Free-fermion Page Curve: Canonical Typicality and Dynamical Emergence

Xie-Hang Yu,1,2 Zongping Gong,1,2 and J. Ignacio Cirac1,2

1Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany
2Munich Center for Quantum Science and Technology, Schellingstraße 4, 80799 München, Germany

We provide further analytical insights into the newly established noninteracting (free-fermion) Page curve, focusing on both the kinematic and dynamical aspects. First, we unveil the underlying canonical typicality and atypicality for random free-fermion states. The former appears for a small subsystem and is exponentially weaker than the well-known result in the interacting case. The latter explains why the free-fermion Page curve differs remarkably from the interacting one when the subsystem is macroscopically large, i.e., comparable with the entire system. Second, we find that the free-fermion Page curve emerges with unexpectedly high accuracy in some simple tight-binding models in long-time quench dynamics. This contributes a rare analytical result concerning quantum thermalization on a macroscopic scale, where conventional paradigms such as the generalized Gibbs ensemble and quasi-particle picture are not applicable.

Introduction.—As a central concept in quantum information science [1], entanglement has been recognized to play vital roles in describing and understanding quantum many-body systems in and out of equilibrium [2–5]. For example, entanglement area laws for ground states of gapped local Hamiltonians enable their efficient descriptions based on tensor networks [6], while their violations may signature quantum phase transitions [7, 8]. The emergence of thermal ensemble from unitary evolution, a process known as quantum thermalization [9], is ultimately attributed to the entanglement generated between a subsystem and the complement [10].

Almost thirty years ago, Page considered the fundamental problem of bipartite entanglement in a fully random many-body system and found a maximal entanglement entropy (EE) up to finite-size corrections [11]. This seminal work was originally motivated by the black-hole information problem [12]. Remarkably, it has been attracted increasing and much broader interest in the past decade, due not only to the new theoretical insights from quantum thermalization [13–21] and quantum chaos [22–24], but also to the practical relevance in light of the rapid experimental development in quantum simulations [25–32]. In particular, the saturation of maximal entropy has been found to be a consequence of canonical typicality [33–35], which means most random states behave locally like the canonical ensemble. This typicality behavior has been argued to emerge in generic interacting many-body systems satisfying the eigenstate thermalization hypothesis [5, 9, 10, 36–38] and can even be rigorously established or ruled out in specific situations [39, 40].

In this Letter, we provide analogous insights into the noninteracting counterpart of Page’s problem. That is, we focus on free fermions or (fermionic) Gaussian states, which are of their own interest in quantum many-body physics, quantum information and computation [41–57]. Somehow surprisingly, in the seemingly simpler noninteracting case, the subsystem-size dependence of averaged EE, which is described by the Page curve pictorially, was not solved until very recently [58–60]. It turns out to be similar to the interacting case for a small subsystem, but differ significantly otherwise. See Fig. 1(b) for an illustration. With the measure concentration results on compact-group manifolds, we establish the correspond-

ing canonical typicality (atypicality) in microscopic (macroscopic) regions for the free-fermion ensemble. Thus we explicitly explain the similarity and difference from the kinematic aspect. In addition, we show that the free-fermion Page curve can be relevant to extremely simple tight-binding models via long-time quench dynamics. By classifying the systems according to their conserved (eigen) mode occupation numbers, we construct two classes of Hamiltonians which can/cannot give rise to a highly similar Page curve. Our finding concerning macroscopic properties which cannot be captured by the generalized Gibbs ensemble or quasi-particle picture and thus goes beyond the conventional paradigm of local thermalization.

Canonical typicality and atypicality.—We start by generalizing the main result in [33] to the random fermionic Gaussian (RFG) ensemble. While [33] already considers possible restrictions, we stress that Gaussianity is inadequate since Gaussian states do not constitute a Hilbert subspace. For simplicity, we consider number-conserving systems with totally N modes occupied by N/2 fermions, i.e., the half-filling case. Compared to the fully random case without number conservation, this setting appears to be more physically comprehensible and experimentally relevant, while displaying exactly the same Page curve [58, 59]. More general ensembles are discussed in Supplemental Material [61].

A pictorial illustration of our setup is shown in Fig. 1(a). Due to Wick’s theorem [62], a fermionic Gaussian state $\rho$ is fully captured by its covariance matrix $C_{j,j'} = \text{Tr}(\rho a_j a_{j'}^\dagger)$. Here $a_j$ is the annihilation operator for mode $j$, which may label, e.g., a lattice site. As the covariance matrix for any RFG-pure state can be related to each other by a unitary transformation, the uniform distribution over this ensemble can be generated at the level of the covariance matrix $\{ C = U C_0 U^\dagger \}$. Here $U$ is taken Haar-randomly over the unitary group $U(N)$ [58, 59] and $C_0$ is an arbitrary reference covariance matrix in the ensemble satisfying $C_0^2 = C_0$ and $\text{Tr} C_0 = N/2$.

An important property of Gaussian states is that their subsystems remain Gaussian. We denote $C_A$ as the $N_A \times N_A$ covariance matrix restricted to subsystem $A$ with $N_A$ modes. The EE $S_A = -\text{Tr}(\rho_A \log_2 \rho_A)$ of the reduced state $\rho_A =$
covariance matrix for RFG ensemble:

allows us to turn the upper bound on the distance averaging ensemble in the thermodynamic limit N → ∞. It is obvious that these two Page curves agree with each other in the microscopic region but show a O(1) deviation in the macroscopic region. The interacting Page curve in the thermodynamic limit is always saturated. (c) The table summarizes the typicality/atypicality property for the Page curves of the RFG and interacting ensembles. Here “Poly.” and “Exp.” indicates polynomial and exponential scalings, respectively.

\[ S_A = -\text{Tr}(C_A \log_2 C_A) - \text{Tr}((I_A - C_A) \log_2 (I_A - C_A)), \]

where \( I_A \) is the identity matrix with dimension \( N_A \).

Our first result is the measure concentration property of the covariance matrix for RFG ensemble:

**Theorem 1** For arbitrary \( \epsilon > 0 \) and subsystem \( A \), the probability that the reduced covariance matrix of a state in the RFG ensemble deviates from the ensemble average satisfies

\[
P(\text{dHS}(C_A, I_A/2) \geq \eta + 2\epsilon) \leq 2e^{-\frac{\epsilon^2}{\eta}} \label{eq:coh2}
\]

and

\[
P(\text{dHS}(C_A, I_A/2) \leq \eta_a - 2\epsilon) \leq 2e^{-\frac{\epsilon^2}{\eta_a}} \label{eq:coh3}
\]

with \( \eta = \sqrt{\frac{N_A^2}{2(N-1)}}, \eta' = \frac{12}{N}, \eta_a = \frac{N^2}{4N^2 + 1}, \eta_a' = \frac{12N}{N} \) and \( \text{dHS}(C, C') = \sqrt{\text{Tr}(C - C')^2} \) being the Hilbert-Schmidt distance.

The proof largely relies on the generalized Levy’s lemma for Riemann manifolds with positive curvature [64–66], which allows us to turn the upper bound on the distance average \( \langle d_{\text{HS}}(C_A, I_A/2) \rangle \) into a probability inequality [61]. From Eq. (2) we can easily see for infinite environments \( N \to \infty \), the local microscopic system will have maximal entropy \( S_A \to N_A \).

