_1 = \frac{V_A!\,(V-V_A)!\,N!}{N_{\rm crit}!\,(N-N_{\rm crit}-1)!\,V!} \times \int_0^1 dx \oint_{|z|=1} \frac{dz}{2\pi i z}\left[\frac{N-N_{\rm crit}-1}{V}x + \left(\frac{N_{\rm crit}+1}{V} - nf\right)(x+z)\right] \times \frac{(1+z)^{V-N}}{z^{V-V_A}}(1-x)^{N_{\rm crit}}(x+z)^{N-N_{\rm crit}-2}. \tag{A31}$$

This time the integral is too involved to compute the results directly in a closed form. We use a saddle point approximation. The derivatives in x and z of the logarithm of the integrand yield in leading order the saddle point equations

$$\frac{N-N_{\rm crit}}{x_0+z_0} - \frac{N_{\rm crit}}{1-x_0} = 0 \quad \text{and} \quad \frac{V-N}{1+z_0} - \frac{V-V_A}{z_0} + \frac{N-N_{\rm crit}}{x_0+z_0} = 0. \tag{A32}$$

Since $N_{\rm crit} = nfV + \Delta N_{\rm crit}$ with $\Delta N_{\rm crit} = O(\sqrt{V})$, and the standard deviation about the saddle points is of order $1/\sqrt{V}$, we can replace $N_{\rm crit}$ by nfV in the saddle point equations without changing the integrals. The same holds for the deviation of f from $\frac{1}{2}$ as $1 - 2f$ is at most of order $1/\sqrt{V}$ in the present case. This yields the approximate saddle point solutions

$$x_0 = 0 \quad \text{and} \quad z_0 = 1. \tag{A33}$$

Thus, we expand $x = \delta x/\sqrt{V}$ and $z = 1 + i\delta z/\sqrt{V}$ with $\delta x \in \mathbb{R}_+$ and $\delta z \in \mathbb{R}$.

For the first moment, this yields a drastic simplification because of the factor

$$\begin{aligned}&\frac{N-N_{\rm crit}-1}{V}x+\left(\frac{N_{\rm crit}+1}{V}-nf\right)(x+z)\\&=\frac{n}{2\sqrt{V}}\delta x+\frac{\Delta N_{\rm crit}}{V}+\frac{1}{2V}\left(2+\sqrt{V}(1-2f)n\delta x+2i\frac{\Delta N_{\rm crit}}{\sqrt{V}}\delta z\right)\\&\quad+O(V^{-3/2}),\end{aligned}\tag{A34}$$

which already starts with order $1/\sqrt{V}$. Therefore, we only need to go to order $1/\sqrt{V}$ in all other factors in $\Delta\mathcal{M}_1$ like in $\Delta\mathcal{M}_0$.

The single terms we need to expand are [note that $1-2f=O(1/\sqrt{V})$ or smaller]

$$\begin{aligned}(1+z)^{V-N}&=2^{(1-n)V}\left(1+i\frac{1}{2\sqrt{V}}\delta z\right)^{(1-n)V}\\&=2^{(1-n)V}\exp\left[i\frac{\sqrt{V}(1-n)}{2}\delta z+\frac{1-n}{8}\delta z^2\right]\left[1-i\frac{1-n}{24\sqrt{V}}\delta z^3+O(V^{-1})\right],\end{aligned}\tag{A35}$$

$$\begin{aligned}z^{-V+V_A-1}&=\left(1+i\frac{\delta z}{\sqrt{V}}\right)^{-(1+1-2f)V/2-1}\\&=\exp\left[-i\frac{\sqrt{V}}{2}\delta z-\frac{1}{4}(2i\sqrt{V}(1-2f)\delta z+\delta z^2)\right]\\&\quad\times\left[1-i\frac{\delta z}{\sqrt{V}}-\frac{1-2f}{4}\delta z^2+i\frac{\delta z^3}{6}+O(V^{-1})\right],\end{aligned}\tag{A36}$$

$$\begin{aligned}(1-x)^{N_{\rm crit}}&=\left(1-\frac{\delta x}{\sqrt{V}}\right)^{n(1-1+2f)V/2+\Delta N_{\rm crit}}\\&=\exp\left[-\frac{n\sqrt{V}}{2}\delta x-\left(\frac{\Delta N_{\rm crit}}{\sqrt{V}}-\frac{\sqrt{V}(1-2f)n}{2}\right)\delta x-\frac{n}{4}\delta x^2\right]\\&\quad\times\left[1-\frac{1}{2\sqrt{V}}\left(\frac{\Delta N_{\rm crit}}{\sqrt{V}}-\frac{\sqrt{V}(1-2f)n}{2}\right)\delta x^2-\frac{n}{6\sqrt{V}}\delta x^3+O(V^{-1})\right],\end{aligned}\tag{A37}$$

and

$$\begin{aligned}(x+z)^{N-N_{\rm crit}-j}&=\left(1+\frac{\delta x+i\delta z}{\sqrt{V}}\right)^{n(1+1-2f)V/2-\Delta N_{\rm crit}-j}\\&=\exp\left[\left(\frac{n\sqrt{V}}{2}-\frac{\Delta N_{\rm crit}}{\sqrt{V}}+\frac{\sqrt{V}(1-2f)n}{2}\right)(\delta x+i\delta z)-\frac{n}{4}(\delta x+i\delta z)^2\right]\\&\quad\times\left[1-\frac{j}{\sqrt{V}}(\delta x+i\delta z)+\left(\frac{\Delta N_{\rm crit}}{2V}-\frac{(1-2f)n}{4}\right)(\delta x+i\delta z)^2\right.\\&\quad\left.+\frac{n}{6\sqrt{V}}(\delta x+i\delta z)^3+O(V^{-1})\right],\end{aligned}\tag{A38}$$

with $j = 1, 2$. Additionally, we expand the factorial prefactors using Stirling's formula,

$$\begin{aligned}
&\frac{V_A!\,(V-V_A)!\,N!}{N_{\rm crit}!\,(N-N_{\rm crit}-1)!\,V!} \\
&\quad= \frac{2^{(n-1)V}(1-1+2f)^{(1-1+2f)V/2}}{[1-1+2f+2\Delta N_{\rm crit}/(nV)]^{n(1-1+2f)V/2+\Delta N_{\rm crit}}} \\
&\qquad\times \frac{(1+1-2f)^{(1+1-2f)V/2}}{[1+1-2f-2\Delta N_{\rm crit}/(nV)]^{n(1+1-2f)V/2-\Delta N_{\rm crit}}} \\
&\qquad\times \left(1+1-2f-\frac{2\Delta N_{\rm crit}}{nV}\right)\frac{V\sqrt{n}}{2}[1+O(V^{-1})] \\
&\quad= 2^{(n-1)V}\frac{V\sqrt{n}}{2}\exp\left[\frac{V}{2}(1-2f)^2-\frac{(V(1-2f)n-2\Delta N_{\rm crit})^2}{2nV}\right] \\
&\qquad\times\left[1+(1-2f)-\frac{2\Delta N_{\rm crit}}{V}+O(V^{-1})\right].
\end{aligned} \tag{A39}$$

Collecting all our previous results, and integrating over $\delta x \in \mathbb{R}_+$ and $\delta z \in \mathbb{R}$, we find that

$$\Delta\mathcal{M}_0 = \frac{1}{2}\mathrm{erfc}\left[\sqrt{\frac{2}{n(1-n)V}}\Delta N_{\rm crit}\right] - \frac{1}{\sqrt{2\pi n(1-n)V}}\exp\left[-\frac{2\Delta N_{\rm crit}^2}{n(1-n)V}\right] + O(V^{-1}) \tag{A40}$$

and

$$\Delta\mathcal{M}_1 = \frac{(1-n)n+2\Delta N_{\rm crit}/V}{\sqrt{8\pi(1-n)nV}}\exp\left[-\frac{2\Delta N_{\rm crit}^2}{n(1-n)V}\right] + O(V^{-3/2}). \tag{A41}$$

These two equations are our second ingredient. In particular, when combining these with Eq. (A12), we find that

$$\begin{aligned}
&\sum_{N_A=N_{\rm crit}+1}^{N}\frac{d_A(N_A)d_B(N-N_A)}{d_N}\ln\left(\frac{d_B(N-N_A)}{d_A(N_A)}\right) \\
&\quad= -V(1-2f)[n\ln(n)+(1-n)\ln(1-n)] \\
&\qquad\times\left(\frac{1}{2}\mathrm{erfc}\left[\sqrt{\frac{2}{n(1-n)V}}\Delta N_{\rm crit}\right]-\frac{1}{\sqrt{2\pi n(1-n)V}}\exp\left[-\frac{2\Delta N_{\rm crit}^2}{n(1-n)V}\right]\right) \\
&\qquad+\sqrt{V}\ln\left(\frac{n}{1-n}\right)\frac{(1-n)n+2\Delta N_{\rm crit}/V}{\sqrt{2\pi(1-n)n}}\exp\left[-\frac{2\Delta N_{\rm crit}^2}{n(1-n)V}\right]+o(1),
\end{aligned} \tag{A42}$$

which is our second intermediate result. As one can readily check, this term starts at most with order $\sqrt{V}$ as $1-2f$ is at most of order $1/\sqrt{V}$ in the present case. Additionally, it will contribute order-one terms.

d. Average over dimension ratios

We now turn to evaluate sum (A6). We need to expand the average

$$\begin{aligned}
&\sum_{N_A=0}^{\min(N,V_A)}\frac{d_A(N_A)d_B(N-N_A)}{d_N}\min\{d_A(N_A)/d_B(N-N_A), d_B(N-N_A)/d_A(N_A)\} \\
&\quad=\sum_{N_A=0}^{N_{\rm crit}}\frac{d_A^2(N_A)}{d_N}+\sum_{N_A=N_{\rm crit}+1}^{\min(N,V_A)}\frac{d_B^2(N-N_A)}{d_N}.
\end{aligned} \tag{A43}$$

For that purpose, it is paramount to take the correct bounds of the summing index as either $d_B(N-N_A)$ or $d_A(N_A)$ are canceled in the weight $\varrho_{N_A} = d_A(N_A)d_B(N-N_A)/d_N$ by the observable $\min\{d_A(N_A)/d_B(N-N_A), d_B(N-N_A)/d_A(N_A)\}$.

Note that, in Eq. (A21), the weight ϱ_{N_A} allowed us to extend the sum from $N_A = 0$ to $N_A = N$ instead of $N_A = \min(N, V_A)$. The weight ϱ_{N_A} implemented the correct terminal.

We need to understand where $d_A(N_A)$ and $d_B(N - N_A)$ become maximal. This is indeed given for $\bar{N}_A^{(1)} = V_A/2 = fV/2$ for $d_A(N_A)$ and $\bar{N}_A^{(2)} = N - (V - V_A)/2 = (2n - 1 + f)V/2$ for $d_B(N - N_A)$ because both are binomial weights whose maximum is always at the center of the distribution.

To see which of the two dimensions is largest, we take their ratio and expand in large V:

$$\begin{aligned}\frac{d_A(\bar{N}_A^{(1)})}{d_B(N - \bar{N}_A^{(2)})} &= \frac{V_A!}{(V - V_A)!}\left(\frac{[(V - V_A)/2]!}{(V_A/2)!}\right)^2 \\ &= \frac{(fV)!}{[(1 - f)V]!}\left(\frac{[(1 - f)V/2]!}{(fV/2)!}\right)^2 \\ &= \sqrt{\frac{1 - f}{f}}\, 2^{(2f-1)V}[1 + O(V^{-1})]. \end{aligned} \tag{A44}$$

Hence, for suitably large V, it is always $d_A(\bar{N}_A^{(1)}) \leq d_B(N - \bar{N}_A^{(2)})$ because $f \leq \frac{1}{2}$. Note that we compare the maximums of the two dimensions at different N_A and not at the same one. However, the ratio of the two dimensions at the same $N_A = \bar{N}_A^{(2)}$ will always be the smaller one, i.e., $d_A(\bar{N}_A^{(2)}) \leq d_A(\bar{N}_A^{(1)}) \leq d_B(N - \bar{N}_A^{(2)})$ [the first inequality follows from the fact that the maximum of $d_A(N_A)$ is achieved at $\bar{N}_A^{(1)}$]. Hence, it is clear that $\bar{N}_A^{(2)} \leq N_{\text{crit}}$. In particular, the maximum of $d_B^2(N - N_A)/d_N$ cannot be achieved at $\bar{N}_A^{(2)}$ but is only given at the lower terminal $N_A = N_{\text{crit}} + 1$. Since $d_B^2(N - N_A)/d_N < \varrho_{N_A}$ for $N_A \geq N_{\text{crit}}$, the second sum can only contribute when N_{crit} has a distance of order $\sqrt{V}$ to the maximum $\bar{n}_A V = nfV$ of ϱ_{N_A}. As we have seen, this is only the case when $1 - 2f$ is of order $1/\sqrt{V}$ or less and $1 - 2n \ll \sqrt{V}(1 - 2f)$.

Let us first consider the case $1 - 2f \gg V^{-1/2}$ as it is still unclear whether sum (A43) yields anything of order one in this case. When $N_A \leq N_{\text{crit}}$, we have, for the first part of the sum,

$$\frac{d_A^2(N_A)}{d_N} < \frac{d_A(N_A)d_B(N - N_A)}{d_N} = \varrho_{N_A}. \tag{A45}$$

This means that $d_A^2(N_A)/d_N$ is exponentially suppressed when N_A has a distance larger than order $\sqrt{V}$ to the maximum $\bar{n}_A V = nfV$ of ϱ_{N_A}. The maximum $\bar{N}_A^{(1)} = fV/2$ of $d_A(N_A)$ is evidently far away when n is not close to $\frac{1}{2}$. When expanding $d_A^2(N_A)/d_N$ about $N_A = nfV + \delta n_A V$ with δn_A of order $1/\sqrt{V}$, the expansion up to order one is

$$\begin{aligned}\frac{d_A^2(N_A)}{d_N} = \frac{1}{\sqrt{2\pi V n(1 - n)f^2}}&\left(\frac{1 - n}{n}\right)^{2V\delta n_A}[(1 - n)^{1-n}n^n]^{V(1-2f)} \\ &\times \exp\left[-\frac{V\delta n_A^2}{n(1 - n)f}\right][1 + O(V^{-1/2})]. \end{aligned} \tag{A46}$$

Since $f, n \in (0, \frac{1}{2}]$, the leading term in V exponentially suppresses all summands if $1 - 2f \gg V^{-1/2}$. Similarly, we have an exponential suppression of $d_B^2(N - N_A)/d_N < d_A(N_A)d_B(N - N_A)/d_N = \varrho_{N_A}$ for any $N_A > N_{\text{crit}}$ because then $\bar{n}_A V = nfV$ is certainly further away than a distance $\sqrt{V}$ from N_A. Therefore, the second part (A6) of the entanglement entropy does not contribute for $1 - 2f \gg V^{-1/2}$ in the thermodynamic limit.

When $1 - 2f = O(V^{-1/2})$, the leading term in Eq. (A46) slightly shifts the maximum of the summands, as we have seen for N_{crit}. Actually, N_{crit}, which is $(nf + \delta n_{\text{crit}})V$ with Eq. (A20) in this scaling, will be the maximum since $(1 - n)/n \geq 1$ for $n \leq \frac{1}{2}$, such that, whenever $N_{\text{crit}}/V - n_A = \delta n_{\text{crit}} - \delta n_A$ is larger than order $1/V$, the ratio $d_A^2(N_A)/d_N$ will again be exponentially suppressed. Also, for $d_B^2(N - N_A)/d_N$, the maximum lies at N_{crit} when $N_A > N_{\text{crit}}$, as we have seen that the maximum $\bar{N}_A^{(2)} < N_{\text{crit}}$ and $d_B(N - N_A)$ is monotonously decreasing for $N_A > \bar{N}_A^{(2)}$. Its expansion about $N_A = (nf +$

$\delta n_A)V$ is equal to

$$\frac{d_B^2(N_A)}{d_N} = \frac{1}{\sqrt{2\pi Vn(1-n)(1-f)^2}}\left(\frac{n}{1-n}\right)^{2V\delta n_A}[(1-n)^{1-n}n^n]^{V(2f-1)} \times \exp\left[-\frac{V\delta n_A^2}{n(1-n)(1-f)}\right][1+O(V^{-1/2})]. \tag{A47}$$

Let us try to simplify both expansions (A46) and (A47) further. We know that they are only valid when N_{crit} has a distance of order $\sqrt{V}$ to $\bar{n}_A V = nfV$. In Appendix A 1 b, we showed that this is only the case when $\sqrt{V}(1-2f)$ is maximally of order $1-2n$. Hence, $1-2f$ must be maximally of order $1/\sqrt{V}$ in which we can expand. However, there are two cases to discuss in the scaling of the variable n depending on the exact scaling of $1-2f$.

When $1-2f = O(V^{-1/2})$, we have already mentioned that N_A takes its maximum at $N_{\text{crit}} = \lfloor nf + \delta n_{\text{crit}})V\rfloor$, which is the largest integer that is smaller than or equal to $(nf + \delta n_{\text{crit}})V = (nf - (\frac{1}{2}-f)\ln[n^n(1-n)^{1-n}]/\ln[(1-n)/n])V + o(1)$; see Eq. (A20). This holds for both sums (A43). One can readily see from the Eqs. (A46) and (A47) that only summands where $|N_A - N_{\text{crit}}|$ is of order one are of the same order as the maximum while the other terms are suppressed. Thus, we substitute $N_A = N_{\text{crit}} - j = nfV + \Delta N_{\text{crit}} - j$ for $N_A \leq N_{\text{crit}}$ and $N_A = N_{\text{crit}} + 1 + j = nfV + \Delta N_{\text{crit}} + 1 + j$ for $N_A > N_{\text{crit}}$ with $j = 0, 1, 2, \ldots$ and

$$\Delta N_{\text{crit}} = N_{\text{crit}} - nfV = \lfloor \delta n_{\text{crit}} V\rfloor + O(1) = \left\lfloor -\frac{V(1-2f)\ln[n^n(1-n)^{1-n}]}{2\ln[(1-n)/n]}\right\rfloor + O(1), \tag{A48}$$

where the error is only a number in $[0, 1)$. Substituting expansions (A46) and (A47) into Eq. (A43), we find that

$$\begin{aligned}
&\sum_{N_A=0}^{\min(N,V_A)} \frac{d_A(N_A)d_B(N-N_A)}{d_N}\min\{d_A(N_A)/d_B(N-N_A), d_B(N-N_A)/d_A(N_A)\} \\
&= \sum_{j=0}^{\infty}\frac{1}{\sqrt{2\pi Vn(1-n)f^2}}\left(\frac{1-n}{n}\right)^{2\Delta N_{\text{crit}}-2\delta n_{\text{crit}}V-2j}\exp\left[-\frac{(\Delta N_{\text{crit}}-j)^2}{n(1-n)fV}\right] \\
&\quad + \sum_{j=0}^{\infty}\frac{1}{\sqrt{2\pi Vn(1-n)(1-f)^2}}\left(\frac{n}{1-n}\right)^{2\Delta N_{\text{crit}}-2\delta n_{\text{crit}}V+2+2j} \\
&\quad \times \exp\left[-\frac{(\Delta N_{\text{crit}}+1+j)^2}{n(1-n)(1-f)V}\right] + O(V^{-1}),
\end{aligned} \tag{A49}$$

since in the first sum it is $\delta n_a V = \Delta N_{\text{crit}} - j$ and in the second sum it is $\delta n_A V = \Delta N_{\text{crit}} + 1 + j$. We have extended the sum to ∞ because we only add exponentially suppressed terms to it, and we have used the expression of δn_{crit} from Eq. (A20). The j dependence in the Gaussian part of Eq. (A49) can be dropped since $j/\sqrt{V}$ is of order $1/\sqrt{V}$; note that $\Delta N_{\text{crit}} = O(\sqrt{V})$. Using the fact that the geometric series converge because $1-2n \in (0, 1)$ is of order one and all the summands are of order $1/\sqrt{V}$ due to the prefactor, it becomes immediate that sum (A6) still vanishes; although it now vanishes like $1/\sqrt{V}$.

