

# Matchgate Circuits Deeply Thermalize

Mircea Bejan,<sup>1</sup> Benjamin Béri<sup>1,2</sup>, and Max McGinley<sup>1</sup>

<sup>1</sup>*T.C.M. Group, Cavendish Laboratory, University of Cambridge, J.J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

<sup>2</sup>*DAMTP, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom*



(Received 13 December 2024; revised 16 May 2025; accepted 4 June 2025; published 7 July 2025)

We study the ensemble of states generated by performing projective measurements on the output of a random matchgate (or free-fermionic) quantum circuit. We rigorously show that this “projected ensemble” exhibits deep thermalization: for large system sizes, it converges toward a universal ensemble that is uniform over the manifold of Gaussian fermionic states. As well as proving momentwise convergence of these ensembles, we demonstrate that the full distribution of any physical observable in the projected ensemble is close to its universal form in Wasserstein-1 distance, which we argue is an appropriate and efficiently computable measure of convergence when studying deep thermalization. Using this metric, we also numerically find that local matchgate circuits deeply thermalize after a timescale  $t \sim L^2$  set by the diffusive spreading of quantum information. Our work opens up new avenues to experimentally accessible protocols to probe the emergence of quantum statistical mechanics and benchmark quantum simulators.

DOI: [10.1103/v8kp-39ry](https://doi.org/10.1103/v8kp-39ry)

Our theoretical understanding of how thermalization and statistical mechanics emerge from underlying microscopic many-body dynamics has evolved greatly over the last century, particularly in the context of isolated quantum systems [1–4]. A standard approach to probing thermalization is to watch how small subsystems  $A$  locally equilibrate under the full dynamics of the system. At late times, statistical mechanics prescribes that the state of  $A$  should have maximal entropy (subject to any physical constraints). Microscopically, this is made possible by the generation of entanglement between  $A$  and its surroundings  $B$  [5,6], the latter acting as a heat bath.

Thanks to advances in quantum simulators [7], it is now possible to experimentally study quantum dynamics beyond the evolution of local observables. This has motivated notions of generalized forms of thermalization [8–22], which pertain to full conditional distributions on  $A$  given measurement outcomes on  $B$ , rather than ignoring  $B$  entirely. Specifically, if one looks at the statistical properties of the ensemble of states on  $A$  after measurements on  $B$ , one gets the so-called *projected ensemble* (PE) [8], see also Refs. [23–28]. This contains more information than traditional measures of thermalization. Results have pointed to the emergence, at late times, of a new phenomenon dubbed *deep thermalization* (DT) [10,11], where

the PE approaches a universal ensemble of states that are maximally random, up to physical constraints.

Probing deep thermalization experimentally comes with a postselection problem, akin to measurement-induced phase transitions in entanglement [29–35]. Since the PE contains  $\sim 2^L$  states (for an  $L$ -qubit system), brute-force tests of DT have exponential sample complexity and have thus been restricted to dozens of qubits [8,9]. At the same time, the only rigorous theoretical results on DT in finite-dimensional Hilbert spaces are for Haar-random qubit states [8,13] and dual-unitary circuits [10,12], for which no known solution to the postselection problem exists. This raises the question of whether one can experimentally realize systems that exhibit deep thermalization, but which do not suffer from a postselection barrier.

In this work, we study DT in free-fermion dynamics and their qubit-based equivalents, matchgate circuits [36–38], cf. Fig. 1. These circuits, which form the basis of fermion linear optics (FLO) [38,39], can capture the time evolution of any quadratic fermionic Hamiltonian. Because matchgates are classically efficiently simulable, one can construct hybrid quantum-classical algorithms for experimentally probing DT in matchgate-dominated systems, which overcome the postselection barrier [40–42]. Accordingly, the physics we study here can be efficiently verified and may serve as a robust benchmark for quantum simulators and computers.

While these circuits are not universal for quantum computation [43] and hence cannot exhibit exactly the same behavior as fully universal systems, here we prove that free fermions do deeply thermalize: in particular, at late times, the projected ensemble tends toward a universal distribution, which we refer to as the Gaussian Haar ensemble (GHE).

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

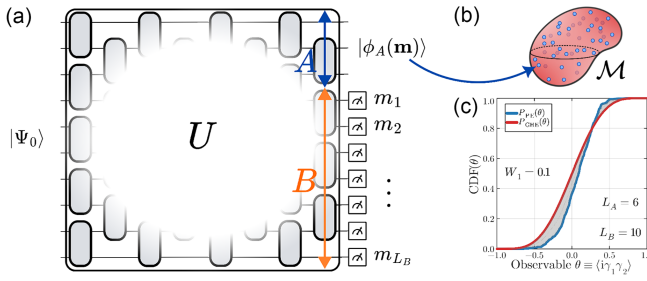


FIG. 1. (a) Starting from an initial state  $|\Psi_0\rangle$  on  $L$  fermionic modes, a Gaussian circuit  $U$  is applied, after which Fock basis measurements are performed on subsystem  $B$ , with outcome  $\mathbf{m}$ . This yields pure states  $|\phi_A(\mathbf{m})\rangle$ , which constitute the PE. (b) The PE is a discrete ( $2^{L_B}$  elements shown as blue circles) probability distribution on the manifold  $\mathcal{M}$  of pure Gaussian states. We prove that the PE covers  $\mathcal{M}$  uniformly for  $L_B \rightarrow \infty$ . (c) Cumulative distribution function for the projected and Haar ensembles vs the observable  $\theta = \langle i\gamma_1\gamma_2 \rangle$  for global  $U$ . The Wasserstein-1 distance  $W_1$  (gray shaded region's area) quantifies the distance between these distributions.