We emphasize that in Eq. (2), \( \eta \) and \( \eta' \) only scales polynomially with the (sub)system size. This contrasts starkly with the exponential scaling canonical typicality for random interacting ensemble [33]. Intuitively, this is because in the interacting case, the Hilbert-space dimension scales exponentially with the (sub)system size, which, however, simply equals to the size of the covariance matrix in the free-fermion case. Physically, the Gaussian constraint makes the ensemble only explore a very limited sub-manifold in the entire Hilbert space. This polynomial scaling means that, for a fixed subsystem size \( N_A \), the reduced state still exhibits canonical typicality, while the atypicality is only polynomially suppressed by the environment size. Accordingly, the averaged EE should achieve the maximal value but with a polynomial finite-size correction. In fact, such an exponentially weaker canonical typicality (2) can also be exploited to explain the qualitatively larger variance of the EE for the RFG ensemble, which is \( O(N^{-2}) \) in comparison to \( e^{-O(N)} \) in the interacting case [58, 59, 61].

On the other hand, if the subsystem is macroscopically large, meaning that \( f = N_A/N \) is \( O(1) \), the concentration inequality (2) becomes meaningless since \( \eta \) is \( O(\sqrt{N}) \), the same order as the Hilbert-Schmidt norm of \( C_A \). Instead, we may take \( \epsilon = O(N^{-\alpha}) \) with \( \alpha \in (0, 1) \) in Eq. (3), finding that the majority of the reduced covariance matrix differs significantly from the ensemble average. In other words, the RFG ensemble exhibits canonical atypicality in this case. In particular, this result implies an \( O(1) \) deviation in the EE density from the maximal value. We recall that, in stark contrast, the canonical typicality for interacting states is exponentially stronger and persists even on any macroscopic scale with \( f < 1/2 \).

The above discussions can be made more straightforward by considering the measure concentration property of \( S_A \). By bounding \( S_A \) using \( d_{\text{HS}}(C_A, I_A/2) \) from both sides, we obtain [61]

\[
P(S_A \leq N_A - \epsilon) \leq 2e^{-\frac{(\epsilon-\xi)^2}{\xi}} \quad \forall \epsilon > 0 \tag{4}
\]

in the microscopic region and

\[
P(S_A \geq N_A - \xi_a + \epsilon) \leq 2e^{-\frac{\epsilon^2}{\xi_a}} \quad \forall \epsilon > 0 \tag{5}
\]

in the macroscopic region. Here \( \xi = \sqrt{\frac{2N_A}{N-1}}, \xi' = \sqrt{\frac{N}{N}} \), \( \xi_a = \frac{N_A^2}{2(N-1)} \) and \( \xi_a' = \frac{12N^2}{N} \). Note that Eq. (4) also becomes meaningless in the macroscopic region since \( \xi^2 \) will be comparable with \( N_A \). Choosing \( \epsilon = (\xi + O(N^{-\alpha/2}))^2 \) for Eq. (4) and \( \epsilon = O(N^{-\alpha}) \) in Eq. (5) with \( \alpha \in (0, 1) \), we fully explain the microscopic similarity and macroscopic difference between the Page curves for the RFG and interacting ensembles. See Fig. 1(b) and (c).
The considerable deviation between the theoretical result and the dynamical Page curve near $f = N / f$ and highly experimentally accessible. See Fig. 2(a-b) for a schematic illustration. The dynamical Page curve is formally given by $S(\rho_A(t))$, where $\rho_A = \text{Tr}_A[e^{-iHt}|\Psi_0\rangle\langle\Psi_0|e^{iHt}]$ and $f(t) = \lim_{T \to \infty} T^{-1} \int_0^T dt f(t)$ denotes the long-time average. It is worth mentioning that the dynamical Page curve is ensured to be concave by translation invariance, as a result of the strong subadditivity of quantum entropy [69].

We primarily focus on the minimal model, i.e., a one-dimensional lattice with nearest-neighbor hopping:

$$H_0 = \sum_j a_j^\dagger a_{j+1} + \text{H.c.},$$

whose band dispersion reads $E_k = 2 \cos k$. We believe that the exact results for the large (spatiotemporal) scale dynamical behaviors of this fundamental model are interesting on their own. Moreover, our method and results actually apply to much broader situations, as will soon become clear below.

Surprisingly, despite the additional translation-invariant and energy-conserving constraints compared to the RFG ensemble, this minimal model (6) turns out to give rise to a dynamical Page curve extremely close to that for the RFG ensemble (see blue and red curves in Fig. 2(c)). To gain some analytic insights, we perturbatively expand the entropy expression (1) around $C_A = \frac{I_A}{2}$, obtaining

$$S_A(t) = N_A - \sum_{n=1}^\infty \frac{\text{Tr}(2C_A(t) - I_A)^{2n}}{2n(2n - 1)!} \ln 2.$$  

Thanks to the translational invariance, $C_A(t)$ can be related to the block-diagonal momentum-space covariance matrix $\tilde{C}(t) = \bigoplus_k \tilde{C}_k(t)$ via $C_A(t) = \Pi_A U(t) \tilde{C}(t) U(t)^\dagger \Pi_A$. Here, $U(t) = \text{e}^{iHt}$.
Theorem 2 The RFG ensemble-like dynamical Page curve (8) emerges for a period-2 short-range free-fermion Hamiltonian if and only if the conserved mode occupation numbers are 1/2.

Discussions.—It is well-known that the generalized Gibbs ensemble (GGE) characterizes the local thermalization of integrable systems including free fermions [70–74]. However, in principle, GGE only predicts the expectation values of observables, which do not include the entropy. Note that the former (latter) is linear (nonlinear) in \( \rho_A \) and thus commutes (does not commute) with time average. Moreover, we also study the macroscopic scale, which can not be captured by GGE as well as its recently proposed refined version [75] concerning the purified subsystem by measuring the complement [76]. In this sense, our study goes well beyond the conventional paradigm of quantum thermalization in integrable systems, pointing out especially the highly nontrivial behaviors on the macroscopic level, where typicality may completely break down.

Finally, let us mention the relation between our strategy and the quasi-particle picture, which is widely used to calculate EE growth [74, 77–83]. It turns out this picture fails to reproduce the dynamical Page curve. Under the period-2 boundary condition, the quasi-particle picture predicts \( S_A = N - N_A^2/4 \), which is as small as about \( 10^{-3} \) even for \( f \) near 1/2.

Interestingly, if we add a perturbation \( H_1 \) to Eq. (6), as long as \( H_1 \) is period-2 and only includes odd-range hopping, as represented by Fig. 2(a), the dynamical Page curve can be analytically demonstrated to be the same as Eq. (8) in the thermodynamic limit, as the same Feynman rules apply [61]. One example is \( H_1 = J (\sum_{j: \text{even}} a_j^+ a_j + 2m+1 - \sum_{j: \text{odd}} a_j^+ a_j + 2n) + \text{H.c.} \) for arbitrary \( J \) and integer \( m \). Thus, we have defined another ensemble of fermionic Gaussian states by dynamical evolution, which covers a wide class of Hamiltonians and this ensemble has remarkably similar Page curve as the RFG ensemble.

However, if \( H_1 \) includes even-range hopping, as represented by Fig. 2(b) the dynamical Page curve will be very different, as shown in Fig. 2(d). This can be easily explained with the canonical typicality property proved above: for this class of Hamiltonians, their conserved (even) mode occupation number \( n_k \) deviates from the average value of RFG ensemble, which is \( 1/2 \). Thus, the dynamical ensemble is naturally "atypical" even for microscopic scale because the local conserved observable is constructed from mode occupation numbers [70]. This result implies the reduced state on a small subsystem deviates considerably from being maximally mixed so that the tangent slope of the dynamical Page curve at \( f = 0 \) is well below that for the RFG ensemble. In contrast, one can show that all the conserved mode occupation number for the class of Hamiltonians mentioned in the last paragraph are \( \frac{1}{2} \).

All the observations above constitute our second main result:

Theorem 2 The RFG ensemble-like dynamical Page curve (8) emerges for a period-2 short-range free-fermion Hamiltonian if and only if the conserved mode occupation numbers are 1/2.
supported by the Max-Planck-Harvard Research Center for Quantum Optics (MPHQ). J.I.C. acknowledges support by the EU Horizon 2020 program through the ERC Advanced Grant QENOCOBA No. 742102.