A similar argument holds when $1-2n \gg 1/\sqrt{V}$ is of order one even if $(1-2f)\sqrt{V}$ is maximally of order $1-2n$ and, hence, rapidly vanishing. In this case the variable $x = (1-2n)j$ becomes quasicontinuous, meaning that j will be of order $1/(1-2n)$. We need to scale j with $1-2n$ since the quadratic part from the Gaussian, namely $j/\sqrt{V} = x/[\sqrt{V}(1-2n)] \ll 1$. This yields an integral over an exponential function showing that both sums in Eq. (A43) vanish like $1/[\sqrt{V}(1-2n)]$. It is certainly much weaker, but it is vanishing nonetheless.

Therefore, only in the case $(1-2f)\sqrt{V} \ll 1-2n \ll 1$ can sum (A6) contribute to the entanglement entropy of order one and above. The two expansions (A46) and (A47) simplify in this case to

$$\begin{aligned}
\frac{d_A^2(N_A)}{d_N} &= 2^{V(2f-1)}\sqrt{\frac{8}{\pi V}}\exp[-8V\delta n_A^2 + 4V(1-2n)\delta n_A][1+O(V^{-1/2})], \\
\frac{d_B^2(N_A)}{d_N} &= 2^{V(1-2f)}\sqrt{\frac{8}{\pi V}}\exp[-8V\delta n_A^2 - 4V(1-2n)\delta n_A][1+O(V^{-1/2})].
\end{aligned} \tag{A50}$$

The variable δn_A becomes quasicontinuous as it is of order $1/\sqrt{V}$, meaning that the sum is replaced by an integral. Thus, the final computation of sum (A43) is

$$\begin{aligned}
&\sum_{N_A=0}^{\min(N,V_A)} \frac{d_A(N_A)d_B(N-N_A)}{d_N} \min\{d_A(N_A)/d_B(N-N_A), d_B(N-N_A)/d_A(N_A)\} \\
&= 2^{V(2f-1)} \int_{-\infty}^{\ln(2)(1-2f)/[4(1-2n)]} \frac{4\sqrt{V}d\delta n_A}{\sqrt{2\pi}} \exp[-8V\delta n_A^2 + 4V(1-2n)\delta n_A] \\
&\quad + 2^{V(1-2f)} \int_{\ln(2)(1-2f)/[4(1-2n)]}^{\infty} \frac{4\sqrt{V}d\delta n_A}{\sqrt{2\pi}} \exp[-8V\delta n_A^2 - 4V(1-2n)\delta n_A] + o(1) \\
&= 2^{V(2f-1)-1} \exp\left[\frac{V(1-2n)^2}{2}\right] \operatorname{erfc}\left[\sqrt{V}\frac{(1-2n)^2 - \ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right] \\
&\quad + 2^{V(1-2f)-1} \exp\left[\frac{V(1-2n)^2}{2}\right] \operatorname{erfc}\left[\sqrt{V}\frac{(1-2n)^2 + \ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right] + o(1),
\end{aligned} \tag{A51}$$

where we have made use of the complementary error function.

Also, in this formula one can readily check the various asymptotic limits. If either $1-2n \gg 1/\sqrt{V}$ or $1-2f \gg 1/V$, the term vanishes, as already pointed out. For the three cases discussed in Appendix A 1 b, the following conclusions hold.

Case 1: $(1-2n) \ll \sqrt{V}(1-2f) = O(1/\sqrt{V})$. It holds that

$$\begin{aligned}
&\sum_{N_A=0}^{\min(N,V_A)} \frac{d_A(N_A)d_B(N-N_A)}{d_N} \min\left\{\frac{d_A(N_A)}{d_B(N-N_A)}, \frac{d_B(N-N_A)}{d_A(N_A)}\right\} \\
&= 2^{V(2f-1)} + o(1).
\end{aligned} \tag{A52}$$

Case 2: $(1-2n) = O[\sqrt{V}(1-2f)] = O(1/\sqrt{V})$. We need to keep the result as is.
Case 3: $O(1/\sqrt{V}) = (1-2n) \gg \sqrt{V}(1-2f)$. We obtain

$$\begin{aligned}
&\sum_{N_A=0}^{\min(N,V_A)} \frac{d_A(N_A)d_B(N-N_A)}{d_N} \min\left\{\frac{d_A(N_A)}{d_B(N-N_A)}, \frac{d_B(N-N_A)}{d_A(N_A)}\right\} \\
&= \exp\left[\frac{V(1-2n)^2}{2}\right] \operatorname{erfc}\left[\sqrt{V}\frac{(1-2n)}{\sqrt{2}}\right] + o(1).
\end{aligned} \tag{A53}$$

The third case covers part of the second case and the third case in Appendix A 1 b. In the first case, the exponential decay is not always completely correct as we have seen in the discussion above. The reason is that the error term still contains algebraic decaying terms that will take over.

e. Resulting formula

Collecting all intermediate results (A27), (A42), and (A51), we arrive at

$$\begin{aligned}
\langle S_A \rangle_N &= -Vf\,[n\ln(n) + (1-n)\ln(1-n)] + \frac{f}{2} + \frac{\ln(1-f)}{2} \\
&\quad - V(1-2f)\ln[n^n(1-n)^{1-n}] \\
&\quad \times \left(\frac{1}{2}\operatorname{erfc}\left[\sqrt{\frac{2}{n(1-n)V}}\Delta N_{\text{crit}}\right] - \frac{1}{\sqrt{2\pi n(1-n)V}} \exp\left[-\frac{2\Delta N_{\text{crit}}^2}{n(1-n)V}\right]\right)
\end{aligned}$$

$$+\sqrt{V}\ln\left(\frac{n}{1-n}\right)\frac{(1-n)n+2\Delta N_{\rm crit}/V}{\sqrt{2\pi(1-n)n}}\exp\left[-\frac{2\Delta N_{\rm crit}^2}{n(1-n)V}\right]$$
$$-2^{V(2f-1)-2}\exp\left[\frac{V(1-2n)^2}{2}\right]\mathrm{erfc}\left[\sqrt{V}\frac{(1-2n)^2-\ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right]$$
$$-2^{V(1-2f)-2}\exp\left[\frac{V(1-2n)^2}{2}\right]\mathrm{erfc}\left[\sqrt{V}\frac{(1-2n)^2+\ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right]+o(1). \tag{A54}$$

What remains to be discussed is the parameter

$$\Delta N_{\rm crit}=\left\lfloor -\frac{V(1-2f)\ln[n^n(1-n)^{1-n}]}{2\ln[(1-n)/n]}\right\rfloor+O(1). \tag{A55}$$

Two terms in Eq. (A54) can be combined with the help of this expansion, i.e.,

$$\frac{V(1-2f)\ln[n^n(1-n)^{1-n}]}{\sqrt{2\pi n(1-n)V}}\exp\left[-\frac{2\Delta N_{\rm crit}^2}{n(1-n)V}\right]$$
$$+\ln\left(\frac{n}{1-n}\right)\frac{2\Delta N_{\rm crit}}{\sqrt{2\pi(1-n)nV}}\exp\left[-\frac{2\Delta N_{\rm crit}^2}{n(1-n)V}\right]$$
$$=o(1), \tag{A56}$$

implying that we can omit these two terms. This yields the simplification

$$\langle S_A\rangle_N=-Vf\left[n\ln(n)+(1-n)\ln(1-n)\right]+\frac{f}{2}+\frac{\ln(1-f)}{2}$$
$$-\frac{V(1-2f)}{2}\ln[n^n(1-n)^{1-n}]\mathrm{erfc}\left[\sqrt{\frac{2}{n(1-n)V}}\Delta N_{\rm crit}\right]$$
$$+\sqrt{\frac{(1-n)nV}{2\pi}}\ln\left(\frac{n}{1-n}\right)\exp\left[-\frac{2\Delta N_{\rm crit}^2}{n(1-n)V}\right]$$
$$-2^{V(2f-1)-2}\exp\left[\frac{V(1-2n)^2}{2}\right]\mathrm{erfc}\left[\sqrt{V}\frac{(1-2n)^2-\ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right]$$
$$-2^{V(1-2f)-2}\exp\left[\frac{V(1-2n)^2}{2}\right]\mathrm{erfc}\left[\sqrt{V}\frac{(1-2n)^2+\ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right]+o(1). \tag{A57}$$

The two terms in the second and third lines can give rise to additional order-one terms from the fact that $\Delta N_{\rm crit}$ has an order-one distance to the smallest integer. To see whether this is indeed the case, we split $\Delta N_{\rm crit}=\Delta N_{\rm crit}^{(1)}+\Delta N_{\rm crit}^{(2)}$ with

$$\Delta N_{\rm crit}^{(1)}=-\frac{V(1-2f)\ln[n^n(1-n)^{1-n}]}{2\ln[(1-n)/n]}. \tag{A58}$$

We need to expand

$$\mathrm{erfc}\left[\sqrt{\frac{2}{n(1-n)V}}\Delta N_{\rm crit}\right]$$
$$=\mathrm{erfc}\left[\sqrt{\frac{2}{n(1-n)V}}\Delta N_{\rm crit}^{(1)}\right]-\sqrt{\frac{8}{\pi n(1-n)V}}\exp\left[-\frac{2(\Delta N_{\rm crit}^{(1)})^2}{n(1-n)V}\right]\Delta N_{\rm crit}^{(2)}+O(V^{-1}) \tag{A59}$$

and

$$\exp\left[-\frac{2\Delta N_{\rm crit}^2}{n(1-n)V}\right] = \exp\left[-\frac{2(\Delta N_{\rm crit}^{(1)})^2}{n(1-n)V}\right] - \frac{4\Delta N_{\rm crit}^{(1)}}{n(1-n)V}\exp\left[-\frac{2(\Delta N_{\rm crit}^{(1)})^2}{n(1-n)V}\right]\Delta N_{\rm crit}^{(2)} + O(V^{-1}). \tag{A60}$$

Plugging in $\Delta N_{\rm crit}^{(1)}$, we note that the leading-order correction due to a nonzero $\Delta N_{\rm crit}^{(2)}$ vanishes, and the next order is of order $1/\sqrt{V}$.

Thence, the final result for the average entanglement entropy reads

$$\begin{aligned}\langle S_A\rangle_N = &-Vf\left[n\ln(n)+(1-n)\ln(1-n)\right]+\frac{f}{2}+\frac{\ln(1-f)}{2}\\ &-\frac{V(1-2f)}{2}\ln[n^n(1-n)^{1-n}]\operatorname{erfc}\left[\sqrt{\frac{2}{n(1-n)V}}\Delta N_{\rm crit}^{(1)}\right]\\ &+\sqrt{\frac{(1-n)nV}{2\pi}}\ln\left(\frac{n}{1-n}\right)\exp\left[-\frac{2(\Delta N_{\rm crit}^{(1)})^2}{n(1-n)V}\right]\\ &-2^{V(2f-1)-2}\exp\left[\frac{V(1-2n)^2}{2}\right]\operatorname{erfc}\left[\sqrt{V}\frac{(1-2n)^2-\ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right]\\ &-2^{V(1-2f)-2}\exp\left[\frac{V(1-2n)^2}{2}\right]\operatorname{erfc}\left[\sqrt{V}\frac{(1-2n)^2+\ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right]+o(1).\end{aligned} \tag{A61}$$

When $n>\frac{1}{2}$ and/or $f>\frac{1}{2}$, we need to apply the symmetries and reflect $(f,n)\leftrightarrow(1-f,n)\leftrightarrow(f,1-n)\leftrightarrow(1-f,1-f)$.

Note that the result in Eq. (A61) naturally holds everywhere as long as $f, n, (1-n), (1-f)$ stay of order one. Only at these boundaries are there significant deviations. When, additionally, $1-2f$ and/or $1-2n$ are of order one, the last four terms vanish naturally as they are exponentially suppressed then.

At fixed f and n, we therefore find that

$$\begin{aligned}\langle S_A\rangle_N = &-Vf\left[n\ln(n)+(1-n)\ln(1-n)\right]+\frac{f+\ln(1-f)}{2}\\ &-\sqrt{\frac{n(1-n)}{2\pi}}\left|\ln\frac{1-n}{n}\right|\delta_{f,1/2}\sqrt{V}-\frac{1}{2}\delta_{f,1/2}\delta_{n,1/2}+o(1),\end{aligned} \tag{A62}$$

as already derived in the main text. Thus, there are discontinuities at $f=\frac{1}{2}$ and $f=n=\frac{1}{2}$, which can be further resolved with the help of Eq. (A61), as discussed next.

f. Resolving the critical regimes

In the main text, we found that the formula at fixed $f, n\in(0,\frac{1}{2})$ comprises additional terms at the line $f=\frac{1}{2}$ as well as at the multicritical point $f=n=\frac{1}{2}$. With the help of Eq. (A61), we can resolve these critical points by considering double scaling limits as follows:

$$\text{zooming in at } f=n=\frac{1}{2}: \quad f=\frac{1}{2}+\frac{\Lambda_f}{V}, \quad n=\frac{1}{2}+\frac{\Lambda_{\tilde n}}{\sqrt{V}}, \tag{A63}$$

$$\text{zooming in at } f=\frac{1}{2}: \qquad f=\frac{1}{2}+\frac{\Lambda_f}{\sqrt{V}}. \tag{A64}$$

Here the relevant scales (i.e., powers of V) are determined based on the discussion in the previous section. When choosing different dependencies in V for the deviations Λ_f and $\Lambda_{\tilde n}$, either create the Kronecker deltas in Eq. (A62) (higher powers

in V) or the terms take their limit for vanishing deviations (lower powers in V). Actually, the multicritical point $f = n = \frac{1}{2}$ is more subtle. While the last two terms in Eq. (A61) scale like those shown in Eq. (A63), the two terms in the second line of Eq. (A61) are still resolved when keeping $\sqrt{V}(1-2f)/(1-2n)$ of order one. We do not discuss this subtlety in the present subsection.

Critical line at $f = \frac{1}{2}$ *and* $n < \frac{1}{2}$. It is important to resolve the Kronecker delta about $f = \frac{1}{2}$ for fixed n. In this case, we need to expand Eq. (A61) with $f = \frac{1}{2} + \Lambda_f/\sqrt{V}$, where we need to take into account the fact that Eq. (A61) is only valid for $\Lambda_f < 0$ because we assumed that $f, n < \frac{1}{2}$ to derive this formula. The last two terms become exponentially suppressed and can be omitted. Thus, the expression simplifies to

$$\begin{aligned}\langle S_A\rangle_N = &-\frac{V}{2}[n\ln(n)+(1-n)\ln(1-n)] + \sqrt{V}[n\ln(n)+(1-n)\ln(1-n)]|\Lambda_f| + \frac{1}{4} - \frac{\ln(2)}{2}\\ &- \sqrt{V}\Bigg[|\Lambda_f|\ln[n^n(1-n)^{1-n}]\operatorname{erfc}\Bigg(-\frac{\sqrt{2}|\Lambda_f|\ln[n^n(1-n)^{1-n}]}{\sqrt{n(1-n)}|\ln[(1-n)/n]|}\Bigg)\\ &+\sqrt{\frac{n(1-n)}{2\pi}}\left|\ln\left[\frac{1-n}{n}\right]\right|\exp\Bigg(-\frac{2\Lambda_f^2\ln^2[n^n(1-n)^{1-n}]}{(1-n)n\ln^2[(1-n)/n]}\Bigg)\Bigg].\end{aligned} \tag{A65}$$

The terms of the first line agree with expansion (A62) around $f = \frac{1}{2} + \Lambda_f/\sqrt{V}$, while the second line resolves the Kronecker delta describing the subleading term $-b\sqrt{V}$ given by

$$\begin{aligned}b = &|\Lambda_f|\ln[n^n(1-n)^{1-n}]\operatorname{erfc}\Bigg(-\frac{\sqrt{2}|\Lambda_f|\ln[n^n(1-n)^{1-n}]}{\sqrt{n(1-n)}|\ln[(1-n)/n]|}\Bigg)\\ &+\sqrt{\frac{n(1-n)}{2\pi}}\left|\ln\left[\frac{1-n}{n}\right]\right|\exp\Bigg(-\frac{2\Lambda_f^2\ln^2[n^n(1-n)^{1-n}]}{(1-n)n\ln^2[(1-n)/n]}\Bigg).\end{aligned} \tag{A66}$$

This term is visualized in Fig. 6(d). Both equations above are already written in such a way that they hold for $f, n \in (0,1)$, meaning that one can also plug in values with $n > \frac{1}{2}$ and $\Lambda_f > 0$.

Multicritical point at $f = n = \frac{1}{2}$. When expanding the final result (A61) about $f = n = \frac{1}{2}$, where we assume that $f = \frac{1}{2} + \Lambda_f/V$ and $n = \frac{1}{2} + \Lambda_{\bar{n}}/\sqrt{V}$, we resolve the Kronecker delta at the point $f = n = \frac{1}{2}$; see Eq. (A62). Now all terms in Eq. (A61) become important and need to be taken into account:

$$\begin{aligned}\langle S_A\rangle_N = &\frac{\ln(2)}{2}V - \Lambda_{\bar{n}}^2 - \sqrt{\frac{2}{\pi}}|\Lambda_{\bar{n}}|\exp\Bigg(-\frac{\Lambda_f^2\ln^2(2)}{2\Lambda_{\bar{n}}^2}\Bigg) - |\Lambda_f|\ln(2)\operatorname{erf}\Bigg(\frac{|\Lambda_f|\ln(2)}{\sqrt{2}|\Lambda_{\bar{n}}|}\Bigg)\\ &-\frac{1}{4}\Bigg(\ln 4 - 1 + e^{2\Lambda_{\bar{n}}^2}\Bigg[4^{\Lambda_f}\operatorname{erfc}\Bigg(\frac{2\Lambda_{\bar{n}}^2+\Lambda_f\ln 2}{\sqrt{2}|\Lambda_{\bar{n}}|}\Bigg) + 4^{-\Lambda_f}\operatorname{erfc}\Bigg(\frac{2\Lambda_{\bar{n}}^2-\Lambda_f\ln 2}{\sqrt{2}|\Lambda_{\bar{n}}|}\Bigg)\Bigg]\Bigg)\\ &+o(1).\end{aligned} \tag{A67}$$

The constant terms of the first line are the result of expanding higher-order terms in n and f around $n = f = \frac{1}{2}$. The second and third lines then contain the resolution of the Kronecker delta at $f = n = \frac{1}{2}$ plus the offset $(\ln 4 - 1)/4$. The offset is not the result of higher-order terms, which is why we have included it in the negative term $-c$ that is of constant order in V. This constant is given by

$$c = \frac{1}{4}\Bigg(\ln 4 - 1 + e^{2\Lambda_{\bar{n}}^2}\Bigg[4^{\Lambda_f}\operatorname{erfc}\Bigg(\frac{2\Lambda_{\bar{n}}^2+\Lambda_f\ln 2}{\sqrt{2}|\Lambda_{\bar{n}}|}\Bigg) + 4^{-\Lambda_f}\operatorname{erfc}\Bigg(\frac{2\Lambda_{\bar{n}}^2-\Lambda_f\ln 2}{\sqrt{2}|\Lambda_{\bar{n}}|}\Bigg)\Bigg]\Bigg). \tag{A68}$$

We plotted this term in Fig. 6(e). Let us mention that only the last two terms in this constant create the Kronecker delta at $f = n = \frac{1}{2}$; cf. Eq. (A62).