This rigorously establishes that the phenomenology seen in the numerical experiments of Ref. [16] holds true in general and complements a related observation made for bosonic Gaussian systems in Ref. [20].

Our main rigorous results—Theorems 1–3—show that, for random free-fermionic circuits, the PE converges to the GHE at the level of *full probability distributions* of physical observables, which is much stronger than previous approaches where low-order moments are compared. To make this comparison concrete, we propose to use a particular measure of closeness between probability distributions known as the Wasserstein-1 distance, which we argue is particularly relevant to this setting. We supplement these results with numerics, which also show that the corresponding probability distributions remain close in the same precise sense also for local random matchgate circuits at late times. We argue that the universal behavior emerges after a timescale  $t \sim L^2$  set by the diffusive spreading of quantum information in the system.

**Physical setup and the projected ensemble**—We consider Gaussian dynamics [38,39] (see also Supplemental Material [44]) on  $L$  fermionic modes, which are spanned by  $2L$  Majorana operators  $\{\gamma_j\}$  satisfying  $\{\gamma_j, \gamma_k\} = 2\delta_{jk}$  [43,45]. Starting from the vacuum  $|0^{\otimes L}\rangle$  without loss of generality, we apply a sequence of FLO gates, without particle number conservation [38,39]. These operations can be specified by an orthogonal matrix  $O \in \text{SO}(2L)$ , to which we associate a Gaussian unitary operator  $U[O]$  in Fock space [45] (or, alternatively, a matchgate circuit on  $L$  qubits [38]). Often, we will consider a  $(1+1)D$  geometry with gates applied in a brickwork architecture as shown in Fig. 1(a). Being Gaussian, the resulting state  $|\Psi\rangle = U[O]|0^{\otimes L}\rangle$  is fully determined by its covariance matrix  $\Gamma_{jk} := (i/2)\langle\Psi|[\gamma_j, \gamma_k]|\Psi\rangle$  [43].

We then projectively measure each mode within a subsystem  $B \subset [L] := \{1, \dots, L\}$  of size  $L_B = |B|$  in the Fock basis. The outcome  $\mathbf{m} \in \{0, 1\}^{L_B}$  occurs with probability  $p_{\mathbf{m}} = \langle\Psi|(\mathbb{1}_A \otimes |\mathbf{m}\rangle\langle\mathbf{m}|_B)|\Psi\rangle$ , conditioned on which the postmeasurement state of the unmeasured modes  $A = [L] \setminus B$  becomes  $|\phi_A(\mathbf{m})\rangle = (\mathbb{1}_A \otimes \langle\mathbf{m}|_B)|\Psi\rangle / \sqrt{p_{\mathbf{m}}}$ , which is also a Gaussian state [38,39,43], and hence is specified by a covariance matrix  $\Gamma_{\mathbf{m}}^A$ . The distribution of postmeasurement states is the PE [see Fig. 1(b)]

$$\mathcal{E}_{\text{PE}} := \{p_{\mathbf{m}}, |\phi_A(\mathbf{m})\rangle\}. \quad (1)$$

Our aim is to characterize universal properties of the PE for  $L_B \rightarrow \infty$ . Recent work showed that for generic dynamics, the PE tends toward the Haar ensemble, i.e., the uniform distribution over all states on  $A$  [8,10–15,17]. More generally, if the dynamics is constrained, e.g., for energy- or charge-conserving dynamics, or in Gaussian bosonic systems, it has been suggested that the appropriate ensemble is the one that is maximally entropic subject to these constraints [19–21].

For free fermions, the authors of Ref. [16] presented numerical evidence, in terms of certain statistical moments, that under Hamiltonian evolution a generalized Gibbs ensemble of a similar kind is reached at late times. However, analytical results on finite-order statistical moments and, especially, on the full distribution of post-measurement states have been lacking. Here, we rigorously prove that the PE does indeed converge (in a sense that we will make precise) to the Gaussian Haar ensemble, which is uniform over all Gaussian states on  $A$

$$\mathcal{E}_{\text{GHE}} := \text{Unif}(\{|\phi^A\rangle \in \mathcal{H}^A : |\phi^A\rangle \text{ is Gaussian}\}). \quad (2)$$

Since we can identify Gaussian states, through their covariance matrix, with points on the symmetric space  $[\text{SO}(2L_A)/\text{U}(L_A)]$ , or DIII in Cartan's classification [46] (see [44]), “Unif” here refers to the corresponding circular ensemble [47].

**Deep thermalization via moments**—Because the post-measurement states are Gaussian, a natural way to characterize the PE (1) is via the covariance matrices of the conditional states  $\Gamma_{\mathbf{m}}^A$ . Viewing the matrix elements  $[\Gamma_{\mathbf{m}}^A]_{ij}$  as a collection of  $(2L_A)^2$  random variables, we define the  $k$ th moments of this multivariate distribution

$$\mathcal{C}_{\text{PE}, \mathbf{ij}}^{(k)} = \sum_{\mathbf{m}} p_{\mathbf{m}} [\Gamma_{\mathbf{m}}^A]_{i_1 j_1} \cdots [\Gamma_{\mathbf{m}}^A]_{i_k j_k}, \quad (3)$$

where  $\mathbf{i} = (i_1, \dots, i_k)$  is a multi-index where each individual index  $i_a \in [2L_A]$  specifies a Majorana operator  $\gamma_{i_a}$  (similar for  $\mathbf{j}$ ). To probe deep thermalization, we compare  $\mathcal{C}_{\text{PE}, \mathbf{ij}}^{(k)}$  with the corresponding GHE moments  $\mathcal{C}_{\text{GHE}, \mathbf{ij}}^{(k)}$ .