Note added.—While finalizing this manuscript, a related work by Issue et al. appeared in Ref. [87], which reported the typicality for random bosonic Gaussian states.
[67] D. Bernard and L. Piroli, Phys. Rev. E 104, 014146 (2021).
[68] B. C. Dias, M. Haque, P. Ribeiro, and P. McClarty (2021), arXiv:2102.09846.
[69] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac, Phys. Rev. Lett. 100, 070502 (2008).
[70] T. Ishii and T. Mori, Phys. Rev. E 100, 012139 (2019).
[71] M. Rigol, V. Dunjko, V. Vyarovskiy, and M. Olshanii, Phys. Rev. Lett. 98, 050405 (2007).
[72] A. C. Cassidy, C. W. Clark, and M. Rigol, Phys. Rev. Lett. 106, 140405 (2011).
[73] T. Langen, S. Erne, R. Geiger, B. Rauer, T. Schweigler, M. Kuhnert, W. Rohringer, I. E. Mazets, T. Gasenzer, and J. Schmiedmayer, Science 348, 207 (2015).
[74] F. H. L. Essler and M. Fagotti, J. Stat. Mech. 2016, 064002 (2016).
[75] M. Lucas, L. Piroli, J. D. Nardis, and A. D. Luca (2022), arXiv:2207.13628.
[76] W. W. Ho and S. Choi, Phys. Rev. Lett. 128, 060601 (2022).
[77] C.-A. Chen, S. Khlebnikov, and C.-L. Hung, Phys. Rev. Lett. 127, 060404 (2021).
[78] P. Jurcevic, B. P. Lanyon, P. Hauke, C. Hempel, P. Zoller, R. Blatt, and C. F. Roos, Nature 511, 202 (2014).
[79] O. A. Castro-Alvaredo, B. Doyon, and T. Yoshimura, Phys. Rev. X 6, 041065 (2016).
[80] F. Calabrese and J. Cardy, J. Stat. Mech. 2005, P04010 (2005).
[81] M. Fagotti and F. Calabrese, Phys. Rev. A 78, 010306 (2008).
[82] B. Bertini, M. Fagotti, L. Piroli, and F. Calabrese, J. Phys. A 51, 39LT01 (2018).
[83] B. Bertini, E. Tartaglia, and F. Calabrese, J. Stat. Mech. 2018, 063104 (2018).
[84] A. Altland and M. R. Zirnbauer, Phys. Rev. B 55, 1142 (1997).
[85] S. Murciano, P. Calabrese, and L. Piroli, Phys. Rev. D 106, 046015 (2022).
[86] P. H. C. Lau, T. Noumi, Y. Takii, and K. Tamaoka (2022), arXiv:2206.09633.
[87] J. T. Iosue, A. Ehrenberg, D. Hangleiter, A. Deshpande, and A. V. Gorshkov (2022), arXiv:2209.06838.
[88] S. Lloyd, Phys. Rev. A 55, 1613 (1997).
[89] M. Hayashi, Group Representation for Quantum Theory (Springer International Publishing, 2017).
[90] P. Stein and R. Shakarchi, Fourier Analysis: An Introduction (Princeton University Press, 2003).
[91] R. Kress, Numerical Analysis, Vol. 181 (Springer New York, 1998).
[92] Noting that in higher $n-$expansion of $\text{Tr}X_A^{2n}$, there will also be contributions to the entropy density $\frac{N}{N_A}$ at the order of $O(f)$.
[93] E. Stein and R. Shakarchi, Fourier Analysis: An Introduction (Princeton University Press, 2003).
[94] Another illustration for general cases can be obtained with the techniques in Subsec. III C.
[95] V. Alba and P. Calabrese, SciPost Physics 4, 017 (2018).
Supplementary Information

Contents

References 5

I. Proof of Canonical Typicality/Atypicality for the RFG ensemble S1

II. Some Applications of Measure Concentration Typicality S3
   A. Measure concentration property for entropy S3
   B. Upper bound on the variance of entropy S3

III. Detailed Calculation of the Dynamical Page Curves S4
   A. Calculation for the minimal model S4
      Second order in $X_A$ S5
      Third order in $X_A$ S5
      Fourth order in $X_A$ S6
   B. General Feynman Rules S6
   C. Proof of Theorem 2 S8
   D. Generalization to the atypical Page curves S9

IV. Calculation of entanglement Entropy in the Quasi-Particle Picture S11

In this Supplemental Material, we provide detailed proof of Theorem 1 and 2 in the main text. We also provide the calculations of other results and conclusions in the main text and discuss their generalization.

I. Proof of Canonical Typicality/Atypicality for the RFG ensemble

In this section, we consider number conserving fermionic Gaussian ensemble with $N$ modes occupied by $m$ fermions. The notation follows the main text. In particular, $\langle \cdots \rangle$ is used to denote the average value over the ensemble. The covariance matrix of the subsystem $A$ for a particular random Gaussian state is

$$ C_A = \Pi_A U C_0 U^\dagger \Pi_A^\dagger, \quad (S.1) $$

where $\Pi_A$ is the projection operator on the subsystem with size $N_A \times N$, $U$ is taken Haar randomly over $U(N)$ and $C_0$ satisfies $C_0^2 = C_0$ and $\text{Tr} C_0 = m$. In the following, the distance between two matrices is measured by Hilbert-Schmidt distance $d_{HS}$. We define a function $f: U(N) \to \mathbb{R}$ as

$$ f(U) = d_{HS}(\Pi_A U C_0 U^\dagger \Pi_A^\dagger, \langle C_A \rangle). \quad (S.2) $$

It is easy to check that $f$ is Lipschitz continuous with constant 2:

$$ |f(U_1) - f(U_2)| \leq d_{HS}(U_1 C_0 U_1^\dagger, U_2 C_0 U_2^\dagger) $$
$$ \leq d_{HS}(U_1 C_0 U_1^\dagger, U_1 C_0 U_2^\dagger) + d_{HS}(U_1 C_0 U_2^\dagger, U_2 C_0 U_2^\dagger) $$
$$ = \|C_0(U_1^\dagger - U_2^\dagger)\|_{HS} + \|(U_1 - U_2)C_0\|_{HS} $$
$$ \leq 2d_{HS}(U_1, U_2). $$

The generalized Levy’s lemma [64–66] states that for any Lipschitz continuous function over some Riemann manifolds with positive curvature, its values are concentrated around the mean one. For the unitary group, we have

$$ \mathbb{P}(|f(U) - \langle f(U) \rangle| \geq \epsilon) \leq 2e^{-\frac{\epsilon^2}{d_{HS}^2}}, \quad (S.3) $$
where $l$ is the Lipschitz constant. In what follows, we will bound $\langle f(U) \rangle = \langle d_{HS}(P A U C_0 U^\dagger U^\dagger, \{C_A\}) \rangle$. First, since $\langle C_A \rangle = \Pi_A \int d_H(U) U C_0 U^\dagger \Pi_A^\dagger$ is invariant under any unitary on $U(N_A)$, according to Schur’s lemma, $\langle C_A \rangle = \frac{m}{N} I_A$ and

$$\langle \Pi_A U C_0 U^\dagger \Pi_A^\dagger - \langle C_A \rangle \rangle_{HS} \leq \sqrt{\langle \Pi_A U C_0 U^\dagger \Pi_A^\dagger - \langle C_A \rangle \rangle_{HS}^2} = \sqrt{\langle \text{Tr}(\Pi_A U C_0 U^\dagger \Pi_A^\dagger - \langle C_A \rangle) \rangle^2} \leq \sqrt{\langle \text{Tr}(\Pi_A U C_0 U^\dagger \Pi_A^\dagger)^2 \rangle} - \frac{m^2}{N^2 N_A}.$$