2. Weighted average over sectors

Finally, we turn to the average over different sectors of fixed particle number $N = 0, \ldots, V$, which is weighted by $\exp[-wN]$. Explicitly, this means that we need to compute

$$\begin{aligned}\langle S_A\rangle_w &= \frac{1}{(1+e^{-w})^V}\sum_{N=0}^{V}\frac{V!}{N!\,(V-N)!}e^{-wN}\ \langle S_A\rangle_N \\ &= \frac{1}{(1+e^{-w})^V}\sum_{N=0}^{V/2}\frac{V!}{N!\,(V-N)!}e^{-wN}\,\langle S_A\rangle_N \\ &\quad+\frac{1}{(1+e^{-w})^V}\sum_{N=0}^{V/2}\frac{V!}{N!\,(V-N)!}e^{w(N-V)}\ \langle S_A\rangle_N\,. \end{aligned} \tag{A69}$$

The second identity follows from the particle-hole symmetry $\langle S_A\rangle_N = \langle S_A\rangle_{V-N}$. It is useful to write the sum in this form since Eq. (A61) is only applicable for $N = nV < V/2$. It becomes symmetric in $n \leftrightarrow 1-n$ when replacing all terms in $\ln[n/(1-n)]$ by $-|\ln[(1-n)/n]|$ and in the last two terms $(1-2n) \to |1-2n|$. All the other terms are already symmetric. Because of this symmetry, we may also assume that w is positive, meaning that the average particle number is smaller than $\frac{1}{2}$.

The average of the first four terms in Eq. (A61) does not involve an exponential behavior in n. Thus, it is suitable to expand n about its mean, which is $\langle n\rangle = \langle N\rangle/V = 1/(1+e^{w}) + \delta n$ in the first sum and $\langle n\rangle = \langle N\rangle/V = 1/(1+e^{-w}) + \delta n$ in the second one, with $\delta n = O(1/\sqrt{V})$. The order of this deviation tells us when the second sum plays a role, namely, when $w = O(1/\sqrt{V})$. Otherwise, it is exponentially suppressed due to the Gaussian tail of the large-V approximation of the binomial weight. Indeed, this Gaussian would also exponentially suppress the last two terms in Eq. (A61) when $w \gg 1/\sqrt{V}$, as those only contribute when $1-2n$ is of order $1/\sqrt{V}$ or smaller.

Hence, for $w \gg 1/\sqrt{V}$, the average entanglement entropy becomes

$$\begin{aligned}\langle S_A\rangle_w &= \frac{1}{(1+e^{-w})^V}\sum_{N=0}^{V/2}\frac{V!}{N!\,(V-N)!}e^{-wN}\,\langle S_A\rangle_N + o(1) \\ &= \frac{1}{(1+e^{-w})^V}\sum_{N=0}^{V/2}\frac{V!}{N!\,(V-N)!}e^{-wN} \\ &\quad\times\left(-Vf\,[n\ln(n)+(1-n)\ln(1-n)] + \frac{f}{2} + \frac{\ln(1-f)}{2}\right. \\ &\quad-\frac{V(1-2f)}{2}\ln[n^n(1-n)^{1-n}]\mathrm{erfc}\left[\sqrt{\frac{2}{n(1-n)V}}\Delta N^{(1)}_{\mathrm{crit}}\right] \\ &\quad\left.+\sqrt{\frac{(1-n)nV}{2\pi}}\ln\left(\frac{n}{1-n}\right)\exp\left[-\frac{2(\Delta N^{(1)}_{\mathrm{crit}})^2}{n(1-n)V}\right]\right)+o(1) \\ &= \langle S_A\rangle_{N=\bar N} - \frac{f}{2V\langle n\rangle(1-\langle n\rangle)}\frac{1}{(1+e^{-w})^V}\sum_{N=0}^{V}\frac{V!}{N!\,(V-N)!}e^{-wN}(N-\langle N\rangle)^2 + o(1) \\ &= \langle S_A\rangle_{N=\bar N} - \frac{f}{2} + o(1)\end{aligned} \tag{A70}$$

with $\bar N = 1/(1+e^{w})$. The $-f/2$ results from the very first term in Eq. (A61), as it is the average of the second-order term in the Taylor expansion in $(n-\langle n\rangle)$. The averages of the first-order terms vanish because of the centered Gaussian approximation of the binomial weight. This is also why the other terms do not contribute as the second-order Taylor terms will be of order $O(1/\sqrt{V})$ or smaller.

The calculation is more complicated for $w = O(1/\sqrt{V})$ or smaller. Then N is concentrated about $V/2$ and its difference is of order $\sqrt{V}$. The question is where we get new contributions in Eq. (A61). Certainly, this can only happen when $1-2f$

is of order $1/V$ or smaller as otherwise the maximum is always sufficiently away from the multicritical point $f = n = \frac{1}{2}$. We go step by step through the single terms.

The terms in the first line of Eq. (A61) are symmetric and smooth about $n = \frac{1}{2}$, so that the Taylor expansion and average will be exactly the same as in the case for $w \gg 1/\sqrt{V}$. The only thing to consider is that $\langle N\rangle = \bar{N} = V\bar{n}$ is not exactly at $V/2$, but $\frac{1}{2} - \bar{n}$ is of order $1/\sqrt{V}$. Therefore, we have an expansion about $N = V/2$, yielding

$$\langle S_A\rangle_{N=\bar{N}} = \langle S_A\rangle_{N=V/2} - V\left(\bar{n} - \frac{1}{2}\right)^2 + O(V^{-1}). \tag{A71}$$

The term in the second line of Eq. (A61) is of order $V(1-2f) = O(1)$ or smaller, so that any Taylor expansion about $n = \frac{1}{2}$ leads to terms that are vanishing in the large-V limit. The term in the third line of Eq. (A61) yields an additional integral as the leading order vanishes and the first-order Taylor expansion gives a term of order one. This integral is

$$\begin{aligned}
&\left\langle \sqrt{\frac{(1-n)nV}{2\pi}} \left| \ln\left(\frac{n}{1-n}\right)\right| \exp\left[-\frac{2(\Delta N^{(1)}_{\rm crit})^2}{n(1-n)V}\right]\right\rangle_N \\
&\quad = 4e^{-Vw^2/8}\int_0^\infty \frac{d\delta n}{\sqrt{2\pi}} \exp[-2\delta n^2]\cosh[\sqrt{V}w\delta n]\sqrt{\frac{2}{\pi}}\delta n \exp\left[-\frac{(\ln 2)^2V^2(1-2f)^2}{8\delta n^2}\right] \\
&\qquad + O\left(\frac{1}{\sqrt{V}}\right) \\
&\quad = \frac{4}{\pi}e^{-Vw^2/8}\int_0^\infty d\delta n \exp\left[-2\delta n^2 - \frac{(\ln 2)^2V^2(1-2f)^2}{8\delta n^2}\right]\cosh[\sqrt{V}w\delta n]\delta n + O\left(\frac{1}{\sqrt{V}}\right),
\end{aligned} \tag{A72}$$

because $\Delta N^{(1)}_{\rm crit} = \ln(2)V^{3/2}(1-2f)/(8\delta n)$ with $n = N/V = \frac{1}{2} - \delta n/\sqrt{V}$. We have not found a way to further simplify this integral, so we have evaluated it numerically.

Also, for the two last terms in Eq. (A61), we can go over to a Gaussian integral as any correction will be of vanishing order in $V \to \infty$ as the two terms are of order one or smaller. Hence, we have

$$\begin{aligned}
&\frac{1}{(1+e^{-w})^V}\sum_{N=0}^{V/2}\frac{V!}{N!\,(V-N)!}(e^{-wN} + e^{w(N-V)})2^{V(2f-1)-2}\exp\left[\frac{V(1-2n)^2}{2}\right] \\
&\quad \times \operatorname{erfc}\left[\sqrt{V}\frac{(1-2n)^2 - \ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right] \\
&\quad = 2^{V(2f-1)}e^{-Vw^2/8}\int_0^\infty \frac{d\delta n}{\sqrt{2\pi}}\cosh[\sqrt{V}w\delta n]\operatorname{erfc}\left[\frac{4\delta n^2 - \ln(2)V(1-2f)}{\sqrt{8}\delta n}\right] + O\left(\frac{1}{\sqrt{V}}\right)
\end{aligned} \tag{A73}$$

and, similarly,

$$\begin{aligned}
&\frac{1}{(1+e^{-w})^V}\sum_{N=0}^{V/2}\frac{V!}{N!\,(V-N)!}(e^{-wN} + e^{w(N-V)})2^{V(1-2f)-2}\exp\left[\frac{V(1-2n)^2}{2}\right] \\
&\quad \times \operatorname{erfc}\left[\sqrt{V}\frac{(1-2n)^2 + \ln(2)(1-2f)}{\sqrt{2}(1-2n)}\right] \\
&\quad = 2^{V(1-2f)}e^{-Vw^2/8}\int_0^\infty \frac{d\delta n}{\sqrt{2\pi}}\cosh[\sqrt{V}w\delta n]\operatorname{erfc}\left[\frac{4\delta n^2 + \ln(2)V(1-2f)}{\sqrt{8}\delta n}\right] + O\left(\frac{1}{\sqrt{V}}\right).
\end{aligned} \tag{A74}$$

We evaluate these integrals numerically as an analytical treatment seems to be out of reach.

In summary, the entanglement entropy averaged over the particle number N at $w = O(1/\sqrt{V})$ or smaller and for $1 - 2f = O(V^{-1})$ is

$$\begin{aligned}\langle S_A\rangle_w = \langle S_A\rangle_{\langle N\rangle=V/2} &- \frac{V(1-2\bar{n})^2}{4} - \frac{1}{4} \\ &- \frac{4}{\pi}e^{-Vw^2/8}\int_0^\infty d\delta n \exp\left[-2\delta n^2 - \frac{(\ln 2)^2V^2(1-2f)^2}{8\delta n^2}\right]\cosh[\sqrt{V}w\delta n]\delta n \\ &- 2^{V(2f-1)}e^{-Vw^2/8}\int_0^\infty \frac{d\delta n}{\sqrt{2\pi}}\cosh[\sqrt{V}w\delta n]\mathrm{erfc}\left[\frac{4\delta n^2 - \ln(2)V(1-2f)}{\sqrt{8}\delta n}\right] \\ &- 2^{V(1-2f)}e^{-Vw^2/8}\int_0^\infty \frac{d\delta n}{\sqrt{2\pi}}\cosh[\sqrt{V}w\delta n]\mathrm{erfc}\left[\frac{4\delta n^2 + \ln(2)V(1-2f)}{\sqrt{8}\delta n}\right] + O\left(\frac{1}{\sqrt{V}}\right). \end{aligned} \tag{A75}$$

Let us stress that

$$\begin{aligned}\langle S_A\rangle_{\langle N\rangle=V/2} &= \ln(2)Vf + \frac{1}{4} - \frac{\ln(2)}{2} + \ln(2)V(1-2f) - 2^{V(2f-1)-1} + O\left(\frac{1}{\sqrt{V}}\right) \\ &= \ln(2)V(1-f) + \frac{1-\ln(4)}{4} - 2^{V(2f-1)-1} + O\left(\frac{1}{\sqrt{V}}\right) \end{aligned} \tag{A76}$$

for $f \leq \frac{1}{2}$.

At last, we would like to consider the particular case $f = \frac{1}{2}$ and $w = 0$ that corresponds to a multicritical point, too. In this case Eq. (A75) simplifies drastically. Indeed, the last two integrals are both equal to

$$\int_0^\infty \frac{d\delta n}{\sqrt{2\pi}}\mathrm{erfc}[\sqrt{2}\delta n] = \frac{1}{2\pi}. \tag{A77}$$

Also, the third, remaining integral can be carried out as

$$\frac{4}{\pi}\int_0^\infty d\delta n e^{-2\delta n^2}\delta n = \frac{1}{\pi}. \tag{A78}$$

Collecting everything we find that, at $w = O(1/\sqrt{V})$ or smaller,

$$\langle S_A\rangle_w = \langle S_A\rangle_{\langle N\rangle=V/2} - \frac{1}{4} - \frac{2}{\pi}. \tag{A79}$$

This result can be combined with Eq. (A71) to get Eq. (67), when choosing f and $\bar{n}$ of order one and fixed without double scaling. Then, the last term $-2/\pi$ only appears when $f = \bar{n} = \frac{1}{2}$, resulting in a Kronecker delta.

APPENDIX B: INTEGRATION FORMULA FOR JACOBI POLYNOMIALS

When considering the average and the variance of the entanglement entropy for ensembles of fermionic Gaussian states, we routinely encounter expressions involving integrals of Jacobi polynomials given by

$$I_{kl}^{(\alpha_1,\beta_1,\alpha_2,\beta_2)} = \left[\frac{d}{d\epsilon}\int_{-1}^{1}(1-x)^{\alpha_2+\epsilon}(1+x)^{\beta_2}\mathcal{P}_k^{(\alpha_1,\beta_1)}\mathcal{P}_l^{(\alpha_2,\beta_2)}\right]_{\epsilon\to 1}. \tag{B1}$$

In order to evaluate this expression, we write the Jacobi polynomials using one of their two representations:

$$\mathcal{P}_k^{(\alpha_1,\beta_1)}(x) = \sum_{m=0}^{k}\underbrace{\frac{(\alpha_1+k)!}{k!\,(\alpha_1+\beta_1+k)!}\binom{k}{m}\frac{(\alpha_1+\beta_1+k+m)!}{(\alpha_1+m)!}\left(-\frac{1}{2}\right)^m}_{=X_m^{(\alpha_1,\beta_1,k)}}(1-x)^m \tag{B2}$$

for the sum representation and

$$\mathcal{P}_l^{(\alpha_2,\beta_2)}(x) = \underbrace{\frac{(-1)^l}{2^l l!}}_{=Y^{(l)}} \frac{1}{(1-x)^{\alpha_2}(1+x)^{\beta_2}} \partial_x^l [(1-x)^{\alpha_2+l}(1+x)^{\beta_2+l}] \tag{B3}$$

for the Rodrigues formula. Note that it is important that $(1-x)^{\alpha_2}(1+x)^{\beta_2}$ in Eq. (B1) matches those appearing in $\mathcal{P}_l^{\alpha_2,\beta_2)}(x)$. The appearing integrals can be evaluated by first an integration by parts and then exploiting the beta function integral [Eq. (5.12.1) of Ref. [164]], which can be identified when substituting $x = 2t - 1$ and $t \in [0, 1]$,

$$\begin{aligned} T_{\alpha_2,\beta_2,m,l}(\epsilon) &= \int_{-1}^{1} (1-x)^{m+\epsilon} \partial_x^l [(1-x)^{\alpha_2+l}(1+x)^{\beta_2+l}] dx \\ &= 2^{\alpha_2+\beta_2+m+\epsilon+l+1} \frac{\Gamma(\alpha_2+m+\epsilon+1)\Gamma(\beta_2+l+1)\Gamma(m+\epsilon+1)}{\Gamma(\alpha_2+\beta_2+m+\epsilon+l+2)} \\ &\quad \times \frac{1}{\Gamma(m-l+\epsilon+1)}, \end{aligned} \tag{B4}$$

where $\Gamma(z)$ is the gamma function. Note that this function has removable singularities at integers when $m \leq l-2$ in the limit $\epsilon \to 1$. When $0 \leq m \leq l-2$, we can use Euler's reflection formula [Eq. (5.5.3) of Ref. [164]],

$$\begin{aligned} &\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)} \\ \iff \quad &\frac{1}{\Gamma(m-l+\epsilon+1)} = \frac{\sin\pi(m-l+\epsilon+1)}{\pi} \Gamma[1-(m-l+\epsilon+1)], \end{aligned} \tag{B5}$$

to remove those singularities, such that

$$T_{\alpha_2,\beta_2,m,l} = \begin{cases} 2^{\alpha_2+\beta_2+m+\epsilon+l+1} \dfrac{\Gamma(\alpha_2+m+\epsilon+1)\Gamma(\beta_2+l+1)}{\Gamma(\alpha_2+\beta_2+m+\epsilon+l+2)\Gamma(m+\epsilon+1)} \\ \times \dfrac{\sin\pi(m-l+\epsilon+1)\Gamma(l-m-\epsilon)}{\pi}, & m \leq l-2, \\ 2^{\alpha_2+\beta_2+m+\epsilon+l+1} \dfrac{\Gamma(\alpha_2+m+\epsilon+1)\Gamma(\beta_2+l+1)}{\Gamma(\alpha_2+\beta_2+m+\epsilon+l+2)\Gamma(m+\epsilon+1)} \\ \times \dfrac{1}{\Gamma(m-l+\epsilon+1)}, & m \geq l-1. \end{cases} \tag{B6}$$

We can evaluate the derivative in ϵ at $\epsilon = 1$ analytically to find that

$$\begin{aligned} &T'_{\alpha_2,\beta_2,m,l}(1) \\ &= \begin{cases} \frac{2^{2+m+l+\alpha_2+\beta_2}(m+1)!(l+\beta_2)!(m+\alpha_2+1)!(l-m-2)!}{(\alpha_2+\beta_2+l+m+2)!}(-1)^{m+l} & \text{for } m \leq l-2, \\ \frac{2^{2+m+l+\alpha_2+\beta_2}(m+1)!(l+\beta_2)!(m+\alpha_2+1)![\ln 2+\Psi(2+m)-\Psi(2+m-l)+\Psi(2+m+\alpha_2)-\Psi(3+m+l+\alpha_2+\beta_2)]}{(\alpha_2+\beta_2+l+m+2)!(m-l+1)!} \\ \quad \text{for } m \geq l-1, \end{cases} \end{aligned} \tag{B7}$$

for the respective integer cases of m, where the divergences have been removed accordingly. This yields

$$I_{kl}^{(\alpha_1,\beta_1,\alpha_2,\beta_2)} = Y^l \sum_{m=0}^{k} \underbrace{X_m^{(\alpha_1,\beta_1,k)} T'_{\alpha_2,\beta_2,m,l}(1)}_{F}. \tag{B8}$$

The summand F for $m \leq l-2$ can be written as

$$\begin{aligned} F &= (-1)^l 2^{2+l+\alpha_2+\beta_2} \frac{k!\,(k+\alpha_1)!\,(l+\beta_2)!}{k!} \frac{(1+m)!}{m!} \frac{(l-m-2)!}{(k-m)!} \frac{(1+m+\alpha_2)!}{(m+\alpha_1)!} \\ &\quad \times \frac{(k+m+\alpha_1+\beta_1)!}{(2+l+m+\alpha_2+\beta_2)!} \\ &= (-1)^l 2^{2+l+\alpha_2+\beta_2} (k+\alpha_1)!\,(l+\beta_2)!\,(1+m) \frac{(l-m-2)!}{(k-m)!} \frac{(1+m+\alpha_2)!}{(m+\alpha_1)!} \\ &\quad \times \frac{(k+m+\alpha_1+\beta_1)!}{(2+l+m+\alpha_2+\beta_2)!}. \end{aligned} \tag{B9}$$

The difficulty lies in the evaluation of this sum, which simplifies drastically for suitable k and l, as well as $\alpha_1+\beta_1$ and $\alpha_2+\beta_2$. While there may exist a method to evaluate Eq. (B8) for general values, it will suffice for our purpose to compute for the specific situation we encounter.