Our first rigorous result (proved in [44]) is that free fermions evolving under globally random Gaussian circuits deeply thermalize at the level of these moments.

*Theorem 1*—If  $|\Psi\rangle$  is prepared by a circuit  $U[O]$  with  $O$  drawn from the Haar measure over  $\text{SO}(2L)$ , then for any choice of  $\mathbf{i}, \mathbf{j}$ , the probability that the moment (3) deviates by more than  $\epsilon$  from the corresponding GHE moment is

$$\mathbb{P}_O\left(|\mathcal{C}_{\text{PE},\mathbf{i}\mathbf{j}}^{(k)} - \mathcal{C}_{\text{GHE},\mathbf{i}\mathbf{j}}^{(k)}| > \epsilon\right) < \exp\left(-\frac{\kappa L_B \epsilon^2}{k^2}\right) \quad (4)$$

for some constant  $\kappa \geq 1/64$ .

This result, observed numerically for  $k \leq 4$  in Ref. [16], establishes that the moments of the covariance matrices themselves are close to those of the GHE with high probability in the large  $L_B$  regime.

We may also consider a more operational notion of deep thermalization, which is more closely related to the non-Gaussian case. Here, we compare the moments of the full many-body states

$$\mathcal{E}_{\text{PE}}^{(k)} := \sum_{\mathbf{m}} p_{\mathbf{m}}(|\phi_A(\mathbf{m})\rangle\langle\phi_A(\mathbf{m})|)^{\otimes k} \quad (5)$$

with those of the GHE,  $\mathcal{E}_{\text{GHE}}^{(k)}$ . Following the nomenclature for fully random quantum states, we say that the projected ensemble forms an  $\epsilon$ -approximate Gaussian  $k$ -design if

$$\Delta^{(k)} := \left\| \mathcal{E}_{\text{PE}}^{(k)} - \mathcal{E}_{\text{GHE}}^{(k)} \right\|_1 \leq \epsilon, \quad (6)$$

where  $\|X\|_1 := \text{Tr}[\sqrt{X^\dagger X}]$  is the trace norm. Concretely, this condition implies that any strategy to distinguish these ensembles using  $k$  copies of the state—including those using non-Gaussian operations—has success probability close to that of a random guess  $p_{\text{succ}} \leq \frac{1}{2} + (\epsilon/4)$  [72]. We can prove the following. (See [44] for a proof.)

*Theorem 2*—There exist constants  $\alpha, \beta$  such that if  $|\Psi\rangle$  is prepared by a circuit  $U[O]$  with  $O$  drawn from the Haar measure over  $\text{SO}(2L)$ , then the PE forms an  $\epsilon$ -approximate Gaussian  $k$ -design with probability at least  $1 - \delta$ , provided  $L_B \geq \alpha \log(1/\delta)/\epsilon^2$  and

$$kL_A \leq \beta \log(L_B). \quad (7)$$

This should be compared to the case with  $|\Psi\rangle$  fully Haar random, where the PE forms an  $\epsilon$ -approximate  $k$  design for  $kL_A \leq \beta' L_B$ , provided  $L_B \geq \alpha' [\log(1/\epsilon) + \log \log(1/\delta)]$  for some constants  $\alpha', \beta'$  [8].

An important concept that we invoke in our proofs of Theorems 1 and 2 is the concentration of measure phenomenon for the Haar measure over  $\text{SO}(2L) \ni O$  [48], which states that for certain well-behaved functions  $F(O)$ , large deviations from the mean are very rare. In brief, we show that the quantities (3) and (5) satisfy the necessary

requirements to invoke these results, and that their mean is equal to their GHE equivalent.

*Deep thermalization via Wasserstein distance*—The  $k$  moments of the covariance matrices (3) and the many-body states (5) allow us to infer certain statistics of a number of physical quantities in the projected ensemble. For instance, the average and variance of the purity of a subsystem of  $A$  can be computed from  $\mathcal{E}_{\text{PE}}^{(4)}$ . However, it would be desirable to characterize the full distribution of observables in the PE more comprehensively, without having to look at a spectrum of distances  $\Delta^{(k)}$  for all  $k$ . For this reason, we look to relate the ensembles  $\mathcal{E}_{\text{PE}}$  (1) and  $\mathcal{E}_{\text{GHE}}$  (2) using a *statistical distance*.

Statistical distances allow one to compare two probability distributions [73]. These include the total variation distance and the Kullback-Leibler divergence [74], which are natural choices in many probability-theoretic settings. However, these particular examples always evaluate to their maximal value whenever we compare a discrete with a continuous distribution, even if the former is a “good” discretization of the latter; this is discussed in [44].