Next, we need to calculate $\langle \text{Tr}(\Pi_A U C_0 U^\dagger \Pi_A^\dagger)^2 \rangle$. The idea is similar as in [33] and originally comes from random quantum channel coding [88]: we introduce another reference space $R'$ which has the same dimension as the original total system $R$. The following equation holds:

$$\langle \text{Tr}(\Pi_A U C_0 U^\dagger \Pi_A^\dagger)^2 \rangle = \int d_H(U) \text{Tr}[(\Pi_A \otimes \Pi_A)(U C_0 U^\dagger \otimes U C_0 U^\dagger)\text{SWAP}_{RR'}(\Pi_A^\dagger \otimes \Pi_A^\dagger)],$$

where $\text{SWAP}_{RR'}$ is the SWAP operation between the original system $R$ and the reference one $R'$. From Schur-Weyl duality [89], we obtain

$$\int d_H(U)(U C_0 U^\dagger \otimes U C_0 U^\dagger)\text{SWAP}_{RR'} = \alpha I_{R'R'} + \beta \text{SWAP}_{RR'}.$$  

(S.4)

Now, for simplicity, we can take $C_0 = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$. The following relations hold

$$\text{Tr}\text{SWAP}_{RR'} = N,$$

$$\text{Tr}[(U \otimes U)(C_0 \otimes C_0)(U^\dagger \otimes U^\dagger)\text{SWAP}_{RR'}] = \text{Tr}[(U \otimes U)(C_0 \otimes C_0)\text{SWAP}_{RR'}(U^\dagger \otimes U^\dagger)] = \text{Tr}C_0^2 = m,$$

$$\text{Tr}[(U \otimes U)(C_0 \otimes C_0)(U^\dagger \otimes U^\dagger)] = m^2.$$

The trace of Eq. (S.4) gives $N^2 \alpha + N \beta = m$. Multiplying Eq. (S.4) by $\text{SWAP}_{RR'}$ and tracing it, we have $N \alpha + N^2 \beta = m^2$.

Solving the equations leads to $\alpha = \frac{N^2 - m^2}{N(N^2 - 1)}$ and $\beta = \frac{N^2 - m}{N(N^2 - 1)}$. As a result,

$$\langle \text{Tr}(\Pi_A U C_0 U^\dagger \Pi_A^\dagger)^2 \rangle = \text{Tr}[(\Pi_A \otimes \Pi_A)(\alpha I_{R'R'} + \beta \text{SWAP}_{RR'})(\Pi_A^\dagger \otimes \Pi_A^\dagger)] = N \alpha^2 + \beta N_A.$$  

(S.5)

Assuming that in the thermodynamic limit $N \to \infty$, the density of charge $\frac{m}{N}$ is fixed as $O(1)$, we obtain

$$\langle f(U) \rangle^2 \leq \frac{m N_A^2}{N(N - 1)} \sim O\left(\frac{N_A^2}{N} \right)$$

and the typicality

$$P \left( d_{HS}(C_A, \frac{m}{N} I_A) \geq 2e + \sqrt{\frac{m N_A^2}{N(N - 1)}} \right) \leq 2e^{-\frac{N_A^2}{2N + 2}}.$$  

(S.6)

For the other direction, we take $f(U) = d_{HS}^2(C_A, \langle C_A \rangle)$, which is also Lipschitz continuous with constant calculated as

$$|f(U_1) - f(U_2)| \leq 2(d_{HS}(P A U_1 C_0 U^\dagger U^\dagger, \{C_A\}) + d_{HS}(P A U_2 C_0 U^\dagger U^\dagger, \{C_A\}))d_{HS}(U_1, U_2) \leq 4 \sqrt{N_A \frac{1 - m}{N}} d_{HS}(U_1, U_2).$$  

(S.7)

Here we assume $m \leq \frac{N}{2}$ due to the particle-hole symmetry (otherwise, we may replace $1 - \frac{m}{N}$ by $\frac{m}{N}$). According to Eq. (S.5), we obtain

$$\langle f(U) \rangle = \langle \text{Tr}(\Pi_A U C_0 U^\dagger \Pi_A^\dagger)^2 \rangle - \frac{m^2}{N^2 N_A} \geq \frac{(N - m)m N_A^2}{N^2(N + 1)}.$$  

(S.8)
If $\frac{N_A}{N}$ and $\frac{N}{2}$ are both fixed as $O(1)$ in the thermodynamic limit, this formula scales linear with $N$. Applying generalized Levy’s lemma leads to

$$
\mathbb{P} \left( d_{HS}^2 \left( C_A, \frac{m}{N} I_A \right) \leq \frac{(N - m)mN^2}{N^2(N + 1)} - 4\sqrt{N_A} \left( 1 - \frac{m}{N} \right) \epsilon \right) \leq 2e^{-\frac{N}{2}\epsilon^2}.
$$

(S.9)

For example, if we choose $\epsilon \sim O(N^{1/4})$, the above inequality means that $C_A$ will deviate from its ensemble average by an $O(N)$ factor with almost unit probability. This is the atypicality discussed in the main text.

**II. SOME APPLICATIONS OF MEASURE CONCENTRATION TYPICALITY**

**A. Measure concentration property for entropy**

In this subsection, we will use Eq. (S.6) and Eq. (S.9) to derive the measure concentration typicality/atypicality for subsystem entropy. For simplicity the half filling condition is assumed. The eigenvalues of $C_A$ are denoted as $\{\frac{1}{2} + \lambda_i, i \in \{1, \cdots, N_A\}\}$ with $\sum_{i=1}^{N_A} \lambda_i^2 = d_{HS}(C_A, \frac{I_A}{2})$ and $\lambda_i \in [-\frac{1}{2}, \frac{1}{2}]$.

If the subsystem is microscopically small, we know the typicality of entropy follows by noting that

$$
S_A = \sum_{i=1}^{N_A} H \left( \frac{1}{2} + \lambda_i, \frac{1}{2} - \lambda_i \right) = \sum_{i=1}^{N_A} \left[ 1 - \sum_{n=1}^{\infty} \frac{(2\lambda_i)^{2n}}{2n(2n - 1) \ln 2} \right] \geq N_A - 4d_{HS}^2 \left( C_A, \frac{I_A}{2} \right),
$$

(S.10)

where we replace $(2\lambda_i)^{2n}$ by $(2\lambda_i)^2$ in the last inequality since $(2\lambda_i)^2 \leq 1$. Here $H(p_0, p_1) = -p_0 \log_2 p_0 - p_1 \log_2 p_1$ is the Shannon entropy. Combined with Eq. (S.6) we obtain

$$
\mathbb{P} (N_A - S_A \geq x) \leq \mathbb{P} \left( d_{HS} \left( C_A, \frac{I_A}{2} \right) \geq \frac{\sqrt{x}}{2} \right)
\leq 2 \exp \left( -\frac{N}{48} \left( \frac{\sqrt{x}}{2} - \sqrt{\frac{N_A^2}{2(N + 1)}} \right)^2 \right), \quad x > \frac{2N^2}{N - 1};
\leq 1, \quad x \leq \frac{2N^2}{N - 1}.
$$

(S.11)

As long as $N \gg N_A^2$, we conclude the subsystem entropy will be nearly maximal.