APPENDIX C: PURE FERMIONIC GAUSSIAN STATES

1. Average entanglement entropy

We start with the one-point function $R_1(x)$, normalized to $\int_0^1 R_1(x)dx = V_A$, which can be shown [136,159] to be given by

$$R_1(x) = \sum_{n=0}^{V_A-1} \psi_n^2(x) = (1-x)^\Delta \sum_{n=0}^{V_A-1} \frac{1}{c_n} [\mathcal{P}_{2n}^{(\Delta,\Delta)}(x)]^2, \tag{C1}$$

where $\Delta = V - 2V_A$, $\mathcal{P}_n^{(\alpha,\beta)}$ refers to the Jacobi polynomials, and the constant c_n is given by

$$c_n = \frac{2^{2\Delta}[(2n+\Delta)!]^2}{(2n)!\,(2n+2\Delta)!\,(4n+2\Delta+1)}. \tag{C2}$$

An important property of orthogonal polynomials is that they satisfy the Christoffel-Darboux relation [Eq. (18.2.12) of Ref. [164]] that expresses the above sum in terms of the polynomials and their first derivatives of the highest orders. This is useful for analyzing the asymptotic behavior. This Christoffel-Darboux formula, for a general set of orthogonal polynomials $p_n(x)$ with normalization c_n, reads

$$\sum_{n=0}^{V_A-1} \frac{p_n^2(x)}{c_n} = \frac{k_{V_A-1}}{c_{V_A-1} k_{V_A}} [p'_{V_A}(x) p_{V_A-1}(x) - p'_{V_A-1}(x) p_{V_A}(x)], \tag{C3}$$

where the coefficients k_n are the leading-order coefficients in $p_n(x)$. In the present case of the Jacobi polynomials, we can exploit an additional recurrence relation [Eq. (18.9.15) of Ref. [164]], and the known coefficient k_n of the highest power in the Jacobi polynomials,

$$\frac{d}{dx}\mathcal{P}_n^{(\alpha,\beta)}(x) = \frac{\alpha+\beta+n+1}{2} \mathcal{P}_{n-1}^{(\alpha+1,\beta+1)}(x) \quad \text{and} \quad k_n = \frac{(2n+\alpha+\beta)!}{2^n (n+\alpha+\beta)!\,n!}. \tag{C4}$$

This allows us to write $R_1(x)$ as

$$R_1(x) = V_A[A_1 F_1(x) - A_2 F_2(x)], \tag{C5}$$

where we have introduced the abbreviations

$$A_1 = \frac{k_{2V_A-2}}{c_{2V_A-2} k_{2V_A}} \frac{2\Delta+2V_A+1}{2}, \quad F_1(x) = (1-x^2)^\Delta \mathcal{P}_{2V_A-1}^{(\Delta+1,\Delta+1)}(x) \mathcal{P}_{2V_A-2}^{(\Delta,\Delta)}(x), \tag{C6}$$

$$A_2 = \frac{k_{2V_A-2}}{c_{2V_A-2}k_{2V_A}}\frac{2\Delta+2V_A-1}{2}, \quad F_2(x) = (1-x^2)^{\Delta}\mathcal{P}^{(\Delta+1,\Delta+1)}_{2V_A-3}(x)\mathcal{P}^{(\Delta,\Delta)}_{2V_A}(x). \tag{C7}$$

We make use of this simplified expression for the level density to compute the average entanglement entropy $\langle S_A\rangle_{G,N} = \int_0^1 R_1(x)s(x)dx$ for Gaussian states, where the empirical entanglement entropy in terms of the eigenvalues x is

$$s(x) = -\left(\frac{1+x}{2}\right)\ln\left(\frac{1+x}{2}\right) - \left(\frac{1-x}{2}\right)\ln\left(\frac{1-x}{2}\right) = \ln 2 - \frac{1}{2}\partial_\epsilon[(1-x)^\epsilon + (1+x)^\epsilon]_{\epsilon=1}, \tag{C8}$$

where $s(x)$ was introduced in Eq. (78). The trick with generating the logarithm by a derivative is a standard one and it is related to the replica trick. The advantage is that we can understand the factors $(1\pm x)^\epsilon$ as a tractable deformation of the original weight $(1-x^2)^\Delta = (1-x)^\Delta(1+x)^\Delta$.

Because of the normalization of the one-point function $\int_0^1 R_1(x)dx = V_A$ and the symmetry $s(x) = s(-x)$, the average becomes

$$\langle S_A\rangle_{G,N} = V_A[\ln 2 - A_1 I_1 + A_2 I_2], \tag{C9}$$

where we have introduced the integrals

$$I_1 = I^{(\Delta+1,\Delta+1,\Delta,\Delta)}_{(2V_A-1),(2V_A-2)}, \quad I_2 = I^{(\Delta+1,\Delta+1,\Delta,\Delta)}_{(2V_A-3),2V_A}, \tag{C10}$$

where $I^{(\alpha_1,\beta_1,\alpha_2,\beta_2)}_{kl}$ was introduced in Eq. (B8) and reduced to a sum in Eq. (B8). Note that the factors of $\frac{1}{2}$ in Eq. (C8) were canceled by the symmetrizing over $x \to -x$. Both sums can be performed analytically, leading to the full expression given by (89).

2. Variance

In Eq. (92), we have seen that the variance of the entanglement entropy can be expressed in terms of the integrals

$$\begin{aligned} s_{ij} &= \int_{-1}^{1} s(x)\psi_i(x)\psi_j(x)dx \\ &= -\frac{1}{2}\frac{d}{d\epsilon}\left[\int_{-1}^{1}[(1+x)^\epsilon + (1-x)^\epsilon]\psi_i(x)\psi_j(x)\right]_{\epsilon=1} \\ &= -I^{(\Delta,\Delta,\Delta,\Delta)}_{ij}, \end{aligned} \tag{C11}$$

where $\psi_i(x)$ and $s(x)$ were introduced in Eqs. (88) and (78), respectively, and we used the fact that $\psi_i(x) = \psi_i(-x)$. For $i<j$,

$$s^2_{ij} = \frac{(2j)!(2\Delta+4i+1)(\Delta+j+1)(2\Delta+2j+1)(2\Delta+4j+1)[2(\Delta+i)]![(1+\Delta-2\Delta^2)i-2(\Delta-1)i^2+(\Delta+1)(2j+1)(\Delta+j)]^2}{2(2i)!(2i-2j+1)^2(i-j)^2(-2i+2j+1)^2[2(\Delta+j+1)]!(\Delta+i+j)^2(\Delta+i+j+1)^2(2\Delta+2i+2j+1)^2}, \tag{C12}$$

where Eq. (C12) is only valid for $i<j$, which is all we need for the sum in Eq. (92). Despite all terms in the sum of Eq. (92) being nonzero for large N, it is dominated by the summand $s^2_{N_A-1,N_A}$, so that it makes sense to consider the limit

$$\bar{s}^2_{lk} = \lim_{N\to\infty} s^2_{N_A-1-l,N_A+k} = \frac{(1/f\ -1)^{-2(k+l+1)}[2k+2l+3-4f\ (k+l+1)]^2}{4(k+l+1)^2(2k+2l+1)^2(2k+2l+3)^2} \tag{C13}$$

with fixed $f\ = V_A/V$. Summing over Eq. (C13) then yields Eq. (93).

APPENDIX D: PURE FERMIONIC GAUSSIAN STATES WITH A FIXED NUMBER OF PARTICLES

1. Average entanglement entropy

The idea behind computing the entanglement entropy for Gaussian states is based on a different random matrix average than that in the Page setting. Since a formula for the entanglement entropy à la Page is not at hand, we derive it in the present section. Let us briefly outline the strategy.

1. The goal is to compute $\langle S_A\rangle_{G,N} = \int_0^1 s(x)R_1(x)dx$ as explained in the main text.
2. We can express $R_1(x)$ as a sum of Jacobi polynomials and $s(x)$ as derivatives in ϵ of powers $(1\pm x)^\epsilon$ evaluated at $\epsilon = 1$. This is a crucial trick that allows us to deal with the logarithm in $s(x)$.
3. In order to simplify the integral, we first apply the Christoffel-Darboux formula (turning the sum of V_A terms into a sum over two terms) and then express one Jacobi polynomial as a sum and the other as derivatives via Rodrigues formula.
4. This gives $\langle S_A\rangle_{G,N}$ as a sum of known integrals (of which some need to be regularized) that can be eventually evaluated.

As already mentioned, we start with the level density R_1 from Eq. (110) that describes the eigenvalues of a truncated unitary matrix. It is given in terms of Jacobi polynomials and the associated weight $(1-x)^\alpha(1+x)^\beta$, with respect to which they form an orthonormal set. In particular, the level density is

$$R_1(x) = \sum_{n=0}^{V_A-1} \psi_n^2(x) = (1-x)^\alpha(1+x)^\beta \sum_{n=0}^{V_A-1} \frac{1}{c_n}[\mathcal{P}_n^{(\alpha,\beta)}(x)]^2 \tag{D1}$$

with $\psi_n(x)$ as in Eq. (109). An important property of orthogonal polynomials is that they satisfy the Christoffel-Darboux relation [Eq. (18.2.12) of Ref. [164]] that expresses the above sum in terms of the polynomials and their first derivatives of the highest orders. This is useful for analyzing the asymptotic behavior. This Christoffel-Darboux formula, for a general set of orthogonal polynomials $p_n(x)$ with normalization c_n, reads

$$\sum_{n=0}^{V_A-1} \frac{p_n^2(x)}{c_n} = \frac{k_{V_A-1}}{c_{V_A-1}k_{V_A}}[p'_{V_A}(x)p_{V_A-1}(x) - p'_{V_A-1}(x)p_{V_A}(x)], \tag{D2}$$

where the coefficients k_n are the leading-order coefficients in $p_n(x)$. In the present case, we can again use the known relations from Eq. (C4). Thus, $R_1(x)$ is expressed as

$$R_1(x) = V_A[A_1F_1(x) - A_2F_2(x)], \tag{D3}$$

where we have introduced the abbreviations

$$A_1 = \frac{k_{V_A-1}(\alpha+\beta+V_A+1)}{2V_Ac_{V_A-1}k_{V_A}}, \quad F_1(x) = (1-x)^\alpha(1+x)^\beta\mathcal{P}_{V_A-1}^{(\alpha+1,\beta+1)}(x)\mathcal{P}_{V_A-1}^{(\alpha,\beta)}(x), \tag{D4}$$

$$A_2 = \frac{k_{V_A-1}}{c_{V_A-1}k_{V_A}}\frac{\alpha+\beta+V_A}{2}, \quad F_2(x) = (1-x)^\alpha(1+x)^\beta\mathcal{P}_{V_A-2}^{(\alpha+1,\beta+1)}(x)\mathcal{P}_{V_A}^{(\alpha,\beta)}(x). \tag{D5}$$

We make use of this simplified expression for the level density to compute the average entanglement entropy $\langle S_A\rangle_{G,N} = \int_0^1 R_1(x)s(x)dx$ for Gaussian states, where the empirical entanglement entropy in terms of the eigenvalues x is

$$s(x) = -\left(\frac{1+x}{2}\right)\ln\left(\frac{1+x}{2}\right) - \left(\frac{1-x}{2}\right)\ln\left(\frac{1-x}{2}\right) = \ln 2 - \frac{1}{2}\partial_\epsilon[(1-x)^\epsilon + (1+x)^\epsilon]_{\epsilon=1}, \tag{D6}$$

where $s(x)$ was introduced in Eq. (78).

Because of the normalization, $\int_0^1 R_1(x)dx = V_A$ and the symmetry $R_1(x) = R_1(-x)$ by construction, the average reduces to

$$\langle S_A\rangle_{G,N} = V_A\big[\ln 2 - \tfrac{1}{2}A_1I_1 + \tfrac{1}{2}A_2I_2\big]. \tag{D7}$$

Here, we have introduced the two symmetrized integrals

$$I_1 = I^{(\alpha+1,\beta+1,\alpha,\beta)}_{V_A-1,V_A-1} + I^{(\beta+1,\alpha+1,\beta,\alpha)}_{V_A-1,V_A-1}, \qquad I_2 = I^{(\alpha+1,\beta+1,\alpha,\beta)}_{V_A-2,V_A} + I^{(\beta+1,\alpha+1,\beta,\alpha)}_{V_A-2,V_A}, \tag{D8}$$

where $I_{kl}^{(\alpha_1,\beta_1,\alpha_2,\beta_2)}$ was introduced in Eq. (B8) and reduced to a sum in Eq. (B8). Note that we used $\mathcal{P}_n^{(\alpha,\beta)}(-x) = (-x)^n\mathcal{P}_n^{(\beta,\alpha)}(x)$ to arrive at this symmetrization.

Only the last two terms in the sum need to be dealt with in the second case because those correspond to $m = n - 1$ and $m = n$ with $n = V_A - 1$. This leads to the following lengthy expression:

$$\begin{aligned} I_1 = &\frac{2^{2+\alpha+\beta}(2V_A+\alpha+\beta)(V_A+\alpha-1)!(V_A+\beta-1)!}{(V_A-1)!\,(V_A+\alpha+\beta+1)!} \\ &\times\Bigg[\sum_{m=0}^{V_A-3}\frac{(m+1)}{(V_A-m-2)(V_A-m-1)} \\ &+(\alpha+\beta+2V_A)(\ln 2 - V_A + \Psi(V_A) - \Psi(1) - \Psi(\alpha+\beta+2V_A)) \\ &+(V_A+\alpha)\Psi(V_A+\alpha)+(V_A+\beta)\Psi(V_A+\beta)+\alpha+\beta+3V_A\Bigg]. \end{aligned} \tag{D9}$$

The cumbersome looking sum can be in fact carried out exactly with the help of the following identity that is based on a series representation of the digamma function [Eq. (5.7.6) of Ref. [164]]:

$$\sum_{m=0}^{n}\frac{(m+1)}{(Z+m)(Z+m+1)} = \frac{Z}{1+Z+n} - 1 + \Psi(Z+n+1) - \Psi(Z). \tag{D10}$$

Thence, we arrive at

$$\partial_\epsilon I_1 = \frac{2^{V+2-2V_A}N!\,(V-N-1)!\,[V_A + V\ln 2 + N\Psi(N) - V\Psi(V) + (V-N)\Psi(V-N)]}{N(V_A-1)!\,(V-V_A+1)!}. \tag{D11}$$

In a very similar way, we can also evaluate I_2 as

$$\begin{aligned} I_2 &= \sum_{m=0}^{V_A-2} Y^{(V_A)}[X_m^{(\alpha+1,\beta+1,V_A-2)}T'_{\alpha,\beta,V_A,m}(1) + X_m^{(\beta+1,\alpha+1,V_A-2)}T'_{\beta,\alpha,V_A,m}(1)] \\ &= 2^{2+\alpha+\beta}\frac{(2V_A+\alpha+\beta)(V_A+\alpha-1)!\,(V_A+\beta-1)!}{V_A!\,(V_A+\alpha+\beta)!} \\ &\quad\times\sum_{m=0}^{V_A-2}\frac{(m+1)}{(V_A+\alpha+\beta+m+1)(V_A+\alpha+\beta+m+2)} \\ &= -\frac{2^{2+\alpha+\beta}\Gamma(V_A+\alpha)\Gamma(V_A+\beta)[V_A-1+(2V_A+\alpha+\beta)(\Psi(1+\alpha+\beta+V_A)-\Psi(2V_A+\alpha+\beta))]}{V_A!\,(V_A+\alpha+\beta)!} \\ &= -\frac{2^{V+2-2V_A}(N-1)!\,(V-N-1)!\,[V_A - 1 + T\Psi(V+1-V_A) - T\Psi(V)]}{V_A!\,(V-V_A)!}. \end{aligned} \tag{D12}$$

In this case, we have not needed to consider different cases of Eq. (B7). It always holds that $m \le n-2$ with $n = V_A - 2$. In the fourth line, we employed anew Eq. (D10) to express the sum in terms of digamma functions.

Finally, we combine both terms (D11) and (D12) according to Eq. (D7). After canceling the various factorials, we eventually arrive at the expression

$$\begin{aligned} \langle S_A\rangle_{G,N} = 1 &- \frac{V_A}{V}(1+V) + V\Psi(V) - \frac{V_A}{V}[(V-N)\Psi(V-N) + N\Psi(N)] \\ &+ (V_A - V)\Psi(V - V_A + 1), \end{aligned} \tag{D13}$$

quoted in the main text. This average has the particle-hole symmetry $N \leftrightarrow V - N$, as can be readily seen. Let us highlight that the symmetry between the number of particles N and system size V_A is not visible since the formula above only holds for $V_A \le N, V - N$. The symmetry is enforced by hand, where one needs to reflect $V_A \leftrightarrow N$ in the formula when $V_A > N$. In this way one also gets the mirror symmetry between two subsystems $A \leftrightarrow B$ reflected in $V_A \leftrightarrow V - V_A$. The latter symmetry and how it is introduced, namely by hand, is shared with Page's setting.

2. Variance

In Eq. (92), we have seen that the variance of the entanglement entropy can be expressed in terms of the integrals

$$\begin{aligned} s_{ij} &= \int_{-1}^{1} s(x)\psi_i(x)\psi_j(x)dx \\ &= -\frac{1}{2}\frac{d}{d\epsilon}\left[\int_{-1}^{1}[(1+x)^\epsilon + (1-x)^\epsilon]\psi_i(x)\psi_j(x)\right]_{\epsilon=1} \\ &= -\frac{1}{2}(I_{ij}^{(\alpha,\beta,\alpha,\beta)} + (-1)^{i+j} I_{ij}^{(\alpha,\beta,\alpha,\beta)}), \end{aligned} \tag{D14}$$

which are only needed for the indices $i < V_A \leq j$. Note that the integral in the second line is of the same form as that in Eq. (B1), which was reduced to a sum in Eq. (B8). Thus, we can express the result as the sum

$$I_{ij} = \sum_{m=0}^{i} Y^{(j)}[X_m^{(\alpha,\beta,i)} T'_{\alpha,\beta,m,j}(1) + (-1)^{i+j} X_m^{(\beta,\alpha,i)} T'_{\beta,\alpha,m,j}(1)], \tag{D15}$$

where Y, X, and T are defined in Eqs. (B2), (B3), and (B7), respectively. In order to compute the sum, we need to distinguish the cases $i+1 = j = V_A$ and $i+1 < j$ due to the case discussion in T. The resulting expressions are rather unwieldy, but we can simplify them by defining $i = V_A - 1 - l$ and $j = V_A + k$ and then taking the thermodynamic limit $V \to \infty$ to find that

$$\begin{aligned} \bar{s}_{00}^2 &= \lim_{V\to\infty} s^2_{V_A-1,V_A} = \frac{f\,[f - 2fn - 2(f-1)(n-1)n\,\ln[(1-n)/n]]^2}{(f-1)(n-1)n}, \\ \bar{s}_{lk}^2 &= \lim_{V\to\infty} s^2_{V_A-1-l,V_A+k} \end{aligned} \tag{D16}$$

$$= \frac{f^{1+k+l}([f(2+k+l-2n)+(2+k+l)(n-1)]n^{1+k+l} - (n-1)^{1+k+l}[(2+k+l)n+f(k+l+2n)])^2}{(f-1)(k+l)^2(1+k+l^2)^2(2+k+l)^2(n-fn-n^2+fn^2)^{1+k+l}}. \tag{D17}$$

Sum (92) is certainly a finite sum of generalized hypergeometric functions after taking this limit. What has been rather surprising for us is that it can be performed exactly, yielding the relatively simple expression

$$\begin{aligned} \lim_{V\to\infty}(\Delta S_A)^2_{G,N} &= \sum_{l,k=0}^{\infty} \bar{s}_{kl}^2 \\ &= \ln(1-f) + f + f^2 + f^2(2n-1)\ln\left(\frac{1-n}{n}\right) \\ &\quad + f(f-1)(n-1)n\ln^2\left(\frac{1-n}{n}\right), \end{aligned} \tag{D18}$$

which is the main result of this section.