Here, instead, we propose to use the Wasserstein-1 distance [49,75] to quantify deep thermalization. Formally, for two distributions  $\mu, \nu$  over a manifold  $\mathcal{M} \ni x$ , the Wasserstein-1 distance is

$$W_1(\mu, \nu) = \sup_{f(x): 1\text{-Lipschitz}} \mathbb{E}_{x \sim \mu}[f(x)] - \mathbb{E}_{x \sim \nu}[f(x)], \quad (8)$$

where  $\mathbb{E}_{x \sim \mu}[f(x)]$  denotes the expectation of  $f$  over the distribution  $\mu$ , and similar for  $\nu$ . Here, we are using the “test function”  $f(x)$  to probe discrepancies between the distributions  $\mu$  and  $\nu$ . In particular, we take the maximum discrepancy over all “sufficiently smooth” functions, specifically those that are  $\eta$ -Lipschitz with  $\eta \leq 1$ ,

$$|f(x_1) - f(x_2)| \leq \eta d(x_1, x_2) \quad \forall x_1, x_2 \in \mathcal{M} \quad (9)$$

where  $d(\cdot, \cdot)$  is a chosen metric on  $\mathcal{M}$ . In our setting, we are comparing distributions  $\mu = \mathcal{E}_{\text{PE}}$  (1) and  $\nu = \mathcal{E}_{\text{GHE}}$  (2), with  $\mathcal{M}$  being the manifold of pure Gaussian states. As we shall see, 1-Lipschitz functions over this manifold encompass physically relevant properties [44].

Computing the Wasserstein-1 distance of  $\mathcal{E}_{\text{PE}}$  and  $\mathcal{E}_{\text{GHE}}$  is a formidable task since we need to optimize over functions  $f$ . However, if we instead look at the distribution of a particular scalar quantity  $\theta = \theta(\Gamma^A)$ , we can make some progress. In this case, we need to compare two univariate distributions, which can be represented by their density functions  $p(\theta)$  and  $q(\theta)$  over some domain  $X \subseteq \mathbb{R}$ . Equation (8) then simplifies significantly to [76]

$$W_1(p(\theta), q(\theta)) = \int_{\theta \in X} d\theta |P(\theta) - Q(\theta)|, \quad (10)$$

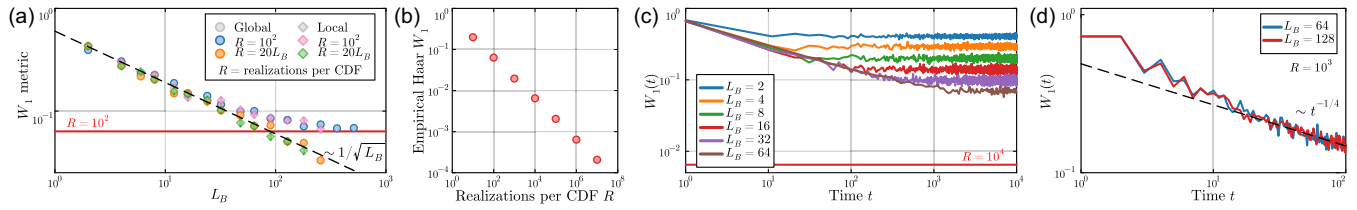


FIG. 2. Deep thermalization of fermionic Gaussian dynamics: Wasserstein-1 distance  $W_1$  between distributions of the correlator  $\theta(\Gamma_{\mathbf{m}}^A) = \langle i\gamma_1\gamma_2 \rangle_{\mathbf{m}}$  for  $L_A = 2$ , averaged over  $N = 10^2$  circuit instances. (a)  $W_1$  decays with the number of measured qubits  $L_B$  as  $W_1 \propto 1/\sqrt{L_B}$  (dashed line), for both global (circles) and late-time local ( $t = 10^4$ , rhombi) dynamics, down to a threshold value (horizontal red line). The error bars, where visible, show the standard error of the mean (SE). (b) Estimate of the finite- $R$  plateau from comparing an empirical CDF using  $R$  GHE realizations and the exact GHE CDF, showing  $W_1 \propto 1/\sqrt{R}$ . (c)  $W_1(t)$  vs time  $t$  for local dynamics. The horizontal red line shows the empirical threshold. Relative SEs are all below 9% (not shown). An empirical PE is constructed every  $10(10^2)$  time steps for  $L_B = 2, \dots, 32(64)$  and  $R = 10^4$ . (d) Closer view of the decay  $W_1(t) \propto t^{-1/4}$  (dashed line) for  $R = 10^3$  and an empirical PE built at each time step. Relative SEs are all below 9%.

with  $P(\theta)$  and  $Q(\theta)$  being the cumulative distribution functions (CDFs) of  $p(\theta)$  and  $q(\theta)$ , respectively, i.e., one takes the unsigned area between the CDFs [see Fig. 1(c)]. This will allow us to numerically compute  $W_1$  in certain instances later on.

Our final rigorous result uses the Wasserstein-1 distance to characterize the *full distribution* of any (sufficiently well behaved) physical quantity  $\theta$ , beyond low-order moments.

**Theorem 3**—Let  $\theta(\Gamma^A)$  be a physical quantity that is bounded in an interval  $[a, b]$ , differentiable, and  $c$ -Lipschitz, i.e.,  $\|\nabla_{\Gamma}\theta\|_2 \leq c$ . If  $|\Psi\rangle$  is prepared by a circuit  $U[O]$  with  $O$  drawn from the Haar measure over  $\text{SO}(2L)$ , then the distributions of  $\theta$  in the projected and Gaussian Haar ensembles  $p_{\text{PE}}(\theta)$  and  $p_{\text{GHE}}(\theta)$  converge in Wasserstein-1 distance for large  $L_B$

$$W_1[p_{\text{PE}}(\theta), p_{\text{GHE}}(\theta)] \leq \kappa \frac{\sqrt{\log L_B}}{L_B^{1/6}}, \quad (11)$$

for some constant  $\kappa \leq 27 \max(c, b - a)$ .