For the other direction, if $N_A$ is macroscopically large, we can upper bound the lhs of Eq. (S.10) by

$$
S_A \leq N_A - \frac{2}{\ln 2} d_{HS}^2 \left( C_A, \frac{I_A}{2} \right).
$$

(S.12)

Following the atypicality of $d_{HS}^2(C_A, \frac{I_A}{2})$ in Eq. (S.9), the subsystem entropy density will show an $O(1)$ deviation from the maximal value:

$$
\mathbb{P} \left( S_A \geq N_A - \frac{N_A^2}{2 \ln 2 (N + 1)} + \frac{2}{\ln 2} \epsilon \right) \leq \mathbb{P} \left( d_{HS}^2 \left( C_A, \frac{I_A}{2} \right) \leq \frac{N_A^2}{4(N + 1) - \epsilon} \right) \leq 2e^{-\frac{N}{2}\epsilon^2}.
$$

(S.13)

We may take $\epsilon \sim O(N^\alpha)$ for arbitrary $\alpha \in (0, 1)$, finding that the majority of subsystem entropy will be comparable or smaller than $N_A - \frac{N_A^2}{2 \ln 2 (N + 1)}$. This clearly illustrates the difference of the Page curve for the RFG ensemble from the interacting one.

**B. Upper bound on the variance of entropy**

At the end of this section, we will discuss the variance of entropy in microscopic region. Here the half filling condition is also assumed. From Eq. (S.11), we obtain

$$
\mathbb{P} ((S_A - N_A)^2 \geq x) \leq \begin{cases} 
1, & x \leq \frac{4N_A^4}{(N - 1)^2}; \\
2 \exp \left[ -\frac{N}{48} \left( \frac{x^{1/2}}{2} - \sqrt{\frac{N_A^2}{2(N + 1)}} \right)^2 \right], & x > \frac{4N_A^4}{(N - 1)^2}.
\end{cases}
$$
Therefore

$$
\langle (S_A - N_A)^2 \rangle = \int P((S_A - N_A)^2 \geq x) \, dx
\leq \frac{4N_A^4}{(N-1)^2} + 2 \int_{\frac{4N_A^4}{(N-1)^2}}^{\infty} dx \exp \left[ -\frac{N}{48} \left( \frac{x^2}{2} - \sqrt{\frac{N_A^2}{2(N-1)}} \right)^2 \right].
$$

For the last line, we can change the integral variable into $t = \sqrt{N} \left( \frac{x^2}{2} - \sqrt{\frac{N_A^2}{2(N-1)}} \right)$, obtaining

$$
2 \int_{\frac{4N_A^4}{(N-1)^2}}^{\infty} dx \exp \left[ -\frac{N}{48} \left( \frac{x^2}{2} - \sqrt{\frac{N_A^2}{2(N-1)}} \right)^2 \right] = \frac{128}{\sqrt{N}} \int_0^{\infty} dt \left( t + \sqrt{\frac{N_A^2}{2(N-1)}} \right)^3 e^{-\frac{t^2}{48}} = \frac{128}{N^2} \int_0^{\infty} dt \left( t + \sqrt{\frac{N_A^2}{2(N-1)}} \right)^3 e^{-\frac{t^2}{48}} \sim O \left( \frac{1}{N^2} \right),
$$

provided that $N_A$ is fixed as $O(1)$. In conclusion, we obtain

$$
\text{Var}(S_A) \leq \langle (S_A - N_A)^2 \rangle \sim O(N^{-2}),
$$

which agrees with [58, 59].

### III. DETAILED CALCULATION OF THE DYNAMICAL PAGE CURVES

As mentioned in the main text, for all the models in this section, we assume the initial state is a period-2 density wave with half filling. Following the same notation in the main text, we further define

$$
X_A(t) = 2 C_A(t) - I_A.
$$

Thus

$$
S_A(t) = N_A - \sum_{n=1}^{\infty} \frac{\text{Tr}[X_A^{2n}(t)]}{2n(2n-1) \ln 2}.
$$

### A. Calculation for the minimal model

We first consider the minimal model. Remember that the minimal model means only nearest neighbor hopping is included. After introducing the Fourier transformed mode $a_k^\dagger = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} e^{-ikj} a_j^\dagger$, we can easily obtain the correlation function in momentum space

$$
\text{Tr}[\rho a_k^\dagger(t) a_k(t)] = \frac{1}{2} \delta_{k,k'} + \frac{1}{2} \delta_{k,k'+\pi} e^{i\theta_k(t)},
$$

where $a_k(t)$ ($a_k^\dagger(t)$) is the annihilation (creation) operator in the Heisenberg picture, $\theta_k(t) = t(E_k - E_{k+\pi})$ and $\rho = |\Psi_0 \rangle \langle \Psi_0|$ corresponds to the initial density matrix. In the following, we may omit the index $t$ if there is no ambiguity.

After the inverse Fourier transformation back to the position space, the covariance matrix for subsystem $A$ reads $[C_A]_{m_1m_2} = \frac{\delta_{m_1m_2}}{2} + \frac{1}{2N} \sum_k e^{i\theta_k} e^{i(k(m_1-m_2))} e^{i\pi m_2}$ and thus

$$
[X_A]_{m_1m_2} = \frac{1}{N} \sum_k e^{i\theta_k} e^{i(k(m_1-m_2))} e^{i\pi m_2}.
$$
Second order in $X_A$

With Eq. (S.16), we can calculate $\overline{\text{Tr} X^2_A}$ as

$$\overline{\text{Tr} X^2_A} = \frac{1}{N^2} \sum_{k_1,k_2,m_1,m_2} e^{i\theta_{k_1}} e^{ik_1(m_1-m_2)} e^{i\pi m_2} e^{i\theta_{k_2}} e^{ik_2(m_2-m_1)} e^{i\pi m_1}$$

$$= \frac{1}{N^2} \sum_{k_1,k_2,m_1,m_2} (\delta_{k_1,k_2+\pi} e^{ik_1(m_1-m_2)} e^{ik_2(m_2-m_1)} e^{-i\pi (m_2-m_1)} e^{i\pi (m_1+m_2)}$$

$$+ \delta_{k_1+k_2,\pi} e^{i2k_1(m_1-m_2)} e^{i\pi (m_1+m_2)} e^{i\pi (m_2-m_1)})$$

$$= \frac{N^2_A}{N} + \frac{1}{N} \sum_{m_1,m_2} e^{i2k_1(m_1-m_2)} = \frac{N^2_A}{N} + \frac{1}{N} \sum_{m_1,m_2} (\delta_{m_1-m_2,0} + \delta_{m_1-m_2,\frac{N}{2}} + \delta_{m_1-m_2,-\frac{N}{2}}).$$

Noting that the first term in the middle step comes from $\theta_{k+\pi} = -\theta_k$, which holds for general Hamiltonians according to the definition. However, the second term is due to the reflection symmetry (which implies $E_k = E_{-k}$) of the minimal model and is thus not universal. Fortunately, this model dependent term will vanish in the thermodynamic limit.

This calculation can be diagrammatically represented as shown in Fig. S1(a), where we draw an arrow from $m_1$ to $m_2$ ($m_2$ to $m_1$) because there is a factor $e^{i k_1 (m_1-m_2)}$ ($e^{i k_2 (m_2-m_1)}$). As will become clear below, such a diagrammatic representation provides a convenient and systematic way for dealing with higher order terms.