While our representation of s_{ij} has not been suitable to perform the full sum for finite size V, the asymptotic result looks as if it is possible to compute the variance $(\Delta S_A)_{G,N}$ as an expression of digamma functions at fixed V, similar to $\langle S_A\rangle$ from Eq. (113). For this, it will likely be beneficial to find closed formulas for the inner product of different orthogonal polynomials along the lines of Ref. [149]. In fact, such methods have already been used to find a closed expression for the variance $(\Delta S_A)_G$ of the entanglement entropy for the ensemble of all Gaussian states [177].

a. Weighted average over sectors

As in the Page setting, we can average the entanglement entropy over all particle-number sectors with a binomial weight for Gaussian states, as well. Employing the same notation as in the Page case, the binomial weight

$$\varrho_w(N) = \binom{V}{N} p^N q^{V-N} = \binom{V}{N} \frac{e^{-Nw}}{(1+e^{-w})^V} \quad \text{with} \quad p = \frac{e^{-w}}{1+e^{-w}} \quad \text{and} \quad q = 1-p = \frac{1}{1+e^{-w}} \tag{D19}$$

depends on the parameter w, which tells us how likely it is to find a system with a particular number of particles in it. Because of the symmetry we can assume anew that $w \geq 0$, where $w = 0$ corresponds to the symmetric probability weight $\varrho_w(N) = 2^{-V}\binom{V}{N}$.

Let us recall that this binomial distribution is characterized by the mean $\bar{N}$, variance $\Delta^2 N$, and skewness ν_N:

$$\bar{N} = Vp = \frac{Ve^{-w}}{1+e^{-w}}, \quad (\Delta N)^2 = Vpq = \frac{Ve^{-w}}{(1+e^{-w})^2}, \quad \text{and} \quad \nu_N = \frac{q-p}{\sqrt{Vqp}} = \frac{2\sinh(w/2)}{\sqrt{V}}. \tag{D20}$$

We need these quantities for our dual approach when averaging, which has been similarly applied to the average over N_A in Appendix A 1 c. As in the Page case, we have to deal with kinks in the entanglement entropy averaged over all Gaussian states, which result from the fact that we have to introduce the symmetry in the particle number $N = nV$ and the subsystem size $V_A = fV$. We would like to highlight that, despite the difference in the physical origin of the average, for Page over N_A and in the present case over N, the mathematical problem is very similar. The average will be split into a sum extending over the whole range of $N = 0, \ldots, V$ for the part of the quantity that the maximum of the binomial weight $\varrho_w(N)$ lies in. For this quantity, we make use of the exact cumulants shown above. In the remaining parts, which are subleading as we will see, we approximate the binomial distribution by a normal distribution and the sum by an integral.

In the particular case $w = 0$, the skewness of the binomial distribution vanishes. This will have an important impact, as we will see since many terms will drop out.

Let us briefly outline the strategy.

1. In Appendix D 2 b, we assume that $f \leq \frac{1}{2}$ and $w \geq 0$ (thus, the average particle number $\bar{n} \leq \frac{1}{2}$) are fixed and do not follow a double scaling in the limit $V \to \infty$.
2. We expand $\langle S_A \rangle_{G,N}$ up to order $1/V$, where we make the surprising observation that there is no term of order 1.
3. To get the correct weighted average up to order $1/V$, it is therefore sufficient to evaluate $\langle S_A \rangle_{G,N}$ at $N = \bar{N} = V\bar{n}$, where the binomial distribution is peaked and then take the binomial average for the leading-order term s_A^G in $\langle S_A \rangle_{G,N} = V s_A^G + O(V^{-1})$ into account.
4. For this, we expand s_A^G around $\bar{n}$ up to fourth order. The nonanalyticities along the symmetry axis $f = n$ have to be dealt, separately, but as the binomial distribution becomes increasingly narrow around $\bar{n}$, we only need to take them into account if $\bar{n} = f$.
5. Combining all terms than yields the weighted average $\langle S_A \rangle_{G,w}$ up to order $O(1/V)$ for fixed f and $\bar{n}$.
6. In Appendix D 2 c, we zoom into the critical line $\bar{n} = f < \frac{1}{2}$ and the multicritical point $\bar{n} = f = \frac{1}{2}$. The critical regime about these points is dictated by the width of the binomial distribution, which, for $n = N/V$, is of order $1/\sqrt{V}$. This allows us to resolve any kinks in the expansion up to order $O(1/\sqrt{V})$.

b. Resulting formula

When we take the averages, we need to respect the validity of Eq. (D13), which is $V_A \leq N \leq V - V_A$ or, equivalently, $f \leq n \leq 1 - f$. Hence, we must take into account the kinks we introduce when we enforce the symmetries $f \leftrightarrow n \leftrightarrow$

TABLE II. The probability distribution $\varrho_w(n)$ is peaked around $\bar{n}$, so we expand $s_A^G = \sum_{m=0}^4 s_A^{(m)} (n-\bar{n})^m + O(n-\bar{n})^5$ from $\langle S_A\rangle_{G,N} = V s_A^G + O(V^{-1})$ up to fourth order for the cases $f < \bar{n}$ and $f > \bar{n}$, showing the nonanalyticity of $\langle S_A\rangle_{G,N}$ around $f = \bar{n}$. This nonanalyticity only shows up starting with the third order. The coefficients for $n > 1-f$ follow from the first column when substituting $\bar{n} \to 1-\bar{n}$ and multiplying by $(-1)^m$, which is a consequence of the particle-hole symmetry $n \leftrightarrow 1-n$ of the entanglement entropy.

m	$s_A^{(m)}$ for $n \le f \le \frac{1}{2}$	$s_A^{(m)}$ for $f \le n \le 1-f$
0	$(f-1)\bar{n}\ln(1-f) - \bar{n}(1+f\ln f)$ $+(\bar{n}-1)\ln(1-\bar{n})$	$(f-1)\ln(1-f)$ $+f[(\bar{n}-1)\ln(1-\bar{n}) - 1 - \bar{n}\ln\bar{n}]$
1	$(f-1)\ln(1-f) - f\ln f + \ln(1-\bar{n})$	$f\ln[(1-\bar{n})/\bar{n}]$
2	$1/2(\bar{n}-1)$	$f/2(\bar{n}-1)\bar{n}$
3	$-1/6(\bar{n}-1)^2$	$(f-2f\bar{n})/6(\bar{n}-1)^2\bar{n}^2$
4	$1/12(\bar{n}-1)^3$	$(f/12)[1/(\bar{n}-1)^3 - 1/\bar{n}^3]$

$1-n$. We have the identifications

$$\langle S_A\rangle_{G,N} = \begin{cases} 1 - \dfrac{N}{V}(1+V) + V\Psi(V) & \\ \quad -\dfrac{N}{V}[(V-V_A)\Psi(V-V_A) + V_A\Psi(V_A)] & \\ \quad +(N-V)\Psi(V-N+1), & N \le V_A, \\ 1 - \dfrac{V_A}{V}(1+V) + V\Psi(V) & \\ \quad -\dfrac{V_A}{V}[(V-N)\Psi(V-N) + N\Psi(N)] & \\ \quad +(V_A-V)\Psi(V-V_A+1), & V_A \le N \le V-V_A, \\ 1 - \dfrac{V-N}{V}(1+V) + V\Psi(V) & \\ \quad -\dfrac{V-N}{V}[(V-V_A)\Psi(V-V_A) + V_A\Psi(V_A)] & \\ \quad -N\Psi(N+1), & V-V_A \le N. \end{cases} \tag{D21}$$

Hence, when denoting the mean $\bar{n} = 1/(1+e^w)$, around which we need to expand $\langle S_A\rangle_{G,N}$, it is necessary to distinguish the three cases $n < f$, $f < n < 1-f$, and $1-f < n$, as each of them requires a different expansion. What is very beneficial is that $\langle S_A\rangle_{G,N}$ agrees up to order $1/V$ in the limit $V \to \infty$ so that the expansion coefficients will be the same.

We begin by expanding $\langle S_A\rangle_{G,N}$ as

$$\langle S_A\rangle_{G,N} = V\sum_{m=0}^4 s_A^{(m)}(n-\bar{n})^m + \frac{s_{A,0}}{V} + o(V^{-1})$$
$$\text{with} \quad s_{A,0} = \begin{cases} \dfrac{f(1-f) + f\bar{n}(1-\bar{n})}{12(1-f)(1-\bar{n})\bar{n}}, & f < \bar{n}, \\ \dfrac{\bar{n}(1-\bar{n}) + f\bar{n}(1-f)}{12(1-f)(1-\bar{n})f}, & f > \bar{n}. \end{cases} \tag{D22}$$

Since $n-\bar{n}$ will be of order $1/\sqrt{V}$ and we multiply by V, this Taylor expansion corresponds to an expansion of up to order $1/V$. We list the respective expansion coefficients $s_A^{(m)}$ in Table II for $n \le 1-f$, which are all of order one when we do not choose any double scaling limit in f and $\bar{n}$. The coefficients for $n \ge 1-f$ can be obtained when employing the symmetry $n \to 1-n$ in the case $n < f$. The constant $s_{A,0}$ is stated separately because it is the only one that is explicitly multiplied by $1/V$ while the V dependence in the other terms only enters via averaging over $(n-\bar{n})^m$.

When computing the leading-order behavior of the respective average $\langle S_A\rangle_{G,w}$, we need to compute the averages $\langle(n-\bar{n})^m\rangle$ of these powers. As long as $|f-\bar{n}|$ is larger than order $1/\sqrt{V}$ and $\bar{n} \le \frac{1}{2}$, which is equivalent to $w \ge 0$,

the power series will be the same on both sides of the maximum, i.e., for $n > \bar{n}$ and $n < \bar{n}$. Then, the kinks are not visible as they are too far away to have a nonexponential suppression. So we can use the known averages of the binomial distribution given by

$$\langle (n-\bar{n})^m \rangle = \begin{cases} 1, & m = 0, \\ 0, & m = 1, \\ \dfrac{pq}{V}, & m = 2, \\ \dfrac{pq(q-p)}{V^2}, & m = 3, \\ \dfrac{p^2q^2}{V^2}\left(3 + \dfrac{1-6qp}{Vpq}\right), & m = 4, \end{cases} \tag{D23}$$

with

$$p = \frac{1}{1+e^w} = \bar{n} \quad \text{and} \quad q = \frac{1}{1+e^{-w}} = 1-\bar{n}.$$

Hence, in the case $|f - \bar{n}| \gg 1/\sqrt{V}$ with $0 < \bar{n} \leq \frac{1}{2}$ the full average is given by

$$\langle S_A \rangle_{G,w} = V s_A^{(0)} + \bar{n}(1-\bar{n}) s_A^{(2)} + \frac{\bar{n}(1-\bar{n})(1-2\bar{n})}{V} s_A^{(3)} + 3\frac{\bar{n}^2(1-\bar{n})^2}{V} s_A^{(4)} + \frac{s_{A,0}}{V} + o(V^{-1}), \tag{D24}$$

where we have used the fact that the next-order term $\langle (n-\bar{n})^5 \rangle$ would be of order V^{-3} since the leading order has a vanishing asymmetry.

In Eq. (D24), only the coefficients $s_A^{(0)}$ and $s_A^{(2)}$ are important when one is interested in terms up to order 1. Table II shows that the expansion coefficients up to order 2 are continuous at $f = \bar{n}$. Thus, Eq. (D24) still holds up to order 1 even when $|f - \bar{n}|$ is of order $1/\sqrt{V}$ or smaller. The key subtlety of this formula lies in the third and fourth moments. The cause for this is a discontinuity showing in the third and higher derivatives of $\langle S_A \rangle_{G,N}$ at $fV = V_A = N = nf$. This requires us to compute the expectation value of $(n-\bar{n})^3$ and $(n-\bar{n})^4$, separately, when $|f - \bar{n}| = O(V^{-1/2})$.

For this purpose, we need to approximate the average over the binomial distribution by a Gaussian integral. However, we need to take particular care beyond the Gaussian case, which is given by the Edgeworth series

$$\varrho_w(Vn) = \frac{\sqrt{V}}{\sqrt{2\pi}[\bar{n}(1-\bar{n}) + O(1/V)]} e^{-V(n-\bar{n})^2/2\bar{n}(1-\bar{n})}[1 + \alpha(n-\bar{n}) + O[(n-\bar{n})^2]]. \tag{D25}$$

Such a series approximates the original probability distribution while cumulants up to a particular order are chosen to be exact. In our case, the constant α is chosen such that we match the skewness of the binomial distribution, i.e., the case $m = 3$ from Eq. (D23). This leads to the requirement that $\alpha = (q-p)/3pq = (1-2\bar{n})/3\bar{n}(1-\bar{n})$. The normalization stays the same up to order $1/V$ as the Gaussian without the Edgeworth series because the first-order correction drops out when integrating over n.

We can stop with the first order since $(n-\bar{n}) = O(V^{-1/2})$ and the zeroth, first, and second moments of the binomial distribution do not need this approximation because the coefficients are the same at $f = \bar{n}$. The fourth moment of the binomial distribution already comes with order $1/V$ [cf. Eq. (D24)], such that the Edgeworth series is not needed for this term as it gives only higher-order corrections. The only term that has been treated with the Edgeworth series is the third-order term $(\bar{n}-n)^3$ that *a priori* starts with an order $1/V^{3/2}$. Thence, the correction via the Edgeworth series only mixes the third with the fourth moments so that the $1/V$ term is correctly attributed.

When expanding s_A^G from $\langle S_A \rangle_{G,N} = V s_A^G + O(V^{-1})$ around $\bar{n} = f$, the expansion coefficients $s_A^{(m)}$ differ at $m = 3$ for $n < f$ and $n > f$. We refer to these different coefficients by $s_A^{(m-)}$ and $s_A^{(m+)}$, respectively. The resulting average is then given by

$$\begin{aligned} \langle S_A \rangle_{G,w} = {} & V \sum_{m=0}^{4} s_A^{(m-)} \langle (n-\bar{n})^m \rangle + \frac{s_{A,0}}{V} \\ & + V(s_A^{(3+)} - s_A^{(3-)}) \int_f^\infty \varrho_w(n)(n-\bar{n})^3 dn + O(V^{-3/2}), \end{aligned} \tag{D26}$$

where we first expand everything for $n < f$ and then correct for $n > f$ at third order. Most expansion coefficients $s_A^{(m\pm)}$ can be read off Table II except for $\bar{n} = f = \frac{1}{2}$. In this case, there is no region with $f < n < 1-f$ such that we have the expansion for $n < f = \frac{1}{2}$ and $n > 1-f = \frac{1}{2}$. The coefficients for the latter can be retrieved via the $n \to 1-n$ mirror symmetry, which implies that $s_A^{(m+)} = (-1)^m s_A^{(m-)}$. When expanding s_A^G around $\bar{n} = f = \frac{1}{2}$, this yields $s_A^{(3+)} - s_A^{(3-)} = \frac{4}{3}$. Recall that we consider $\bar{n} \leq \frac{1}{2}$ and hold n and f fixed, when taking the limit $V \to \infty$. The double scaling limits, when zooming into the critical points, will be discussed in Appendix D 2 c.

Whenever $f \neq n$, the error is exponentially suppressed for $V \to \infty$, which is why we only need to take this term for $f = \bar{n}$ into account. While the averages $\langle (n-\bar{n})^m \rangle$ from Table II are exact, the integrals in Eq. (D26) have an additional error due to the fact that we approximate the binomial sum by a continuous integral. However, at third order this will only induce a subleading error of order $V^{-5/2}$ (eventually of order $V^{-3/2}$ after multiplying by V).

In summary, we evaluate Eq. (D26) to find Eq. (118) of the main text, where we included all terms up to order $1/V$ and indicated the correct next order based on the numerical results shown in Fig. 11.

c. Resolving the critical regimes

The average $\langle S_A \rangle_w$ is described by a continuous function at linear and constant order in V, but the term of order $1/\sqrt{V}$ only appears when $f = \bar{n}$. Based on our previous analysis, it is therefore a natural question to analyze this term to understand how this critical regime is resolved when being close to $\bar{n} = f$. The width of the approximate binomial distribution $\varrho_w(n)$ scales as $1/\sqrt{V}$, so only if $\bar{n} = f + \Lambda_{\bar{n}}/\sqrt{V}$ will the discontinuity in the third derivative contribute to the average. We therefore analyze this limit to resolve the term of order $1/\sqrt{V}$ around $f = \bar{n}$.

In the following, we only analyze the contribution towards the $1/\sqrt{V}$ term. While there is also a discontinuity at order $1/V$, the actual calculation is rather tedious and the result is quite lengthy, but can be carried out with the same techniques presented here. For our purpose, it is indeed sufficient to resolve all discontinuities up to order $1/\sqrt{V}$.

Transition at $\bar{n} = f = \frac{1}{2}$. Around the multicritical point $\bar{n} = f = \frac{1}{2}$, the correction of order $1/\sqrt{V}$ is due to the discontinuities of the third derivative in $s_A^G(f,n)$ with the two cases $\bar{n} < f$ and $1-f > \bar{n} > f$ that need to be distinguished. We recall that $\langle S_A \rangle_{G,N} = V s_A^G + O(V^{-1})$. The relevant scalings of $\bar{n}$ and f are given by

$$\bar{n} = \frac{1}{2} + \frac{\Lambda_{\bar{n}}}{\sqrt{V}} \quad \text{and} \quad f = \frac{1}{2} + \frac{\Lambda_f}{\sqrt{V}}. \tag{D27}$$

The two kinks result in two discontinuities in the third derivative of $s_A^G(f,n)$, namely at $n = \frac{1}{2} \pm \Lambda_f/\sqrt{V}$. Those imply different formulas in the entanglement in the following regions (see Fig. 18):

$$\begin{aligned} s_A^G(n,f) \quad & \text{in region 1}, n \leq \frac{1}{2} - \frac{|\Lambda_f|}{\sqrt{V}}, \\ s_A^G(f,n) \quad & \text{in region 2}, \frac{1}{2} - \frac{|\Lambda_f|}{\sqrt{V}} < n < \frac{1}{2} + \frac{|\Lambda_f|}{\sqrt{V}}, \\ s_A^G(1-n,f) \quad & \text{in region 3}, \frac{1}{2} + \frac{|\Lambda_f|}{\sqrt{V}} \leq n. \end{aligned}$$

Here the function $s_A^G(f,n)$ represents the leading order in $\langle S_A \rangle_{G,N} = V s_A^G(f,n) + O(1)$ (computed for $f \leq n \leq \frac{1}{2}$ and then analytically continued), and is given by

$$s_A^G(f,n) = (f-1)\ln(1-f) + f\left[(n-1)\ln(1-n) - n\ln n - 1\right] \tag{D28}$$

in the region $f \leq n \leq \frac{1}{2}$, which is actually also valid for $\frac{1}{2} \leq n \leq f$ (so that we did not have to split up region 2).