Our proof, given in [44], uses finite-degree polynomial approximations of Lipschitz functions [50], combined with aforementioned concepts from concentration of measure [48]. Theorem 3 implies that, for large  $L_B$ , a full histogram of  $\theta$  in the PE will be very close to the corresponding histogram in the GHE (provided the bins are not too narrow) [77]. Physically relevant functions  $\theta$  that satisfy the criteria above include  $n$ -point correlation functions  $\langle \gamma_{j_1} \cdots \gamma_{j_n} \rangle$ , subsystem Rényi entropies, and von Neumann entanglement entropies, the latter of which is not directly related to any integer moment  $\mathcal{E}_{\text{PE}}^{(k)}$ .

**Numerics**—Our rigorous results support the claim that a state  $|\Psi\rangle$  prepared by a random Gaussian unitary will typically exhibit deep thermalization with respect to  $\mathcal{E}_{\text{GHE}}$  (2). Here, we supplement these theorems with numerical results, both for globally and locally random Gaussian  $(1+1)D$  circuits. Thanks to (10), we can evaluate the Wasserstein-1 distance for the distributions of physical

quantities  $\theta(\Gamma_{\mathbf{m}}^A)$  efficiently (i.e., without having to consider all  $2^{L_B}$  possible outcomes), as follows. We first get an estimate of the CDF  $P_{\text{PE}}(\theta)$  by generating measurement outcomes  $\mathbf{m}^{(i)}$  ( $i = 1, \dots, R$ ) on  $B$  [44], computing  $\Gamma_{\mathbf{m}^{(i)}}^A$ , and from these  $\theta^{(i)} = \theta(\Gamma_{\mathbf{m}^{(i)}}^A)$ . This determines an empirical CDF  $\hat{P}_{\text{PE}}(\theta) = (1/R) \sum_i \Theta(\theta^{(i)} - \theta)$ , which can be combined with analytically known or empirically estimated forms for  $P_{\text{GHE}}$  to get an estimate  $W_1(\hat{P}_{\text{PE}}, P_{\text{GHE}})$ . We then average this over  $N$  circuit instances to estimate  $\mathbb{E}_{\Psi} W_1$ .

**Global dynamics**—First, we take global Gaussian random unitaries, and look at the distribution of a particular correlation function  $\theta = \langle i\gamma_1\gamma_2 \rangle_{\mathbf{m}} = [\Gamma_{\mathbf{m}}^A]_{12}$ , using analytically known expressions for  $P_{\text{GHE}}(\theta)$  [44]. Taking  $L_A = 2$ , our results agree well with a power-law decay  $W_1 \propto 1/\sqrt{L_B}$ . When  $R$  is finite, this saturates at large  $L_B$  due to the finite sample size [Figs. 2(a) and 2(b)]. Similar results hold for other values of  $L_A$  [44]. This behavior is consistent with the bound  $W_1 \lesssim L_B^{-1/6}$  from Theorem 3, but does not saturate it; thus, it may be possible to quantitatively improve (11). By analogy to Ref. [8], an estimate of the plateau value can be obtained by considering an empirical CDF constructed from  $R$  GHE realizations; its discrepancy is shown as the horizontal line in Fig. 2(a), and plotted in Fig. 2(b). Taking  $R$  to scale linearly  $L_B$  alleviates this numerical artifact, see End Matter.

**Local dynamics**—We now take local  $(1+1)D$  circuits, with each gate a random FLO unitary, which can be implemented by matchgates [36–38]. Since any Gaussian unitary can be implemented by a linear-depth  $(1+1)D$  circuit [78,79], we anticipate that random local Gaussian unitaries will converge toward globally random Gaussian unitaries reasonably quickly; thus, we expect DT at long timescales. Indeed, our simulations suggest a power-law decay  $W_1(t) \propto t^{-1/4}$  [Figs. 2(c) and 2(d)] down to a late-time value set by the size of the measured subsystem  $L_B$  following  $W_1(t \gg 1) \propto 1/\sqrt{L_B}$  [Figs. 2(a) and 2(c)]. Here, we choose  $R$  large enough such that the empirical CDF  $\hat{P}_{\text{PE}}$

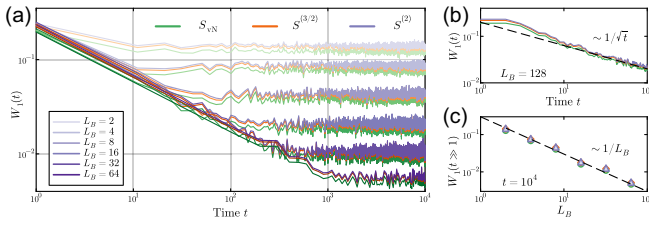


FIG. 3. Deep thermalization of local fermionic Gaussian dynamics: Wasserstein-1 distance  $W_1$  between distributions of subsystem von Neumann or Rényi entropy  $\theta(\Gamma_m^A) = S^{(\alpha)}(\rho_m^{A_1})$  for  $L_{A_1} = (L_A/2) = 1$ ,  $N = 10^2$  circuit instances. (a)  $W_1(t)$  saturates as  $\propto 1/\sqrt{t}$ , down to a value set by the measured subsystem size  $L_B$ . Here,  $R = 10^4$ , and relative errors are all below 12%. (b) Closer view of the decay  $W_1(t) \propto 1/\sqrt{t}$  (dashed line) for  $R = 10^3$ . (c) Scaling of  $W_1(t \gg 1) \propto 1/L_B$  (dashed line), imperceptible error bars (SE).

approximates well  $P_{\text{PE}}$ , see End Matter. We also verified this  $W_1(t) \propto t^{-1/4}$  behavior for other choices of  $L_A$  [44].