Third order in $X_A$

We move on to calculate $\overline{\text{Tr} X^3_A}$ and will see it vanishes in thermodynamic limit. The expression is

$$\overline{\text{Tr} X^3_A} = \frac{1}{N^3} \sum_{k_{1,2,3},m_{1,2,3}} e^{i\theta_{k_1}+i\theta_{k_2}+i\theta_{k_3}} e^{ik_1(m_1-m_2)+ik_2(m_2-m_3)+ik_3(m_3-m_1)} e^{i\pi (m_1+m_2+m_3)}.$$

The non-zero contribution of $e^{i\theta_{k_1}+i\theta_{k_2}+i\theta_{k_3}}$ in the minimal model only comes from two cases:

1. $k_2 = k_1 + \pi$, $k_3 = \pm \frac{\pi}{2}$ and cyclic permutations. The contribution is proportional to

$$\sim \frac{1}{N^3} \sum_{k_1,m_1,m_2,m_3} e^{ik_1(m_1-m_3)} e^{\frac{i\pi}{2} (m_1+m_3)} = \frac{N_A}{N^2} \sum_{m_1,m_3} \delta_{m_1,m_3} e^{i\pi m_1} \sim O \left( \frac{1}{N} \right).$$

2. $k_1, k_2$ satisfy $|\cos k_1 + \cos k_2| \leq 1$ and $k_3 = \arccos(-\cos k_1 - \cos k_2)$. The contribution is proportional to

$$\sim \frac{1}{N^3} \sum_{k_1,k_2} \sum_{m_1} e^{im_1(k_1-k_3+\pi)} \sum_{m_2} e^{im_2(k_2-k_1+\pi)} \sum_{m_3} e^{im_3(k_3-k_2+\pi)}.$$

With Hölder inequality [90]: $\sum_i |a_i||b_i||c_i| \leq \left[ (\sum_i |a_i|^2)(\sum_i |b_i|^2)(\sum_i |c_i|^2) \right]^{\frac{1}{2}}$, we can upper bound the above contribution as

$$\leq \frac{1}{N^3} \left[ \sum_{k_1,k_2} \sum_{m_1} e^{im_1(k_1-k_3+\pi)} \right]^{\frac{3}{2}} \left[ \sum_{k_1,k_2} \sum_{m_2} e^{im_2(k_2-k_1+\pi)} \right]^{\frac{3}{2}} \left[ \sum_{k_3,k_2} \sum_{m_3} e^{im_3(k_3-k_2+\pi)} \right]^{\frac{3}{2}}.$$

Since here, no pair of two $k$ differ by $\pi$ (as it is the case already considered in the first case), the sum of $m_1, m_2, m_3$ only contributes to $O(1)$, so the total contribution will be upper bounded by $O \left( \frac{1}{N^\frac{3}{2}} \right)$.

In conclusion

$$\overline{\text{Tr} X^3_A} \sim O \left( \frac{1}{N} \right)$$

and thus vanishes in the thermodynamic limit.

The above discussion can be generalized to higher orders: as long as the degeneracy point of a Hamiltonian is not dense, we can safely ignore the model dependent contribution and put $e^{i\theta_{k_1}+i\theta_{k_2}} = \delta_{k_1,k_2+\pi}$, which we call a contraction. In the following, we will directly use this contraction rule.[91]
From the above discussion, we can see that the contraction rules here are obviously different from Wick’s theorem.

### B. General Feynman Rules

The method presented in the previous subsection allows us to calculate Eq. (S.14) to arbitrary orders. Here we summarize our Feynman rules for contraction. A general Feynman diagram for calculating \( \text{Tr} X^n_A \) is shown in Fig. S2(a). We will use the integer \( i \) to label the legs associated with momentum \( k_i \).
1. All the legs in Fig. S2(a) must be contracted. Each contraction leads to a delta function of momenta $k$s and has to include even number of legs, where half of the legs should be labeled as even and the other half as odd. This last requirement arises from the phase factor $e^{i\pi m}$ in each leg and is to ensure the diagram not to vanish in the thermodynamic limit (we recall that Fig. S1(d) does not contribute as this requirement is not satisfied). For totally $2n$ legs with $l$ contractions, the $N, N_A$ dependence for this diagram is $\frac{N^{2n-l+1} N_A^l}{N^{2n-l}}$.

2. Each contraction with $2j$ legs should also be assigned a multiple factor $a_{2j}$, accounting for the multiple calculations. $a_2$ and $a_4$ are obtained in previous discussion while higher $a_{2j}$ can be obtained iteratively, as shown below.

3. For each diagram, multiply the term $\frac{N^{2n-l+1}}{N^{2n-l}}$ and factors $a_{2j}$ obtained in Rule 1 and Rule 2 together. Some diagrams also need to multiply by subsystem correction factor $\beta$ (see details below). Sum over all possible diagrams leads to the desired result.

The subsystem correction factor $\beta$ does not appear in $\text{Tr} X^X_A$ and $\text{Tr} X^2_A$, but will appear in calculating $\text{Tr} X^n_A$. For pedagogical purpose, we will now show how to calculate $\text{Tr} X^n_A$. Some diagrams are shown in Fig. S2(b-e). In diagram (b), there are three contractions, each contracts two legs. The contribution for this diagram is $\frac{N_A^3}{N^3} a_2^3$. In diagram (c), there are two contractions, one contracts four legs together and the other contracts two legs. The contribution is $\frac{N_A^4}{N^4} a_4 a_2$. In diagram (d), all the six legs are contracted together, contributing to $\frac{N_A^6}{N^6} a_6$. Attentions should be payed to diagram (e). After contracting the three pairs of legs, we obtain

$$\sum_{k_1, k_2, k_3, k_4, k_5, k_6} e^{ik_1(m_1-m_2+m_4-m_5)} e^{ik_2(m_2-m_3+m_5-m_6)} e^{ik_3(m_3-m_4+m_6-m_1)}$$

$$= N^3 \sum_{m_1, m_2, m_3, m_4, m_5, m_6} \delta_{m_1+m_4=m_2+m_5=m_3+m_6} \text{mod } N.$$ \hspace{1cm} (S.22)

Naively, one may conjecture the result of the last sum to be $N_A^4$. However, this is only true when $f = N_N^N = 1$. If $f \leq \frac{1}{2}$, the sum will be much smaller. After carefully counting the pairs satisfying the delta function, we obtain for $f \leq \frac{1}{2}$

$$\sum_{k_1, k_2, k_3, k_4, k_5, k_6} e^{ik_1(m_1-m_2+m_4-m_5)} e^{ik_2(m_2-m_3+m_5-m_6)} e^{ik_3(m_3-m_4+m_6-m_1)} = \frac{1}{2} N^3 (N_A^4 + N_A^3) \to \beta_1 N^3 = N_A^4,$$ \hspace{1cm} (S.23)

where $\beta_1$ is defined as the subsystem correction factor in thermodynamic limit: $\beta_1 = \begin{cases} \frac{7}{2} & f \leq \frac{1}{2} \\ 1 & f = 1 \end{cases}$. Summing over all possible diagrams, the total contribution is

$$\text{Tr} X_A^X = (5 + \beta_1) \frac{N_A^4}{N^3} - (6 + 3\beta_2) \frac{N_A^5}{N^4} + \frac{N_A^6}{N^5} a_0,$$ \hspace{1cm} (S.24)

where $\beta_2$ is another subsystem correction factor $\beta_2 = \begin{cases} \frac{7}{2} & f \leq \frac{1}{2} \\ 1 & f = 1 \end{cases}$. Here $a_0$ can be determined by considering the case when $N_A = N$ (i.e. $f = 1$). In this case, $\text{Tr} X_A^5 = N$, resulting in $a_0 = 4$. Therefore, $\text{Tr} X_A^5 = \frac{11}{2} \frac{N_A^4}{N^3} - 8 \frac{N_A^5}{N^4} + 4 \frac{N_A^6}{N^5}$ if $f \leq \frac{1}{2}$.

$$\text{Tr} X_A^5 = \frac{11}{2} \frac{N_A^4}{N^3} - 8 \frac{N_A^5}{N^4} + 4 \frac{N_A^6}{N^5} \text{ if } f \leq \frac{1}{2}.$$ \hspace{1cm} (S.25)
Following the above procedure for calculating the Feynman diagrams, we arrive at, up to the order $\mathcal{O}(f^5)$,
\[
\frac{S_A}{N} = f - \frac{1}{12} \left( \frac{1}{2} f^2 + \frac{1}{6} f^3 + \frac{1}{10} f^4 + 0.06 f^5 \right) + \mathcal{O}(f^6). \tag{S.26}
\]

C. Proof of Theorem 2

In this subsection, we go beyond the minimal model and consider the general Hamiltonians satisfying the condition in Theorem 2 in the main text. We will find the Feynman rule as well as the subsystem entropy is indeed the same as the previous subsection.