The main idea of the ensuing computation is to choose the region where $\bar{n}$ lies in and then use Eq. (118) for this particular region. Since there are also contributions from the other regions because the width of the binomial weight also covers parts therein, we need to additionally compute the average over the difference between the corresponding cases in Eq. (118) over these regions. Let us underline that, when $w \geq 0$, it holds that $\bar{n} \leq \frac{1}{2}$. Thence, it holds that $\Lambda_{\bar{n}} < 0$ and there are the two cases, $\bar{n} < f$ and $1-f > \bar{n} > f$, to distinguish that are reflected in $\Lambda_{\bar{n}} < -|\Lambda_f|$ and $0 > \Lambda_{\bar{n}} > -|\Lambda_f|$.

Case 1: $\Lambda_{\bar{n}} < -|\Lambda_f|$. As $\bar{n} < f$, the maximum of the binomial distribution lies in region 1. Hence, we rewrite the three sums that constitute the entanglement entropy into a sum where the index for $s_A^G(n,f)$ runs over the whole range from 0

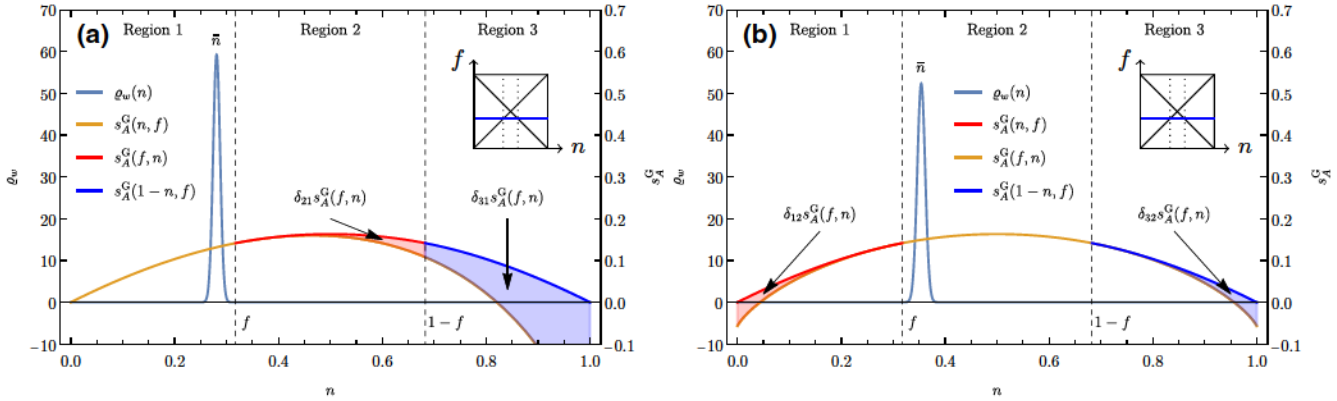


FIG. 18. Illustration of integrals. We illustrate the calculation of $d\langle S_A\rangle_w$ around $\bar{n}=f=\frac{1}{2}$, where (a) refers to case 1 (with $\Lambda_f=-1$, $\Lambda_{\bar{n}}=-1.2$, $V=30$) and (b) refers to case 2 (with $\Lambda_f=-1$, $\Lambda_{\bar{n}}=-0.8$, $V=30$). Note that we use different scales for ϱ_w and s_A^G. Also, the proportions are slightly off since we have used a relatively small V. The size of region 2 is comparable to the width of the binomial distribution when $V\gg 1$.

to V and then sum over the correction, which is the difference to $s_A^G(n,f)$, in the other two sums. In particular, we write

$$\begin{aligned}\langle S_A\rangle_{G,w} &= V\sum_{N=0}^{V_A}\varrho_w(N)s_A^G(n,f)+V\sum_{N=V_A+1}^{V-V_A-1}\varrho_w(N)s_A^G(f,n)\\ &\quad+V\sum_{N=V-V_A}^{V}\varrho_w(N)s_A^G(1-n,f)+O(V^{-1})\\ &= V\sum_{N=0}^{V}\varrho_w(N)s_A^G(n,f)+V\sum_{N=V_A+1}^{V-V_A-1}\varrho_w(N)\delta_{21}s_A^G(f,n)\\ &\quad+V\sum_{N=V-V_A}^{V}\delta_{31}\varrho_w(N)s_A^G(f,n)+O(V^{-1}).\end{aligned}\tag{D29}$$

The differences are defined as

$$\delta_{21}s_A^G(f,n)=s_A^G(f,n)-s_A^G(n,f),\qquad \delta_{31}s_A^G(f,n)=s_A^G(1-n,f)-s_A^G(n,f),\tag{D30}$$

and they are illustrated in Fig. 18. For the first sum, we can employ the first case in Eq. (118). The other two sums simplify when noting that the differences $\delta_{21}s_A^G(f,n)$ and $\delta_{31}s_A^G(f,n)$ have vanishing partial derivatives at $f=\bar{n}=\frac{1}{2}$ up to order 2. Thus, we can replace $\varrho_w(Vn)$ by its Gaussian approximation without the Edgeworth series correction because the terms will already start with order $1/\sqrt{V}$. The total correction can then be summarized as the integral

$$\begin{aligned}d\langle S_A\rangle_{G,w}^{\bar{n}=f=1/2} &= V\sum_{m=0}^{3}\left(\delta_{21}s_A^{(m)}\int_f^{1-f}\varrho_w(Vn)(n-\bar{n})^m dn+\delta_{31}s_A^{(m)}\int_{1-f}^{\infty}\varrho_w(Vn)(n-\bar{n})^m dn\right)\\ &\quad+o(1/\sqrt{V}),\end{aligned}\tag{D31}$$

where $\delta_{ij}s_A^{(m)}$ are the respective expansion coefficients from Eq. (D30) expanded in $(n-\bar{n})^m$. We directly combine the result with the second case.

Case 2: $0 > \Lambda_{\bar{n}} > -|\Lambda_f|$. Now, the main contribution comes from region 2 where $\frac{1}{2} - \Lambda_f/\sqrt{V} \le n \le \frac{1}{2} + \Lambda_f/\sqrt{V}$. Therefore, we split the sums as

$$\langle S_A\rangle_{G,w} = V\sum_{N=0}^{V} \varrho_w(N) s_A^G(f,n) + V\sum_{N=0}^{V_A} \varrho_w(N)\delta_{12}s_A^G(f,n) + V\sum_{N=V-V_A}^{V} \delta_{31}\varrho_w(N)\delta_{32}s_A^G(f,n) + O(V^{-1}) \tag{D32}$$

with

$$\delta_{12}s_A^G(f,n) = s_A^G(n,f) - s_A^G(f,n), \qquad \delta_{32}s_A^G(f,n) = s_A^G(1-n,f) - s_A^G(f,n). \tag{D33}$$

The latter are illustrated in Fig. 18(b). As in case 1, we replace the first sum by the second case in Eq. (118), and approximate the remaining two sums by the Gaussian integrals

$$d\,\langle S_A\rangle_{G,w}^{\bar{n}=f=\frac{1}{2}} = V\sum_{m=0}^{3}\left(\delta_{12}s_A^{(m)}\int_{-\infty}^{f}\varrho_w(Vn)(n-\bar{n})^m dn + \delta_{32}s_A^{(m)}\int_{1-f}^{\infty}\varrho_w(Vn)(n-\bar{n})^m dn\right) + o(1/\sqrt{V}), \tag{D34}$$

where $\delta_{ij}s_A^{(m)}$ are the respective expansion coefficients from Eq. (D33) expanded in $(n-\bar{n})^m$.

We combine both cases and extend the result to positive and negative Λ_f and $\Lambda_{\bar{n}}$ by applying the symmetries of the entanglement entropy resulting in the absolute values of Λ_f and $\Lambda_{\bar{n}}$:

$$\begin{aligned} d\,\langle S_A\rangle_{G,w}^{\bar{n}=f=1/2} = \frac{1}{12}\Bigg[&\sqrt{\frac{2}{\pi}}\big(e^{-2(|\Lambda_f|-|\Lambda_{\bar{n}}|)^2}[1+2(|\Lambda_f|-|\Lambda_{\bar{n}}|)^2] \\ &+ e^{-2(|\Lambda_f|+|\Lambda_{\bar{n}}|)^2}[1+2(|\Lambda_f|+|\Lambda_{\bar{n}}|)^2]\big) \\ &- (|\Lambda_f|+|\Lambda_{\bar{n}}|)[3+4(|\Lambda_f|+|\Lambda_{\bar{n}}|)^2]\mathrm{erfc}[\sqrt{2}(|\Lambda_f|+|\Lambda_{\bar{n}}|)] \\ &- ||\Lambda_f|-|\Lambda_{\bar{n}}||[3+4(|\Lambda_f|-|\Lambda_{\bar{n}}|)^2]\mathrm{erfc}(\sqrt{2}||\Lambda_f|-|\Lambda_{\bar{n}}||)\Bigg]\frac{1}{\sqrt{V}} \\ &+ o(1/\sqrt{V}). \end{aligned} \tag{D35}$$

We still need to expand result (118) around $f = \frac{1}{2} + \Lambda_f/\sqrt{V}$ and $n = \frac{1}{2} + \Lambda_{\bar{n}}/\sqrt{V}$, where we expand the case $f < \bar{n} \le 1/2$ in Eq. (118) for the case 1 of the present subsection and the case $\bar{n} < f \le \frac{1}{2}$ for the case 2 of the present subsection. Adding the correction from Eq. (D34) leads to the final result

$$\begin{aligned} \langle S_A\rangle_{G,w}^{n=f=\frac{1}{2}} = &\left(\ln 2 - \frac{1}{2}\right)V - \left(\frac{1}{4} + \Lambda_f^2 + \Lambda_{\bar{n}}^2\right) \\ &- \frac{\max(|\Lambda_f|,|\Lambda_{\bar{n}}|)}{6}(3 + 12\min(\Lambda_f^2,\Lambda_{\bar{n}}^2) + 4\max(\Lambda_f^2,\Lambda_{\bar{n}}^2))\frac{1}{\sqrt{V}} \\ &+ d\,\langle S_A\rangle_{G,w}^{\bar{n}=f=1/2} + O(1/V). \end{aligned} \tag{D36}$$

This result reflects the fact that the nonanalyticities only show in the $1/\sqrt{V}$ corrections and lower orders. In Fig. 10(d), we show $d\,\langle S_A\rangle_{G,w}^{n=f=1/2}$, which resolves the Kronecker delta $\delta_{f,\bar{n}}\delta_{f,1/2}$ contained in Eq. (118).

The computation above could have also been computed in different ways. For instance, an expansion of all quantities about $f = \bar{n} = n = \frac{1}{2}$ would have led to the same result. This approach would not need any case discussion at the expense that the error functions involved have to be expanded as they contain terms that are proportional to $\sqrt{V}$.

Transition at $\bar{n} = f < \frac{1}{2}$. Around $\bar{n} = f < \frac{1}{2}$ with $1 - 2f \gg 1/\sqrt{V}$, the correction of order $1/\sqrt{V}$ is exclusively due to the discontinuity of the third derivative in $\langle S_A\rangle_{GN}$, so we only need to analyze this case. Applying the same strategy as before, by looking for the main contribution in the sum, we need to distinguish if $\bar{n} < f$ and $\bar{n} > f$. We emphasize that

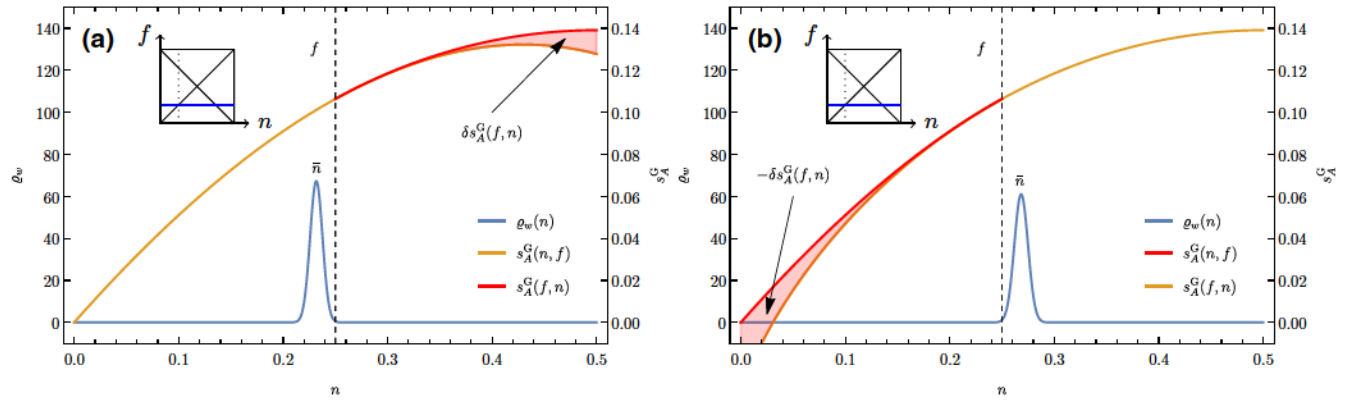


FIG. 19. Illustration of integrals. We illustrate the calculation of $d\langle S_A\rangle_w$ around $f=\frac{1}{4}$, where (a) refers to case 1 (with $\Lambda_f=-0.1$, $V=30$) and (b) refers to case 2 (with $\Lambda_f=0.1$, $V=30$). In contrast to the previous case around $f=\bar{n}=\frac{1}{2}$, we can ignore the third region $n>1-f$, as the effect is exponentially suppressed. Note that we use different scales for ϱ_w and s_A^G. As in Fig. 18, the proportions are a bit off due to the relatively small V that has been chosen (to enhance readability). Actually, the dashed vertical line lies inside the width of the binomial distribution for $V\gg 1$.

contributions from region 3, where $n>1-f$, are always exponentially suppressed because this region lies at a distance that is much larger than the width of the binomial distribution.

Case 1: $\bar{n}<f$. In this case, the main contribution lies in region 1 [see Fig. 19(a)] so that we express the average as

$$\langle S_A\rangle_{G,w}=V\sum_{N=0}^{V}\varrho_w(N)s_A^G(f,n)+V\sum_{N=V_A+1}^{V}\varrho_w(N)\delta_{12}\delta s_A^G(f,n)+O(V^{-1}) \tag{D37}$$

with $\delta s_A^G(f,n)=s_A^G(n,f)-s_A^G(f,n)$ and $s_A^G(f,n)$ given by the first case in Eq. (118). We recall that the sum over the third region is exponentially small. When expanding about $\bar{n}<f$, we need to take into account the nonanalyticity that shows up with the third partial derivatives in n and f. As before, these orders of the Taylor expansion come with the order $1/\sqrt{V}$ once we have multiplied with the prefactor V. Therefore, the correction is given by

$$d\langle S_A\rangle_{G,w}^{\bar{n}=f<1/2}=V\sum_{m=0}^{3}\delta s_A^{(m)}\int_f^\infty\varrho_w(n)(n-\bar{n})^m dn+o(1/V), \tag{D38}$$

where the additional error by approximating the discrete binomial sum by an integral over $\varrho_w(n)$ is subleading, as it will contribute towards the overall error of order $1/V$. The coefficients $\delta s_A^{(m)}$ are the Taylor expansion coefficients of the difference $\delta s_A^G(f,n)$. Those are given by

$$\begin{aligned}\delta s_A^G(f,n)=&-\frac{(f-\bar{n})^3[f(4f-2\bar{n}-3)+\bar{n}]}{12[(f-1)^2f^2]}+\frac{(f-\bar{n})^2[f(7f-4\bar{n}-5)+2\bar{n}]}{6(f-1)^2f^2}(n-\bar{n})\\&-\frac{(f-\bar{n})[3f^2-2f(\bar{n}+1)+\bar{n}]}{2[(f-1)^2f^2]}(n-\bar{n})^2+\frac{[f(5f-4\bar{n}-3)+2\bar{n}]}{6(f-1)^2f^2}(n-\bar{n})^3\\&+O[(n-\bar{n})^4].\end{aligned} \tag{D39}$$

These coefficients are the final ingredient to evaluate Eq. (D38). Before we do this, we also consider the second case and then directly combine the results.

Case 2: $\bar{n}>f$. Now, the main contribution lies in region 2 [see Fig. 19(b)], and we employ the splitting

$$\langle S_A\rangle_{G,w}=V\sum_{N=0}^{V}\varrho_w(N)s_A^G(n,f)-V\sum_{N=0}^{V_A}\varrho_w(N)\delta_{12}\delta s_A^G(f,n)+O(V^{-1}) \tag{D40}$$

with the very same $\delta s_A^G(f,n) = s_A^G(n,f) - s_A^G(f,n)$ as in case 1. Repeating the same steps as in the previous case, we find the correction term

$$d\langle S_A\rangle_w^{\bar{n}=f<1/2} = -V\sum_{m=0}^{3}\delta s_A^{(m)}\int_{-\infty}^{f}\varrho_w(n)(n-\bar{n})^m dn + o(1/\sqrt{V}), \tag{D41}$$

where the overall minus sign comes from the fact that the error for $n<f$ is just opposite to that previously calculated, i.e., the expansion coefficients $\delta s_A^{(m)}$ are still those listed in Eq. (D39).

Using the absolute value to unify the previous results of both cases for a general $\bar{n} = f + \Lambda_{\bar{n}}/\sqrt{V}$, where $\Lambda_{\bar{n}}$ can be both positive or negative, we arrive at

$$\begin{aligned} d\langle S_A\rangle_{G,w}^{\bar{n}=f<1/2} = \Bigg(&\frac{e^{-\Lambda_{\bar{n}}^2/2f(1-f)}(2f^{5/2}-2f^{3/2}-f^{1/2}\Lambda_{\bar{n}}^2)}{6\sqrt{1-f}f\sqrt{2\pi}} \\ &+\frac{|\Lambda_{\bar{n}}|[3(1-f)f+\Lambda_{\bar{n}}^2]\mathrm{erfc}(|\Lambda_{\bar{n}}|/\sqrt{2f(1-f)})}{12(1-f)f}\Bigg)\frac{1}{\sqrt{V}} + o(1/\sqrt{V}). \end{aligned} \tag{D42}$$

We can relate this result to our previous finding in Eq. (D34). If we set $\Lambda_{\bar{n}} \to \Lambda_f + \Lambda_{\bar{n}}$ in Eq. (D34) and then consider the limit $\Lambda_f \to \infty$, we reproduce Eq. (D42) at $f = \frac{1}{2}$ from above. Consequently, the different limits connect at $f = \frac{1}{2}$, as expected.

The full asymptotic of $\langle S_A\rangle_{G,w}^{\bar{n}=f=1/2}$ is given by first expanding formula (118) in the respective regions for $f<n$ and $f>n$ (depending on the sign of $\Lambda_{\bar{n}}$) and then adding the corrections (D42). For $n = f + \Lambda_{\bar{n}}/\sqrt{V}$, we eventually find that

$$\begin{aligned} \langle S_A\rangle_{G,w}^{\bar{n}=f<1/2} = &[(f^2-1)\ln(1-f) - f(1+f\ln f)]V + \Lambda_{\bar{n}} f[\ln(1-f) - \ln f]\sqrt{V} \\ &-\frac{1}{2}\left(\frac{\Lambda_{\bar{n}}^2}{1-f} - f\right) \\ &+\left[\frac{\Theta(\Lambda_{\bar{n}})\Lambda_{\bar{n}}^3(1-2f)}{6f(1-f)^2} - \frac{\Theta(-\Lambda_{\bar{n}})}{6}\left(3+\frac{\Lambda_{\bar{n}}^2}{(1-f)^2}\right)\right]\frac{1}{\sqrt{V}} \\ &+ d\langle S_A\rangle_{G,w}^{\bar{n}=f<1/2} + O(1/V), \end{aligned} \tag{D43}$$

where Θ is Heaviside step function [with $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ otherwise]. We show $d\langle S_A\rangle_{G,w}^{n=f<1/2}$ in Fig. 10(e), which resolves the Kronecker delta $\delta_{f,\bar{n}}$ contained in Eq. (118).