*From local to global dynamics via diffusion*—How fast information spreads under local dynamics sets the time dependence of  $W_1(t)$ , and we can relate  $W_1(t) \propto t^{-1/4}$  for local dynamics to the observed  $W_1 \propto 1/\sqrt{L_B}$  for global dynamics. Since the dynamics does not conserve the particle number, this spreading is not associated with any conserved charge. Instead, it simply quantifies the spreading of correlations. We support diffusive spreading with numerical evidence in [44]. Diffusive spreading implies that at time  $t$ , the effective number of measured modes influencing the PE is  $\ell \sim \sqrt{t}$ . By  $W_1 \propto 1/\sqrt{L_B}$  for  $L_B$  measured modes for global dynamics, we expect  $W_1(t) \propto 1/\sqrt{\ell} \sim t^{-1/4}$  for local dynamics, which agrees with our simulations [Fig. 2(d)].

*Entanglement entropies*—To further test DT, we consider the distribution of subsystem von Neumann and Rényi entanglement entropies  $\theta = S_{\text{vN}}, S^{(\alpha)}$ . Since we cannot access the analytical form for  $P_{\text{GHE}}(\theta)$ , we build an empirical  $\hat{P}_{\text{GHE}}(\theta)$  from  $10^8$  realizations. Taking  $L_A = 2$ ,  $L_{A_1} = L_A/2$  and computing  $S^{(\alpha)}(\rho_m^{A_1})$  for  $\alpha \in \{1, \frac{3}{2}, 2\}$ , our numerics suggest a power-law decay  $W_1(t) \propto 1/\sqrt{t}$  [Figs. 3(a) and 3(b)], approaching a late-time value  $W_1(t \gg 1) \propto 1/L_B$  [Fig. 3(c)], for all  $\alpha \geq 1$ . We can again understand the relaxation from local to global dynamics behavior of  $W_1$  through the diffusion argument  $W_1(t) \propto 1/\ell \sim 1/\sqrt{t}$ . The late-time decay of  $W_1$  for  $\alpha \in \{1, \frac{3}{2}, 2\}$  is consistent with the bound  $W_1 \lesssim L_B^{-1/6}$  from Theorem 3, but does not saturate it.

*Discussion and outlook*—We have rigorously shown that random free-fermion states deeply thermalize. Specifically, we showed that the projected ensemble tends toward a particular universal distribution—the Gaussian Haar ensemble—at the level of both statistical moments (in keeping with previous methodology [8,10–13]), but also

using the more general Wasserstein-1 distance  $W_1$ . Our rigorous theorems are complemented by numerical simulations, which also show a power-law convergence with depth  $t$  for local circuits, which we attribute to the diffusive spreading of quantum information. A key distinction between Gaussian and universal dynamics seems to be that the former typically shows power-law decay of  $W_1$  and moments, while the latter shows exponential decay, with respect to  $L_B$  or  $t$ .

Because matchgate circuits are efficiently simulable, these phenomena can be experimentally probed in a scalable manner, using classical side-processing to avoid postselection issues [41,42]. This form of DT, therefore, offers a versatile way to benchmark quantum simulators. While similar approaches might apply to Clifford circuits as well, we find that the PE there exhibits a much more trivial structure, forming at most a 1-design [44]. Moving forward, it would be interesting to study how these phenomena are enriched upon doping with a small number of non-Gaussian gates. In addition, using this formalism to identify observables whose distributions show a stark contrast between Gaussian and universal dynamics could help study phase transitions between classically efficiently simulable and hard-to-simulate regimes of quantum dynamics. Finally, expanding our analysis to account for realistic experimental imperfections by including the effects of noise and measurement errors would be appealing.

*Acknowledgments*—We thank Jan Behrends and Wen Wei Ho for useful discussions. This work was supported by EPSRC Ph.D. Studentship 2606484, EPSRC Grant No. EP/V062654/1, and by Trinity College, Cambridge. Our simulations used resources at the Cambridge Service for Data Driven Discovery operated by the University of Cambridge Research Computing Service [80], provided by Dell EMC and Intel using EPSRC Tier-2 funding via Grant No. EP/T022159/1, and STFC DiRAC funding [81].

*Data availability*—The data that support the findings of this Letter are not publicly available. The data are available from the authors upon reasonable request.

- [1] J. M. Deutsch, *Phys. Rev. A* **43**, 2046 (1991).
- [2] M. Srednicki, *Phys. Rev. E* **50**, 888 (1994).
- [3] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, *Science* **353**, 794 (2016).
- [4] C. Neill *et al.*, *Nat. Phys.* **12**, 1037 (2016).
- [5] S. Popescu, A. J. Short, and A. Winter, *Nat. Phys.* **2**, 754 (2006).
- [6] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghi, *Phys. Rev. Lett.* **96**, 050403 (2006).
- [7] E. Altman *et al.*, *PRX Quantum* **2**, 017003 (2021).
- [8] J. S. Cotler, D. K. Mark, H.-Y. Huang, F. Hernández, J. Choi, A. L. Shaw, M. Endres, and S. Choi, *PRX Quantum* **4**, 010311 (2023).