Since the Hamiltonian is period-2, we can use a modified Fourier transformation to block diagonalize it:
\[
A_k^1 = \sqrt{\frac{2}{N}} \sum_{j=1}^{\frac{N}{2}} e^{-i(kj-1)} a_{2j-1}^1, \quad B_k^1 = \sqrt{\frac{2}{N}} \sum_{j=1}^{\frac{N}{2}} e^{-i(kj)j} a_{2j}^1, \quad k \in \left\{ \frac{2n\pi}{N} \right\}_{n=0}^{\frac{N}{2}-1}. \tag{S.27}
\]

Here $A_k^1, B_k^1$ are related to the conserved (eigen) modes $P_k^1, Q_k^1$ via a $2 \times 2$ unitary transformation $U_k$ as $A_k^1 = U_{11}^k P_k^1 + U_{12}^k Q_k^1$ and $B_k^1 = U_{21}^k P_k^1 + U_{22}^k Q_k^1$. Substituting into Eq. (S.27) leads to
\[
a_{2m}^1 = \sqrt{\frac{2}{N}} \sum_{k=0}^{\pi} e^{ik2m} (U_{21}^k P_k^1 + U_{22}^k Q_k^1), \quad a_{2m+1}^1 = \sqrt{\frac{2}{N}} \sum_{k=0}^{\pi} e^{ik(2m+1)} (U_{11}^k P_k^1 + U_{12}^k Q_k^1). \tag{S.28}
\]

If we define $Q_{k+\pi} = P_k$ and
\[
Z_k^m = \begin{cases} \sqrt{2} U_{22}^k, & \text{if } m \text{ is even and } k < \pi; \\ \sqrt{2} U_{12}^k, & \text{if } m \text{ is odd and } k < \pi; \\ -\sqrt{2} U_{11}^k, & \text{if } m \text{ is odd and } k \geq \pi, \end{cases} \tag{S.29}
\]

the above inverse Fourier transformation (S.28) can be rewritten as
\[
a_m^1 = \frac{1}{\sqrt{N}} \sum_{k=0}^{2\pi} Z_k^m Q_k^1 e^{ikm}.
\]

In the following, we will simplify $\sum_{k=0}^{2\pi}$ as $\sum_k$ and $k$ should be understood as module $2\pi$. By assumption, all the conserved quantity $\text{Tr}(\rho Q_k^1 Q_k^1) = \frac{1}{2}$ for $k \in [0, 2\pi)$, thus
\[
[C_A]_{ml} = \frac{1}{2N} \sum_k Z_k^m Z_k^{l*} e^{ik(m-l)} + \frac{1}{2N} \sum_k e^{i\theta_k} Z_k^m Z_{k+\pi}^{l*} e^{i(k+\pi)(m-l)} e^{i\pi l} = \frac{1}{2} \delta_{m+l} + \frac{1}{2N} \sum_k e^{i\theta_k} Z_k^m Z_{k+\pi}^{l*} e^{i(k+\pi)(m-l)} e^{i\pi l}
\]

where $\theta_{k+\pi} = -\theta_k$ by definition. In the last equality, we have used the unitarity of $U_k^k$. Namely, if $m - l$ is odd:
\[
\frac{1}{2N} \sum_k Z_k^m Z_k^{l*} e^{i(k(m-l))} = \frac{1}{4N} \sum_k [Z_k^m Z_k^{l*} e^{i(k(m-l))} + Z_{k+\pi}^m Z_{k+\pi}^{l*} e^{i(k+\pi)(m-l)}]
\]
\[
= \frac{1}{4N} \sum_k (Z_k^m Z_k^{l*} - Z_{k+\pi}^m Z_{k+\pi}^{l*}) e^{i(k(m-l))} = 0.
\]

A similar calculation can be carried out for the case in which $m - l$ is even. Therefore, we have
\[
[X_A]_{ml} = \frac{1}{N} \sum_k e^{i\theta_k} Z_k^m Z_{k+\pi}^{l*} e^{i(k(m-l))} e^{i\pi l}.
\]

This expression is very similar to Eq. (S.16) except for the extra factors $Z_k$. Nonetheless, we will show those extra $Z_k$’s do not contribute in the thermodynamic limit. As a result, the same Feynman rules and dynamical Page curve follows.
When evaluating a Feynman diagram in the thermodynamic limit with $N_A, N$ both going to infinity, we can first sum over the position indices. Introducing $s_j = k_j - k_{j-1}$, $k_0 = k_{2n}$ and $m_{2n+1} = m_1$, we obtain

$$
\sum_{m_{1,2\ldots 2n}} \prod_{j=1}^{2n} e^{i k_j (m_j - m_{j+1})} e^{i m_j Z_{k_j}^m Z_{k_{j-1}+\pi}^m} \\
= \sum_{m_{1,2\ldots 2n}} \prod_{j=1}^{2n} e^{i m_j s_j} e^{i m_j Z_{k_j}^m Z_{k_{j-1}+\pi}^m} \\
= 2^n \prod_{j=1}^{2n} \sum_{m_j: \text{even}} e^{i m_j (s_j + \pi)} Z_{k_j}^{0*} Z_{k_{j-1}+\pi}^{0*} + \sum_{m_j: \text{odd}} e^{i m_j (s_j + \pi)} Z_{k_j}^{1*} Z_{k_{j-1}+\pi}^{1*} \\
= 2^n \prod_{j=1}^{2n} \left[ 1 - e^{i N_A (s_j + \pi)} \right] e^{i \frac{N_A (s_j + \pi)}{2 \sin (s_j + \pi)}} \prod_{j=1}^{2n} \left[ \frac{\sin N_A (s_j + \pi)}{\sin (s_j + \pi)} \right] \prod_{j=1}^{2n} (Z_{k_j}^{0*} Z_{k_{j-1}+\pi}^{0*} + e^{i (s_j + \pi)} Z_{k_j}^{1*} Z_{k_{j-1}+\pi}^{1*}) \\
= 2^n \prod_{j=1}^{2n} \left[ \frac{\sin N_A (s_j + \pi)}{\sin (s_j + \pi)} \right] \prod_{j=1}^{2n} (Z_{k_j}^{0*} Z_{k_{j-1}+\pi}^{0*} + e^{i (s_j + \pi)} Z_{k_j}^{1*} Z_{k_{j-1}+\pi}^{1*})
$$

In the above calculation, we have assumed $N_A$ to be even for simplicity. We also emphasize that $s_1 \cdots s_{2n}$ is not independent since $\sum_{i=1}^{2n} s_i = 0$. The factor $\frac{\sin N_A (s_j + \pi)}{\sin (s_j + \pi)}$ will be dominated by the contribution from $s_j = 0$ or $s_j = \pi$ if $Z_k$ is smooth enough. This is due to the convergence of Fourier series, see [92]. In [93] the difference between the discrete sum over momenta and the integration is also upper bounded. However, if $s_j \approx 0$, it will lead to

$$
Z_{k_j}^{0*} Z_{k_{j-1}+\pi}^{0*} + e^{i (s_j + \pi)} Z_{k_j}^{1*} Z_{k_{j-1}+\pi}^{1*} \approx Z_{k_j}^{0*} Z_{k_{j-1}+\pi}^{0*} + Z_{k_j}^{1*} Z_{k_{j-1}+\pi}^{1*} = 0
$$

due to the unitarity of $U$. In the end, we obtain

$$
\sum_{m_{1,2\ldots 2n}} \prod_{j=1}^{2n} e^{i k_j (m_j - m_{j+1})} e^{i m_j Z_{k_j}^m Z_{k_{j-1}+\pi}^m} \approx (Z_{k_j}^{0*} + Z_{k_j}^{1*}) \prod_{j=1}^{2n} \left[ \frac{\sin N_A (s_j + \pi)}{2 \sin (s_j + \pi)} \right] s_j \neq 0
$$

(S.31)

Now we can see in the final expression Eq. (S.31) that the model-dependent factor $Z$ disappears. Therefore, those Hamiltonians satisfying the conditions in Theorem 2 in the main text will have the same Feynman rules and dynamical Page curve as the minimal model.