Anew, one could have taken again a different approach without case discussion and an expansion about $\bar{n} = n = f$. This would have given the same result only the Heaviside step function would have been encoded in additional error functions whose argument would have been proportional to $\sqrt{V}$.

[1] D. N. Page, Information in Black Hole Radiation, Phys. Rev. Lett. **71**, 3743 (1993).

[2] J. Eisert, M. Cramer, and M. B. Plenio, Colloquium: Area laws for the entanglement entropy, Rev. Mod. Phys. **82**, 277 (2010).

[3] J. S. Bell, On the Einstein Podolsky Rosen paradox, Phys. Phys. Fiz. **1**, 195 (1964).

[4] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[5] M. B. Plenio and S. Virmani, An introduction to entanglement measures, Quantum Info. Comput. **7**, 1 (2007).

[6] For fermionic systems, as considered later, one needs to work with a fermionic generalization of the tensor product, which also gives rise to a fermionic notion of the partial trace [217].

[7] R. Islam, R. Ma, P. M. Preiss, M. E. Tai, A. Lukin, M. Rispoli, and M. Greiner, Measuring entanglement entropy in a quantum many-body system, Nature **528**, 77 (2015).

[8] 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).

[9] L. Amico, R. Fazio, A. Osterloh, and V. Vedral, Entanglement in many-body systems, Rev. Mod. Phys. **80**, 517 (2008).

[10] I. Peschel and V. Eisler, Reduced density matrices and entanglement entropy in free lattice models, J. Phys. A **42**, 504003 (2009).

[11] P. Calabrese and J. Cardy, Entanglement entropy and conformal field theory, J. Phys. A **42**, 504005 (2009).

[12] K. Audenaert, J. Eisert, M. B. Plenio, and R. F. Werner, Entanglement properties of the harmonic chain, Phys. Rev. A **66**, 042327 (2002).

[13] A. Osterloh, L. Amico, G. Falci, and R. Fazio, Scaling of entanglement close to a quantum phase transition, Nature **416**, 608 (2002).

[14] T. J. Osborne and M. A. Nielsen, Entanglement in a simple quantum phase transition, Phys. Rev. A **66**, 032110 (2002).

[15] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, Entanglement in Quantum Critical Phenomena, Phys. Rev. Lett. **90**, 227902 (2003).

[16] J. I. Latorre, E. Rico, and G. Vidal, Ground state entanglement in quantum spin chains, Quantum Info. Comput. **4**, 48 (2004).

[17] M. B. Hastings, An area law for one-dimensional quantum systems, J. Stat. Mech. **2007**, 08024 (2007).

[18] P. Calabrese and J. Cardy, Entanglement entropy and quantum field theory, J. Stat. Mech. **2004**, 06002 (2004).

[19] M. M. Wolf, Violation of the Entropic Area Law for Fermions, Phys. Rev. Lett. **96**, 010404 (2006).

[20] D. Gioev and I. Klich, Entanglement Entropy of Fermions in Any Dimension and the Widom Conjecture, Phys. Rev. Lett. **96**, 100503 (2006).

[21] T. Barthel, M.-C. Chung, and U. Schollwöck, Entanglement scaling in critical two-dimensional fermionic and bosonic systems, Phys. Rev. A **74**, 022329 (2006).

[22] W. Li, L. Ding, R. Yu, T. Roscilde, and S. Haas, Scaling behavior of entanglement in two- and three-dimensional free-fermion systems, Phys. Rev. B **74**, 073103 (2006).

[23] M. Cramer, J. Eisert, and M. B. Plenio, Statistics Dependence of the Entanglement Entropy, Phys. Rev. Lett. **98**, 220603 (2007).

[24] L. Ding, N. Bray-Ali, R. Yu, and S. Haas, Subarea Law of Entanglement in Nodal Fermionic Systems, Phys. Rev. Lett. **100**, 215701 (2008).

[25] E. Fradkin and J. E. Moore, Entanglement Entropy of 2D Conformal Quantum Critical Points: Hearing the Shape of a Quantum Drum, Phys. Rev. Lett. **97**, 050404 (2006).

[26] A. B. Kallin, M. B. Hastings, R. G. Melko, and R. R. P. Singh, Anomalies in the entanglement properties of the square-lattice Heisenberg model, Phys. Rev. B **84**, 165134 (2011).

[27] M. A. Metlitski and T. Grover, Entanglement Entropy of systems with spontaneously broken continuous symmetry (2011), ArXiv:1112.5166.

[28] A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore, Colloquium: Nonequilibrium dynamics of closed interacting quantum systems, Rev. Mod. Phys. **83**, 863 (2011).

[29] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics, Adv. Phys. **65**, 239 (2016).

[30] C. Gogolin and J. Eisert, Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems, Rep. Prog. Phys. **79**, 056001 (2016).

[31] C. Mejía-Monasterio, G. Benenti, G. G. Carlo, and G. Casati, Entanglement across a transition to quantum chaos, Phys. Rev. A **71**, 062324 (2005).

[32] V. Alba, M. Fagotti, and P. Calabrese, Entanglement entropy of excited states, J. Stat. Mech. **2009**, 10020 (2009).

[33] J. M. Deutsch, Thermodynamic entropy of a many-body energy eigenstate, New J. Phys. **12**, 075021 (2010).

[34] L. F. Santos, A. Polkovnikov, and M. Rigol, Weak and strong typicality in quantum systems, Phys. Rev. E **86**, 010102 (2012).

[35] J. M. Deutsch, H. Li, and A. Sharma, Microscopic origin of thermodynamic entropy in isolated systems, Phys. Rev. E **87**, 042135 (2013).

[36] M. Storms and R. R. P. Singh, Entanglement in ground and excited states of gapped free-fermion systems and their relationship with Fermi surface and thermodynamic equilibrium properties, Phys. Rev. E **89**, 012125 (2014).

[37] J. Mölter, T. Barthel, U. Schollwöck, and V. Alba, Bound states and entanglement in the excited states of quantum spin chains, J. Stat. Mech. **2014**, 10029 (2014).

[38] H.-H. Lai and K. Yang, Entanglement entropy scaling laws and eigenstate typicality in free fermion systems, Phys. Rev. B **91**, 081110 (2015).

[39] W. Beugeling, A. Andreanov, and M. Haque, Global characteristics of all eigenstates of local many-body Hamiltonians: Participation ratio and entanglement entropy, J. Stat. Mech. **2015**, 02002 (2015).

[40] Z.-C. Yang, C. Chamon, A. Hamma, and E. R. Mucciolo, Two-Component Structure in the Entanglement Spectrum of Highly Excited States, Phys. Rev. Lett. **115**, 267206 (2015).

[41] S. Nandy, A. Sen, A. Das, and A. Dhar, Eigenstate Gibbs ensemble in integrable quantum systems, Phys. Rev. B **94**, 245131 (2016).

[42] L. Vidmar, L. Hackl, E. Bianchi, and M. Rigol, Entanglement Entropy of Eigenstates of Quadratic Fermionic Hamiltonians, Phys. Rev. Lett. **119**, 020601 (2017).

[43] L. Vidmar and M. Rigol, Entanglement Entropy of Eigenstates of Quantum Chaotic Hamiltonians, Phys. Rev. Lett. **119**, 220603 (2017).

[44] Y. Zhang, L. Vidmar, and M. Rigol, Information measures for a local quantum phase transition: Lattice fermions in a one-dimensional harmonic trap, Phys. Rev. A **97**, 023605 (2018).

[45] A. Dymarsky, N. Lashkari, and H. Liu, Subsystem eigenstate thermalization hypothesis, Phys. Rev. E **97**, 012140 (2018).

[46] J. R. Garrison and T. Grover, Does a single eigenstate encode the full Hamiltonian?, Phys. Rev. X **8**, 021026 (2018).

[47] Y. O. Nakagawa, M. Watanabe, H. Fujita, and S. Sugiura, Universality in volume-law entanglement of scrambled pure quantum states, Nat. Commun. **9**, 1635 (2018).

[48] L. Vidmar, L. Hackl, E. Bianchi, and M. Rigol, Volume Law and Quantum Criticality in the Entanglement Entropy

of Excited Eigenstates of the Quantum Ising Model, Phys. Rev. Lett. **121**, 220602 (2018).
[49] Y. Huang, Universal eigenstate entanglement of chaotic local Hamiltonians, Nucl. Phys. B **938**, 594 (2019).
[50] L. Hackl, L. Vidmar, M. Rigol, and E. Bianchi, Average eigenstate entanglement entropy of the XY chain in a transverse field and its universality for translationally invariant quadratic fermionic models, Phys. Rev. B **99**, 075123 (2019).
[51] T.-C. Lu and T. Grover, Renyi entropy of chaotic eigenstates, Phys. Rev. E **99**, 032111 (2019).
[52] C. Murthy and M. Srednicki, Structure of chaotic eigenstates and their entanglement entropy, Phys. Rev. E **100**, 022131 (2019).
[53] A. Jafarizadeh and M. A. Rajabpour, Bipartite entanglement entropy of the excited states of free fermions and harmonic oscillators, Phys. Rev. B **100**, 165135 (2019).
[54] H. Wilming, M. Goihl, I. Roth, and J. Eisert, Entanglement-Ergodic Quantum Systems Equilibrate Exponentially Well, Phys. Rev. Lett. **123**, 200604 (2019).
[55] T. LeBlond, K. Mallayya, L. Vidmar, and M. Rigol, Entanglement and matrix elements of observables in interacting integrable systems, Phys. Rev. E **100**, 062134 (2019).
[56] D. Faiez and D. Šafránek, How much entanglement can be created in a closed system, Phys. Rev. B **101**, 060401 (2020).
[57] R. Modak and T. Nag, Many-body dynamics in long-range hopping models in the presence of correlated and uncorrelated disorder, Phys. Rev. Res. **2**, 012074 (2020).
[58] K. Kaneko, E. Iyoda, and T. Sagawa, Characterizing complexity of many-body quantum dynamics by higher-order eigenstate thermalization, Phys. Rev. A **101**, 042126 (2020).
[59] D. S. Bhakuni and A. Sharma, Entanglement and thermodynamic entropy in a clean many-body-localized system, J. Phys.: Cond. Mat. **32**, 255603 (2020).
[60] D. Faiez, D. Šafránek, J. M. Deutsch, and A. Aguirre, Typical and extreme entropies of long-lived isolated quantum systems, Phys. Rev. A **101**, 052101 (2020).
[61] P. Łydżba, M. Rigol, and L. Vidmar, Eigenstate Entanglement Entropy in Random Quadratic Hamiltonians, Phys. Rev. Lett. **125**, 180604 (2020).
[62] P. Łydżba, M. Rigol, and L. Vidmar, Entanglement in many-body eigenstates of quantum-chaotic quadratic Hamiltonians, Phys. Rev. B **103**, 104206 (2021).
[63] M. Haque, P. A. McClarty, and I. M. Khaymovich, Entanglement of midspectrum eigenstates of chaotic many-body systems: Reasons for deviation from random ensembles, Phys. Rev. E **105**, 014109 (2022).
[64] Q. Miao and T. Barthel, Eigenstate entanglement scaling for critical interacting spin chains, Quantum **6**, 642 (2022).
[65] B. Sutherland, *Beautiful Models* (World Scientific, Singapore, 2004).
[66] P. Calabrese, F. H. L. Essler, and G. Mussardo, Introduction to 'quantum integrability in out of equilibrium systems', J. Stat. Mech. **2016**, 064001 (2016).
[67] M. A. Cazalilla, R. Citro, T. Giamarchi, E. Orignac, and M. Rigol, One dimensional bosons: From condensed matter systems to ultracold gases, Rev. Mod. Phys. **83**, 1405 (2011).
[68] X.-W. Guan, M. T. Batchelor, and C. Lee, Fermi gases in one dimension: From Bethe ansatz to experiments, Rev. Mod. Phys. **85**, 1633 (2013).
[69] H. Moritz, T. Stöferle, M. Köhl, and T. Esslinger, Exciting Collective Oscillations in a Trapped 1D Gas, Phys. Rev. Lett. **91**, 250402 (2003).
[70] T. Kinoshita, T. Wenger, and D. S. Weiss, Observation of a one-dimensional Tonks-Girardeau gas, Science **305**, 1125 (2004).
[71] B. Paredes, A. Widera, V. Murg, O. Mandel, S. Fölling, I. Cirac, G. V. Shlyapnikov, T. W. Hänsch, and I. Bloch, Tonks-Girardeau gas of ultracold atoms in an optical lattice, Nature **429**, 277 (2004).
[72] T. Kinoshita, T. Wenger, and D. S. Weiss, Local Pair Correlations in One-Dimensional Bose Gases, Phys. Rev. Lett. **95**, 190406 (2005).
[73] T. Kinoshita, T. Wenger, and D. S. Weiss, A quantum Newton's cradle, Nature **440**, 900 (2006).
[74] A. H. van Amerongen, J. J. P. van Es, P. Wicke, K. V. Kheruntsyan, and N. J. van Druten, Yang-Yang Thermodynamics on an Atom Chip, Phys. Rev. Lett. **100**, 090402 (2008).
[75] M. Gring, M. Kuhnert, T. Langen, T. Kitagawa, B. Rauer, M. Schreitl, I. Mazets, D. A. Smith, E. Demler, and J. Schmiedmayer, Relaxation and prethermalization in an isolated quantum system, Science **337**, 1318 (2012).
[76] T. Fukuhara, P. Schauß, M. Endres, S. Hild, M. Cheneau, I. Bloch, and C. Gross, Microscopic observation of magnon bound states and their dynamics, Nature **502**, 76 (2013).
[77] G. Pagano, M. Mancini, G. Cappellini, P. Lombardi, F. Schäfer, H. Hu, X.-J. Liu, J. Catani, C. Sias, and M. Inguscio, *et al.*, A one-dimensional liquid of fermions with tunable spin, Nat. Phys. **10**, 198 (2014).
[78] T. Langen, S. Erne, R. Geiger, B. Rauer, T. Schweigler, M. Kuhnert, W. Rohringer, I. E. Mazets, T. Gasenzer, and J. Schmiedmayer, Experimental observation of a generalized Gibbs ensemble, Science **348**, 207 (2015).
[79] M. Boll, T. A. Hilker, G. Salomon, A. Omran, J. Nespolo, L. Pollet, and I. Bloch, Spin- and density-resolved microscopy of antiferromagnetic correlations in Fermi-Hubbard chains, Science **353**, 1257 (2016).
[80] 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).
[81] M. Schemmer, I. Bouchoule, B. Doyon, and J. Dubail, Generalized Hydrodynamics on an Atom Chip, Phys. Rev. Lett. **122**, 090601 (2019).
[82] J. M. Wilson, N. Malvania, Y. Le, Y. Zhang, M. Rigol, and D. S. Weiss, Observation of dynamical fermionization, Science **367**, 1461 (2020).
[83] P. N. Jepsen, J. Amato-Grill, I. Dimitrova, W. W. Ho, E. Demler, and W. Ketterle, Spin transport in a tunable Heisenberg model realized with ultracold atoms, Nature **588**, 403 (2020).
[84] W. Kao, K.-Y. Li, K.-Y. Lin, S. Gopalakrishnan, and B. L. Lev, Topological pumping of a 1D dipolar gas into strongly correlated prethermal states, Science **371**, 296 (2021).

[85] N. Malvania, Y. Zhang, Y. Le, J. Dubail, M. Rigol, and D. S. Weiss, Generalized hydrodynamics in strongly interacting 1D Bose gases, Science **373**, 1129 (2021).

[86] C. Liu, X. Chen, and L. Balents, Quantum entanglement of the Sachdev-Ye-Kitaev models, Phys. Rev. B **97**, 245126 (2018).

[87] Y. Huang and Y. Gu, Eigenstate entanglement in the Sachdev-Ye-Kitaev model, Phys. Rev. D **100**, 041901 (2019).

[88] P. Zhang, C. Liu, and X. Chen, Subsystem Rényi entropy of thermal ensembles for SYK-like models, SciPost Phys. **8**, 94 (2020).

[89] S. C. Morampudi, A. Chandran, and C. R. Laumann, Universal Entanglement of Typical States in Constrained Systems, Phys. Rev. Lett. **124**, 050602 (2020).

[90] O. Bohigas, M. J. Giannoni, and C. Schmit, Characterization of Chaotic Quantum Spectra and Universality of Level Fluctuation Laws, Phys. Rev. Lett. **52**, 1 (1984).

[91] R. Mondaini and M. Rigol, Eigenstate thermalization in the two-dimensional transverse field ising model. II. Off-diagonal matrix elements of observables, Phys. Rev. E **96**, 012157 (2017).

[92] D. Jansen, J. Stolpp, L. Vidmar, and F. Heidrich-Meisner, Eigenstate thermalization and quantum chaos in the Holstein polaron model, Phys. Rev. B **99**, 155130 (2019).

[93] J. Richter, A. Dymarsky, R. Steinigeweg, and J. Gemmer, Eigenstate thermalization hypothesis beyond standard indicators: Emergence of random-matrix behavior at small frequencies, Phys. Rev. E **102**, 042127 (2020).

[94] C. Schönle, D. Jansen, F. Heidrich-Meisner, and L. Vidmar, Eigenstate thermalization hypothesis through the lens of autocorrelation functions, Phys. Rev. B **103**, 235137 (2021).

[95] E. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, Ann. Math. **62**, 548 (1955).

[96] E. Wigner, Characteristic vectors of bordered matrices of infinite dimensions II, Ann. Math. **65**, 203 (1957).

[97] E. P. Wigner, Results and theory of resonance absorption, Oak Ridge Nat'l. Lab. Report ORNL **2309**, 59 (1957).

[98] E. Wigner, On the distribution of the roots of certain symmetric matrices, Ann. Math. **67**, 325 (1958).

[99] L. Landau and Y. Smorodmsky, Lectures on the theory of the atomic nucleus, State Tech Theo Lit Press (Moscow) (1955), translation (Consultants Bureau, New-York 1958), p 55.

[100] M. E. Berbenni-Bitsch, S. Meyer, A. Schäfer, J. J. M. Verbaarschot, and T. Wettig, Microscopic Universality in the Spectrum of the Lattice Dirac Operator, Phys. Rev. Lett. **80**, 1146 (1998).

[101] P. H. Damgaard, U. M. Heller, R. Niclasen, and K. Rummukainen, Low-lying eigenvalues of the QCD Dirac operator at finite temperature, Nucl. Phys. B **583**, 347 (2000).

[102] F. Farchioni, P. de Forcrand, I. Hip, C. B. Lang, and K. Splittorff, Microscopic universality and the chiral phase transition in two flavor QCD, Phys. Rev. D **62**, 014503 (2000).

[103] A. Deuzeman, U. Wenger, and J. Wuilloud, Spectral properties of the Wilson Dirac operator in the ε-regime, JHEP **12**, 109 (2011).

[104] M. Kieburg and T. R. Würfel, Shift of symmetries of naive and staggered fermions in QCD-like lattice theories, Phys. Rev. D **96**, 034502 (2017).

[105] F. J. Dyson, The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics, J. Math. Phys. **3**, 1199 (1962).

[106] A. Altland and M. R. Zirnbauer, Nonstandard symmetry classes in mesoscopic normal-superconducting hybrid structures, Phys. Rev. B **55**, 1142 (1997).