- [9] J. Choi *et al.*, *Nature (London)* **613**, 468 (2023).
- [10] W. W. Ho and S. Choi, *Phys. Rev. Lett.* **128**, 060601 (2022).
- [11] M. Ippoliti and W. W. Ho, *Quantum* **6**, 886 (2022).
- [12] P. W. Claeys and A. Lamacraft, *Quantum* **6**, 738 (2022).
- [13] M. McGinley and M. Fava, *Phys. Rev. Lett.* **131**, 160601 (2023).
- [14] T. Bhore, J.-Y. Desaules, and Z. Papić, *Phys. Rev. B* **108**, 104317 (2023).
- [15] M. Ippoliti and W. W. Ho, *PRX Quantum* **4**, 030322 (2023).
- [16] M. Lucas, L. Piroli, J. De Nardis, and A. De Luca, *Phys. Rev. A* **107**, 032215 (2023).
- [17] H. Shrotriya and W. W. Ho, *SciPost Phys.* **18**, 107 (2025).
- [18] C. Vairogs and B. Yan, *arXiv:2402.10181*.
- [19] D. K. Mark, F. Surace, A. Elben, A. L. Shaw, J. Choi, G. Refael, M. Endres, and S. Choi, *Phys. Rev. X* **14**, 041051 (2024).
- [20] C. Liu, Q. C. Huang, and W. W. Ho, *Phys. Rev. Lett.* **133**, 260401 (2024).
- [21] R.-A. Chang, H. Shrotriya, W. W. Ho, and M. Ippoliti, *arXiv:2408.15325*.
- [22] N. D. Varikuti and S. Bandyopadhyay, *Quantum* **8**, 1456 (2024).
- [23] F. Verstraete, M. Popp, and J. I. Cirac, *Phys. Rev. Lett.* **92**, 027901 (2004).
- [24] M. Popp, F. Verstraete, M. A. Martín-Delgado, and J. I. Cirac, *Phys. Rev. A* **71**, 042306 (2005).
- [25] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghì, *J. Stat. Phys.* **125**, 1193 (2006).
- [26] S. Goldstein, J. L. Lebowitz, C. Mastrodonato, R. Tumulka, and N. Zanghì, *Commun. Math. Phys.* **342**, 965 (2015).
- [27] P. S. Turner and D. Markham, *Phys. Rev. Lett.* **116**, 200501 (2016).
- [28] R. Mezher, J. Ghalbouni, J. Dgheim, and D. Markham, *Phys. Rev. A* **97**, 022333 (2018).
- [29] Y. Li, X. Chen, and M. P. A. Fisher, *Phys. Rev. B* **98**, 205136 (2018).
- [30] Y. Li, X. Chen, and M. P. A. Fisher, *Phys. Rev. B* **100**, 134306 (2019).
- [31] S. Choi, Y. Bao, X.-L. Qi, and E. Altman, *Phys. Rev. Lett.* **125**, 030505 (2020).
- [32] M. Ippoliti, M. J. Gullans, S. Gopalakrishnan, D. A. Huse, and V. Khemani, *Phys. Rev. X* **11**, 011030 (2021).
- [33] M. P. A. Fisher, V. Khemani, A. Nahum, and S. Vijay, *Annu. Rev. Condens. Matter Phys.* **14**, 335 (2023).
- [34] B. Skinner, J. Ruhman, and A. Nahum, *Phys. Rev. X* **9**, 031009 (2019).
- [35] L. Versini, K. A. El-Din, F. Mintert, and R. Mukherjee, *arXiv:2305.01465*.
- [36] L. G. Valiant, *SIAM J. Comput.* **31**, 1229 (2002).
- [37] R. Jozsa and A. Miyake, *Proc. R. Soc. A* **464**, 3089 (2008).
- [38] B. M. Terhal and D. P. DiVincenzo, *Phys. Rev. A* **65**, 032325 (2002).
- [39] S. Bravyi, *Quantum Inf. Comput.* **5**, 216 (2005).
- [40] M. Ippoliti and V. Khemani, *Phys. Rev. Lett.* **126**, 060501 (2021).
- [41] S. J. Garratt and E. Altman, *PRX Quantum* **5**, 030311 (2024).
- [42] M. McGinley, *PRX Quantum* **5**, 020347 (2024).
- [43] S. B. Bravyi and A. Y. Kitaev, *Ann. Phys. (Amsterdam)* **298**, 210 (2002).
- [44] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/v8kp-39ry> for a review of the covariance matrix formalism; a sampling procedure for the projected ensemble; a discussion of the singular values of subsystem covariance matrix; proofs of Theorems 1, 2, and 3; a more in-depth discussion of the Wasserstein-1 distance; details on the diffusive spreading of correlations; additional numerics; and a comparison between matchgate-, Clifford-, and dual-unitary-dominated circuits. Contains Refs. [8,10,12,15,39,45–71].
- [45] A. Y. Kitaev, *Phys. Usp.* **44**, 131 (2001).
- [46] A. Altland and M. R. Zirnbauer, *Phys. Rev. B* **55**, 1142 (1997).
- [47] J. P. Dahlhaus, B. Béri, and C. W. J. Beenakker, *Phys. Rev. B* **82**, 014536 (2010).
- [48] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices* (Cambridge University Press, Cambridge, England, 2009).
- [49] C. Villani, *Optimal Transport: Old and New* (Springer, New York, 2009).
- [50] D. Jackson, *Trans. Am. Math. Soc.* **13**, 491 (1912).
- [51] N. Fournier and A. Guillin, *Probab. Theory Relat. Fields* **162**, 707 (2015).
- [52] F. Mezzadri, *Not. Am. Math. Soc.* **54**, 592 (2007).
- [53] E. Bianchi, L. Hackl, and M. Kieburg, *Phys. Rev. B* **103**, L241118 (2021).
- [54] N. Schuch and B. Bauer, *Phys. Rev. B* **100**, 245121 (2019).
- [55] C.-M. Jian, B. Bauer, A. Keselman, and A. W. W. Ludwig, *Phys. Rev. B* **106**, 134206 (2022).
- [56] C. W. J. Beenakker, *Rev. Mod. Phys.* **69**, 731 (1997).
- [57] I. C. Fulga, F. Hassler, A. R. Akhmerov, and C. W. J. Beenakker, *Phys. Rev. B* **83**, 155429 (2011).
- [58] A. Botero and B. Reznik, *Phys. Lett. A* **331**, 39 (2004).
- [59] J. Surace and L. Tagliacozzo, *SciPost Phys. Lect. Notes* **54** (2022).
- [60] L. Kantorovich and G. S. Rubinstein, *Vestn. Leningr. Univ. Fiz. Khim.* **13**, 52 (1958).
- [61] R. M. Gray, *Entropy and Information Theory* (Springer, New York, 2011).
- [62] N. Carothers, A short course on approximation theory, lecture notes for Bowling Green State University (2009), <http://fourier.math.uoc.gr/~mk/approx1011/carothers.pdf> (accessed Oct. 2024).
- [63] P. Hayden, D. W. Leung, and A. Winter, *Commun. Math. Phys.* **265**, 95 (2006).
- [64] G. Roósz, R. Juhász, and F. Iglói, *Phys. Rev. B* **93**, 134305 (2016).
- [65] X.-H. Yu, Z. Gong, and J. I. Cirac, *Phys. Rev. Res.* **5**, 013044 (2023).
- [66] D. N. Page, *Phys. Rev. Lett.* **71**, 1291 (1993).
- [67] E. Bianchi and P. Donà, *Phys. Rev. D* **100**, 105010 (2019).
- [68] B. Dias and R. Koenig, *Quantum* **8**, 1350 (2024).
- [69] I. L. Markov and Y. Shi, *SIAM J. Comput.* **38**, 963 (2008).
- [70] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, England, 2010).
- [71] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, *J. Math. Phys. (N.Y.)* **45**, 2171 (2004).
- [72] C. W. Helstrom, *J. Stat. Phys.* **1**, 231 (1969).