In Fig. S3, we plotted the dynamical Page curve for the Hamiltonian

$$
H = \sum_i a_i^a a_{i+1} + 0.3 \sum_{i: \text{even}} a_i^a a_{i+3} - 0.3 \sum_{i: \text{odd}} a_i^a a_{i+3} + H.C.
$$

(S.32)

This dynamical Page curve is nearly the same as the one of minimal model in the main text.

### D. Generalization to the atypical Page curves

In this subsection, we further consider the case beyond the condition in Theorem 2 in the main text, namely $\text{Tr}(\rho Q_k^A Q_k) \neq \frac{1}{2}$ (we recall that $Q_{k+\pi} = P_k$). We denote $n_k = \text{Tr}(\rho Q_k^A Q_k)$ and $\eta_k = \sqrt{n_k (1 - n_k)}$. Following the half filling condition, we have

$$
n_{k+\pi} = 1 - n_k, \quad \eta_{k+\pi} = \eta_k
$$
Figure S3. The dynamical Page curve for Hamiltonian (S.32) (blue curve) and its comparison with the one for the RFG ensemble (red curve) and the theoretical result (green curve). Here \( N = 200 \). The theoretical result is truncated up to order \( \mathcal{O}(f^5) \), the same as in the main text.

and still \( a_m^l = \frac{1}{\sqrt{N}} \sum_{k=0}^{2\pi} Z_k^m Q_k^l e^{ikm} \) with \( Z_k^m \) defined in Eq. (S.29). The covariance matrix can be calculated as

\[
[C_A]_{ml} = \frac{1}{N} \sum_{k_{1,2}} Z_{k_1}^m Z_{k_2}^{l*} e^{ik_1m} e^{-ik_2l} \text{Tr}(\rho Q_{k_1}^l Q_{k_2})
\]

\[
= \frac{1}{N} \sum_k Z_k^m Z_k^{l*} e^{ik(m-l)} n_k + \frac{1}{N} \sum_k e^{i\theta_k} Z_k^m Z_{k+\pi}^{l*} e^{ik(m-l)} e^{i\pi l} \eta_k,
\]

Since \( n_k \neq \frac{1}{2} \), in general there is no simple expression for \( X_A \).

Using the same techniques as in the previous subsection, we can still establish the Feynman rules for this case. However, we have to distinguish two kinds of legs, one like \( Z_k^m Z_{k+\pi}^{l*} e^{ik(m_j-m_j+1)} n_{k_j} \) and the other like

\( e^{i\theta_{k_j}} Z_k^{m_j} Z_{k_j+\pi}^{m_j+1*} e^{ik_j(m_j-m_{j+1})} e^{i\pi m_{j+1}} \eta_{k_j} \). There is no dynamical phase \( e^{i\theta_k} \) in the former, namely no delta functions associated with contraction. Also, there is no extra \( e^{i\pi m_{j+1}} \) phase term in the former. Due to the difference between these two kinds of legs, the rule is more complicated than the previous case.

As an example, we can calculate the first three non-trivial terms to obtain:

\[
\text{Tr} C_A^2 = \frac{N_A}{N} \sum_k n_k^2 + \frac{N_A^2}{N^2} \sum_k \eta_k^2,
\]

\[
\text{Tr} C_A^3 = \frac{N_A}{N} \sum_k n_k^3 + \frac{3N_A^2}{N^2} \sum_k n_k \eta_k^2,
\]

\[
\text{Tr} C_A^4 = \frac{N_A}{N} \sum_k n_k^4 + \frac{3N_A^2}{N^2} \sum_k n_k^2 \eta_k^2 + \frac{4N_A^3}{N^3} \sum_k \eta_k^4 + \frac{2N_A^2}{N^2} \sum_k \eta_k^4 - \frac{N_A^4}{N^4} \sum_k \eta_k^4.
\]

Therefore, Up to \( \text{Tr} X_A^4 \), \( [94] \)

\[
S_A \simeq N_A (\ln 2 + \frac{3}{4}) - \frac{N_A}{N} \sum_k (4n_k^2 - \frac{8}{3} n_k^3 + \frac{4}{3} n_k^4) - \frac{N_A^2}{N^2} \sum_k (4n_k^2 - 8n_k \eta_k^2 + \frac{16}{3} n_k^2 \eta_k^2 + \frac{8}{3} \eta_k^4) - \frac{N_A^3}{N^3} \sum_k n_k \eta_k^4 + \frac{N_A^4}{N^4} \sum_k \eta_k^4.
\]
IV. CALCULATION OF ENTANGLEMENT ENTROPY IN THE QUASI-PARTICLE PICTURE

In the quasi-particle picture, a nonequilibrium initial state is a source for generating quasi-particles with opposite momenta, which travel ballistically through the system. Here the main assumption is those quasi-particle pairs generated at different locations and times are incoherent. Therefore, the entanglement entropy of subsystem $A$ is proportional to the number of pairs shared between $A$ and its complement. Without loss of generality, we can assume the subsystem $A$ is located in $[0, N_A)$, $N_A \leq \frac{N}{2}$. For a certain type of pairs with velocity $\pm v(k)$ ($v(k) > 0$), if the right-end of the pair is at position $0 \leq x < N_A$, i.e. within subsystem $A$, only when $x$ satisfies

$$N_A - N \leq x - 2v(k)t < 0$$

can this pair contribute to the entanglement entropy of $A$. Here the periodic boundary condition is taken into account and $2v(k)t$ should be understood as modulo $N$. The solutions of this inequality is a continuous range $x \in [x_{\text{min}}, x_{\text{max}})$, where $x_{\text{min}} = \max\{0, 2v(k)t + N_A - N\}$ and $x_{\text{max}} = \min\{N_A, 2v(k)t\}$. Accordingly, we obtain

$$\Delta x = x_{\text{max}} - x_{\text{min}} = \begin{cases} 2v(k)t, & 2v(k)t \leq N_A; \\ N_A, & N_A < 2v(k)t < N - N_A; \\ N - 2v(k)t, & 2v(k)t \geq N - N_A. \end{cases}$$

A similar argument holds if the left-end is in subsystem $A$. We assume that after a sufficiently long time, the quasi-particle pairs will distribute uniformly among the system. Hence, the contribution of quasi-particle pairs with momentum $k$ to the entanglement entropy upon the long-time average is given by

$$\frac{S_A(k)}{N^2} \int_0^N d(2v(k)t)\Delta x = S_A(k) \left[ \frac{N_A}{N} - \left( \frac{N_A}{N} \right)^2 \right],$$

where the coefficient $S_A(k)$ is to be determined. Summing over all types of pairs, we obtain

$$S_A^{\text{qp}} = \left( \frac{N_A}{N} - \frac{N_A^2}{N^2} \right) \sum_k S_A(k).$$

If $N_A \to 0$, the limit $S_A^{\text{qp}} \to N_A \sum_k \frac{H(n_k)}{N}$ should hold [95]. Therefore, $S_A(k) = H(n_k)$. If $n_k = \frac{1}{2}$ for all $k$’s, the entanglement entropy for subsystem $A$ is

$$S_A^{\text{qp}} = N_A - \frac{N_A^2}{N},$$

which deviates considerably from the dynamical Page curve discussed in the main text.