[107] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Classification of topological insulators and superconductors in three spatial dimensions, Phys. Rev. B **78**, 195125 (2008).

[108] J. J. M. Verbaarschot, The Spectrum of the QCD Dirac Operator and Chiral Random Matrix Theory: The Threefold Way, Phys. Rev. Lett. **72**, 2531 (1994).

[109] A. M. García García, L. Sá, and J. J. M. Verbaarschot, Symmetry classification and universality in non-Hermitian many-body quantum chaos by the Sachdev-Ye-Kitaev model (2021), ArXiv:2110.03444.

[110] T. Kanazawa and T. Wettig, Complete random matrix classification of SYK models with $\mathcal{N} = 0$, 1 and 2 supersymmetry, JHEP **09**, 050 (2017).

[111] M. V. Berry and M. Tabor, Level clustering in the regular spectrum, Proc. R. Soc. A **356**, 375 (1977).

[112] V. Oganesyan and D. A. Huse, Localization of interacting fermions at high temperature, Phys. Rev. B **75**, 155111 (2007).

[113] N. Rosenzweig and C. E. Porter, Repulsion of energy levels in complex atomic spectra, Phys. Rev. **120**, 1698 (1960).

[114] J. French and S. Wong, Some random-matrix level and spacing distributions for fixed-particle-rank interactions, Phys. Lett. B **35**, 5 (1971).

[115] O. Bohigas and J. Flores, Spacing and individual eigenvalue distributions of two-body random Hamiltonians, Phys. Lett. B **35**, 383 (1971).

[116] T. Guhr, A. Müller–Groeling, and H. A. Weidenmüller, Random-matrix theories in quantum physics: Common concepts, Phys. Rep. **299**, 189 (1998).

[117] G. Akemann, J. Baik, and P. Di Francesco, *Random Matrices* (Academic Press, Amsterdam, 2004), 3rd ed.

[118] G. Akemann, J. Baik, and P. Di Francesco, *The Oxford Handbook of Random Matrix Theory* (Oxford University Press, Oxford, 2011).

[119] F. Haake, S. Gnutzmann, and M. Kuś, *Quantum Signature of Chaos* (Springer, Berlin, 2019).

[120] O. Bohigas and J. Flores, Two-body random Hamiltonian and level density, Phys. Lett. B **34**, 261 (1971).

[121] J. French and S. Wong, Validity of random matrix theories for many-particle systems, Phys. Lett. B **33**, 449 (1970).

[122] K. K. Mon and J. B. French, Statistical properties of many-particle spectra, Ann. Phys. (NY) **95**, 90 (1975).

[123] L. Benet, T. Rupp, and A. Weidenmuller, Nonuniversal Behavior of the k-Body Embedded Gaussian Unitary Ensemble of Random Matrices, Phys. Rev. Lett. **87**, 010601 (2001).

[124] T. A. Brody, J. Flores, J. B. French, P. A. Mello, A. Pandey, and S. S. M. Wong, Random-matrix physics: Spectrum and strength fluctuations, Rev. Mod. Phys. **53**, 385 (1981).
[125] V. K. B. Kota, Embedded random matrix ensembles for complexity and chaos in finite interacting particle systems, Phys. Rep. **347**, 223 (2001).
[126] V. K. B. Kota, *Embedded Random Matrix Ensembles in Quantum Physics* (Spriinger, Heidelberg, 2014), lecture Notes in Physics 884.
[127] S. Sachdev and J. Ye, Gapless Spin-Fluid Ground State in a Random Quantum Heisenberg Magnet, Phys. Rev. Lett. **70**, 3339 (1993).
[128] J. Maldacena and D. Stanford, Remarks on the Sachdev-Ye-Kitaev model, Phys. Rev. D **94**, 106002 (2016).
[129] A. M. García García and J. J. M. Verbaarschot, Spectral and thermodynamic properties of the Sachdev-Ye-Kitaev model, Phys. Rev. D **94**, 126010 (2016).
[130] A. M. García García and J. J. M. Verbaarschot, Analytical spectral density of the Sachdev-Ye-Kitaev model at finite N, Phys. Rev. D **96**, 066012 (2017).
[131] A. M. García García, Y. Jia, and J. J. M. Verbaarschot, Exact moments of the Sachdev-Ye-Kitaev model up to order $1/N^2$, JHEP **04**, 146 (2018).
[132] L. Erdős and D. Schröder, Phase transition in the density of states of quantum spin glasses, Math. Phys., Anal. Geom. **17**, 441 (2014).
[133] M. Vyas and V. K. B. Kota, Quenched many-body quantum dynamics with k-body interactions using q-Hermite polynomials, J. Stat. Mech. **1910**, 103103 (2019).
[134] F. Borgonovi and F. M. Izrailev, in *Nuclei and Mesoscopic Physics 2017*, American Institute of Physics Conference Series, Vol. 1912 (2017), p. 020003.
[135] L. F. Santos and E. J. Torres-Herrera, in *Thermodynamics in the Quantum Regime: Fundamental Aspects and New Directions*, Vol. 195, edited by F. Binder, L. A. Correa, C. Gogolin, J. Anders, and G. Adesso (2018), p. 457.
[136] P. J. Forrester, *Log-Gases and Random Matrices (LMS-34)* (Princeton University Press, Princeton, 2010).
[137] C. Giraud, *Introduction to Highdimensional Statistics* (CRC Press, London, 2015), monographs on Statistics and Applied Probability 139.
[138] R. Couillet, *Random Matrix Methods for Wireless Communications* (Cambridge University Press, Cambridge, 2011).
[139] V. A. Marčcenko and L. A. Pastur, Distribution of eigenvalues for some sets of random matrices, Math. USSR-Sbornik **1**, 457 (1967).
[140] C. E. Porter and R. G. Thomas, Fluctuations of nuclear reaction widths, Phys. Rev. **104**, 483 (1956).
[141] B. Collins and I. Nechita, Random matrix techniques in quantum information theory, J. Math. Phys. **57**, 015215 (2016).
[142] M. J. W. Hall, Random quantum correlations and density operator distributions, Phys. Lett. A **242**, 123 (1998).
[143] H.-J. Sommers and K. Zyczkowski, Statistical properties of random density matrices, J. Phys. A Math. Gen. **37**, 8457 (2004).
[144] D. Bures, An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^*-algebras, Trans. Amer. Math. Soc. **135**, 199 (1969).
[145] K. Zyczkowski and H.-J. Sommers, Induced measures in the space of mixed quantum states, J. Phys. A Math. Gen. **34**, 7111 (2001).
[146] H.-J. Sommers and K. Zyczkowski, Bures volume of the set of mixed quantum states, J. Phys. A Math. Gen. **36**, 10083 (2003).
[147] V. A. Osipov, H. J. Sommers, and K. Życzkowski, Random Bures mixed states and the distribution of their purity, J. Phys. A Math. Gen. **43**, 055302 (2010).
[148] P. J. Forrester and M. Kieburg, Relating the Bures measure to the Cauchy two-matrix model, Commun. Math. Phys. **342**, 151 (2016).
[149] L. Wei and N. Witte, Quantum interpolating ensemble: Average entropies and orthogonal polynomials (2021), ArXiv:2103.04231.
[150] E. Lubkin, Entropy of an n-system from its correlation with a k-reservoir, J. Math. Phys. **19**, 1028 (1978).
[151] S. Lloyd and H. Pagels, Complexity as thermodynamic depth, Ann. Phys. **188**, 186 (1988).
[152] D. N. Page, Average Entropy of a Subsystem, Phys. Rev. Lett. **71**, 1291 (1993).
[153] S. K. Foong and S. Kanno, Proof of Page's Conjecture on the Average Entropy of a Subsystem, Phys. Rev. Lett. **72**, 1148 (1994).
[154] J. Sánchez-Ruiz, Simple proof of Page's conjecture on the average entropy of a subsystem, Phys. Rev. E **52**, 5653 (1995).
[155] S. Sen, Average Entropy of a Subsystem, Phys. Rev. Lett. **77**, 1 (1996).
[156] C. Rovelli, *Quantum Gravity* (Cambridge University Press, Cambridge, 2004).
[157] E. Bianchi and P. Donà, Typical entanglement entropy in the presence of a center: Page curve and its variance, Phys. Rev. D **100**, 105010 (2019).
[158] L. G. Valiant, Quantum circuits that can be simulated classically in polynomial time, SIAM J. Comput. **31**, 1229 (2002).
[159] E. Bianchi, L. Hackl, and M. Kieburg, Page curve for fermionic Gaussian states, Phys. Rev. B **103**, L241118 (2021).
[160] Y. Chen, D.-Z. Liu, and D.-S. Zhou, Smallest eigenvalue distribution of the fixed-trace Laguerre beta-ensemble, J. Phys. A Math. Gen. **43**, 315303 (2010).
[161] D.-Z. Liu and D.-S. Zhou, Local statistical properties of Schmidt eigenvalues of bipartite entanglement for a random pure state (2009), ArXiv:0912.3999.
[162] C. Nadal, S. N. Majumdar, and M. Vergassola, Statistical distribution of quantum entanglement for a random bipartite state, J. Stat. Phys. **142**, 403 (2011).
[163] K. Zyczkowski and H.-J. Sommers, Truncations of random unitary matrices, J. Phys. A Math. Gen. **33**, 2045 (2000).
[164] DLMF, *NIST Digital Library of Mathematical Functions*, http://dlmf.nist.gov/, Release 1.1.3 of 2021-09-15 (2021), f. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.
[165] P. Vivo, M. P. Pato, and G. Oshanin, Random pure states: Quantifying bipartite entanglement beyond the linear statistics, Phys. Rev. E **93**, 052106 (2016).

[166] L. Wei, Proof of Vivo-Pato-Oshanin's conjecture on the fluctuation of von Neumann entropy, Phys. Rev. E **96**, 022106 (2017).

[167] Our computation of the variance for Gaussian states at fixed particle number presented in Appendix D shows how to deal with the double sums, and can also be used in the general setting. Basically, one needs to replace the Jacobi polynomials and their corresponding weight by the Laguerre polynomials and the weight function $x^{d_B-d_A}e^{-x}$.

[168] A. Serafini, O. Dahlsten, D. Gross, and M. Plenio, Canonical and micro-canonical typical entanglement of continuous variable systems, J. Phys. A: Math. Theor. **40**, 9551 (2007).

[169] M. Fukuda and R. Koenig, Typical entanglement for Gaussian states, J. Math. Phys. **60**, 112203 (2019).

[170] R. H. Jonsson, L. Hackl, and K. Roychowdhury, Entanglement dualities in supersymmetry, Phys. Rev. Res. **3**, 023213 (2021).

[171] I. Peschel, Calculation of reduced density matrices from correlation functions, J. Phys. A **36**, L205 (2003).

[172] L. F. Hackl, *Aspects of Gaussian States Entanglement, Squeezing and Complexity* (The Pennsylvania State University, University Park, 2018).

[173] L. Hackl and E. Bianchi, Bosonic and fermionic Gaussian states from Kähler structures, SciPost Phys. Core **4**, 025 (2021).

[174] M. Kieburg, P. J. Forrester, and J. R. Ipsen, Multiplicative convolution of real asymmetric and real antisymmetric matrices, Adv. Pure Appl. Math. **10**, 467 (2019).

[175] Note that Ref. [159] uses a different normalization constant, where the level density is $\rho(x) = R_1(x)$. In contrast, we use the normalization $\int_0^1 R_1(x)dx = V_A$, which is standard in random matrix theory, and the level density is $\rho(x) = R_1(x)/V_A$.

[176] A. Borodin, Biorthogonal ensembles, Nucl. Phys. B **536**, 704 (1998).

[177] N. Witte, private communication (2021).

[178] Y. Huang and L. Wei, Second-order statistics of fermionic Gaussian states, J. Phys. A **55**, 105201 (2022).

[179] C. W. J. Beenakker, Random-matrix theory of quantum transport, Rev. Mod. Phys. **69**, 731 (1997).

[180] P. A. Mello, P. Pereyra, and T. H. Seligman, Information theory and statistical nuclear reactions. I. General theory and applications to few-channel problems, Ann. Phys. (N. Y) **161**, 254 (1985).

[181] D. Bernard and L. Piroli, Entanglement distribution in the quantum symmetric simple exclusion process, Phys. Rev. E **104**, 014146 (2021).

[182] M. R. Zirnbauer, Riemannian symmetric superspaces and their origin in random-matrix theory, J. Math. Phys. **37**, 4986 (1996).

[183] R. Bott, The stable homotopy of the classical groups, Ann. Math. **70**, 313 (1959).

[184] E. Onorati, O. Buerschaper, M. Kliesch, W. Brown, A. H. Werner, and J. Eisert, Mixing properties of stochastic quantum Hamiltonians, Commun. Math. Phys. **355**, 905 (2017).

[185] M. Kieburg, S.-H. Li, J. Zhang, and P. J. Forrester, Cyclic Pólya ensembles on the unitary matrices and their spectral statistics (2020), ArXiv:2012.11993.

[186] K. Liechty and D. Wang, Nonintersecting Brownian motions on the unit circle, Annals of Probability **44**, (2016), ArXiv:1312.7390.

[187] M. Bauer, D. Bernard, and T. Jin, Stochastic dissipative quantum spin chains (I): Quantum fluctuating discrete hydrodynamics, SciPost Phys. **3**, 033 (2017).

[188] M. Bauer, D. Bernard, and T. Jin, Equilibrium fluctuations in maximally noisy extended quantum systems, SciPost Phys. **6**, 45 (2019).

[189] A. M. Rey, I. I. Satija, and C. W. Clark, Noise correlations of hard-core bosons: Quantum coherence and symmetry breaking, J. Phys. B: At. Mol. Opt. Phys. **39**, S177 (2006).

[190] K. He and M. Rigol, Scaling of noise correlations in one-dimensional-lattice-hard-core-boson systems, Phys. Rev. A **83**, 023611 (2011).

[191] L. F. Santos and M. Rigol, Onset of quantum chaos in one-dimensional bosonic and fermionic systems and its relation to thermalization, Phys. Rev. E **81**, 036206 (2010).

[192] L. F. Santos and M. Rigol, Localization and the effects of symmetries in the thermalization properties of one-dimensional quantum systems, Phys. Rev. E **82**, 031130 (2010).

[193] T. LeBlond, D. Sels, A. Polkovnikov, and M. Rigol, Universality in the onset of quantum chaos in many-body systems, Phys. Rev. B **104**, L201117 (2021).

[194] M. Rigol, Breakdown of Thermalization in Finite One-Dimensional Systems, Phys. Rev. Lett. **103**, 100403 (2009).

[195] M. Rigol, Quantum quenches and thermalization in one-dimensional fermionic systems, Phys. Rev. A **80**, 053607 (2009).

[196] B. Kramer and A. MacKinnon, Localization: Theory and experiment, Rep. Prog. Phys. **56**, 1469 (1993).

[197] P. Markoš, Numerical analysis of the Anderson localization, Acta Phys. Slovaca **56**, 561 (2006).

[198] F. Evers and A. D. Mirlin, Anderson transitions, Rev. Mod. Phys. **80**, 1355 (2008).

[199] J. Šuntajs, T. Prosen, and L. Vidmar, Spectral properties of three-dimensional Anderson model, Ann. Phys. **435**, 168469 (2021).

[200] B. Altshuler and B. Shklovskii, Repulsion of energy levels and conductivity of small metal sample, Zh. Eksp. Teor. Fiz. **91**, 220 (1986).

[201] B. Altshuler, I. Zharekeshev, S. Kotochigova, and B. Shklovskii, Repulsion between energy levels and the metal-insulator transition, Zh. Eksp. Teor. Fiz. **94**, 343 (1988).

[202] B. I. Shklovskii, B. Shapiro, B. R. Sears, P. Lambrianides, and H. B. Shore, Statistics of spectra of disordered systems near the metal-insulator transition, Phys. Rev. B **47**, 11487 (1993).

[203] T. Prosen and M. Žnidarič, Is the efficiency of classical simulations of quantum dynamics related to integrability?, Phys. Rev. E **75**, 015202 (2007).

[204] V. Alba, J. Dubail, and M. Medenjak, Operator Entanglement in Interacting Integrable Quantum Systems: The

Case of the Rule 54 Chain, Phys. Rev. Lett. **122**, 250603 (2019).

[205] V. Alba, Diffusion and operator entanglement spreading, Phys. Rev. B **104**, 094410 (2021).

[206] L. F. Santos, A. Polkovnikov, and M. Rigol, Entropy of Isolated Quantum Systems after a Quench, Phys. Rev. Lett. **107**, 040601 (2011).

[207] M. Rigol, Fundamental Asymmetry in Quenches between Integrable and Nonintegrable Systems, Phys. Rev. Lett. **116**, 100601 (2016).

[208] R. Modak, V. Alba, and P. Calabrese, Entanglement revivals as a probe of scrambling in finite quantum systems, J. Stat. Mech. **2020**, 083110 (2020).

[209] P. Hosur, X.-L. Qi, D. A. Roberts, and B. Yoshida, Chaos in quantum channels, J. High Energy Phys. **2016**, 4 (2016).

[210] J. Kudler-Flam, M. Nozaki, S. Ryu, and M. T. Tan, Entanglement of local operators and the butterfly effect, Phys. Rev. Res. **3**, 033182 (2021).

[211] T.-C. Lu and T. Grover, Entanglement transitions as a probe of quasiparticles and quantum thermalization, Phys. Rev. B **102**, 235110 (2020).

[212] We evaluate Eq. (49), where $\varrho(n_A)$ and $\varphi(n_A)$ slightly change from expanding Eq. (140) via a saddle point approximation. This yields the normal distribution $\varrho(n_A)$, with mean $\bar{n}_A = fn$ and variance $\sigma^2 = (1-f)f\,(1+n)n/V$, and $\varphi(n_A)$ in Eq. (53) becomes $\varphi(n_A) = \{n_A \ln(n_A) + f \ln(f) + n\ln[(1+n)/n] + \ln(1+n) - (f+n_A)\ln(f+n_A)\}V + \frac{1}{2}\ln[n_A(f+n_A)/f\,(1+n)n] - \frac{1}{2}\delta_{f,1/2}\delta_{n_A,n/2} + o(1)$ for $n_A \geq n_{\text{crit}}$ with $n_{\text{crit}} = N_{\text{crit}}$ again given by $d_A(N_{\text{crit}}) = d_B(N - N_{\text{crit}})$. For $n_A \leq n_{\text{crit}}$, one needs to apply the symmetry $(n_A, f) \leftrightarrow (n - n_A, 1-f)$. The summand at $N_A = N/2$ reflected by the term $\delta_{n_A,n/2}$ has to be taken as it is and is not integrated. Nevertheless, one can check numerically that it yields a term of order $1/\sqrt{V}$ and is thus subleading in Eq. (143).

[213] H. Fujita, Y. O. Nakagawa, S. Sugiura, and M. Watanabe, Page curves for general interacting systems, JHEP **12**, 112 (2018).

[214] J. De Boer, J. Järvelä, and E. Keski-Vakkuri, Aspects of capacity of entanglement, Phys. Rev. D **99**, 066012 (2019).

[215] I. Ulčakar and L. Vidmar, Tight-binding billiards, arXiv:2206.07078.

[216] B. Bhattacharjee, P. Nandy, and T. Pathak, Eigenstate capacity and Page curve in fermionic Gaussian states, Phys. Rev. B **104**, 214306 (2021).

[217] S. Szalay, Z. Zimborás, M. Máté, G. Barcza, C. Schilling, and Ö. Legeza, Fermionic systems for quantum information people, J. Phys. A: Math. Theor. **54**, 393001 (2021).