- [73] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, Ltd, New York, 2005).
- [74] S. Kullback, *Information Theory and Statistics* (Wiley & Sons, Ltd., New York, 1959).
- [75] L. Kantorovic, *Dokl. Akad. Nauk SSSR* **37**, 227 (1942).
- [76] S. S. Vallender, *Theory Probab. Appl.* **18**, 784 (1974).
- [77] J. Niles-Weed and Q. Berthet, *Ann. Stat.* **50**, 1519 (2022).
- [78] Z. Jiang, K. J. Sung, K. Kechedzhi, V. N. Smelyanskiy, and S. Boixo, *Phys. Rev. Appl.* **9**, 044036 (2018).
- [79] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, Garnet Kin-Lic Chan, and R. Babbush, *Phys. Rev. Lett.* **120**, 110501 (2018).
- [80] <http://www.csd3.cam.ac.uk>
- [81] <http://www.dirac.ac.uk>

## End Matter

*Numerical sample complexity*—In this Appendix, we emphasize that accurate numerical estimates of the Wasserstein-1 distance can, in general, be obtained using a reasonable sample size  $R$ . Theorem 2 of Ref. [51] implies that for finite  $R$ , the empirical and exact CDFs,  $\hat{P}_{\text{PE}}^{(R)}$  and  $P_{\text{PE}}$ , for the projected ensemble are themselves close in Wasserstein-1 distance with high probability:  $\mathbb{P}(W_1(\hat{P}_{\text{PE}}^{(R)}, P_{\text{PE}}) > \epsilon) \leq e^{-cR\epsilon^2}$  for a constant  $c$ . Combined with the triangle inequality

$W_1(\hat{P}_{\text{PE}}^{(R)}, P_{\text{GHE}}) \leq W_1(\hat{P}_{\text{PE}}^{(R)}, P_{\text{PE}}) + W_1(P_{\text{PE}}, P_{\text{GHE}})$  (Ref. [49], Chap. 6), this implies that our estimate will be  $\epsilon$ -close to the true value  $W_1(P_{\text{PE}}, P_{\text{GHE}})$  with probability at least  $1 - \delta$  provided  $R \geq c'\epsilon^{-2} \log(1/\delta)$  for a constant  $c'$ . Given that  $W_1(P_{\text{PE}}, P_{\text{GHE}})$  is observed to decay as  $1/\text{poly}(L_B)$ , we can get an arbitrarily small relative error using  $R = \text{poly}(L_B)$ . For the data in Fig. 2(a), suggesting  $W_1(P_{\text{PE}}, P_{\text{GHE}}) \sim L_B^{-1/2}$ , taking  $R \propto L_B$  suffices.