Editors' Suggestion   Featured in Physics

Quantum engine based on many-body localization

Nicole Yunger Halpern,[1,2,*] Christopher David White,[1,2,†] Sarang Gopalakrishnan,[1,2,3,4,‡] and Gil Refael[1,2,3,§]

[1]*Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA*
[2]*Department of Physics, California Institute of Technology, Pasadena, California 91125, USA*
[3]*Walter Burke Institute, California Institute of Technology, Pasadena, California 91125, USA*
[4]*College of Staten Island, City University of New York, Staten Island, New York 10314, USA*

Many-body-localized (MBL) systems do not thermalize under their intrinsic dynamics. The athermality of MBL, we propose, can be harnessed for thermodynamic tasks. We illustrate this ability by formulating an Otto engine cycle for a quantum many-body system. The system is ramped between a strongly localized MBL regime and a thermal (or weakly localized) regime. The difference between the energy-level correlations of MBL systems and of thermal systems enables mesoscale engines to run in parallel in the thermodynamic limit, enhances the engine's reliability, and suppresses worst-case trials. We estimate analytically and calculate numerically the engine's efficiency and per-cycle power. The efficiency mirrors the efficiency of the conventional thermodynamic Otto engine. The per-cycle power scales linearly with the system size and inverse-exponentially with a localization length. This work introduces a thermodynamic lens onto MBL, which, having been studied much recently, can now be considered for use in thermodynamic tasks.

DOI: 10.1103/PhysRevB.99.024203

I. INTRODUCTION

Many-body localization (MBL) has emerged as a unique phase in which an isolated interacting quantum system resists internal thermalization for long times. MBL systems are integrable and have local integrals of motion [1], which retain information about initial conditions for long times or even indefinitely [2]. This and other aspects of MBL were recently observed experimentally [3–10]. In contrast, in thermalizing isolated quantum systems, information and energy can diffuse easily. Such systems obey the eigenstate thermalization hypothesis (ETH) [11–14].

A tantalizing question is whether the unique properties of MBL could be utilized. So far, MBL has been proposed to be used in robust quantum memories [15]. We believe, however, that the potential of MBL is much greater. MBL systems behave athermally, and athermality (lack of thermal equilibrium) facilitates thermodynamic tasks [16–26]. When a cold bath is put in contact with a hot environment, for instance, work can be extracted from the heat flow. Could MBL's athermality have thermodynamic applications?

We present one by formulating, analyzing, and numerically simulating an Otto engine cycle for a quantum many-body system that has an MBL phase. The engine contacts a hot bath and a narrow-bandwidth cold bath, as sketched in Fig. 1.

This application unites the growing fields of quantum thermal machines [27–39] and MBL [1,15,40–43]. Our proposal could conceivably be explored in ultracold-atom [3,4,6,7,10], nitrogen-vacancy-center [8], trapped-ion [9], and possibly doped-semiconductor [44] experiments.

Our engine relies on two properties that distinguish MBL from thermal systems: its spectral correlations [43,45] and its localization. The spectral-correlation properties enable us to build a mesoscale level-statistics engine. The localization enables us to link mesoscale engines together, creating a large engine with an extensive work output.

Take an interacting finite spin chain as an example. Consider the statistics of the gaps between consecutive energy eigenvalues far from the energy band's edges. A gap distribution $P(\delta)$ encodes the probability that any given gap has size δ. The MBL gap distribution enables small (and large) gaps to appear much more often than in ETH spectra [46]. This difference enables MBL to enhance our quantum many-body Otto cycle.

Let us introduce the MBL and ETH distributions in greater detail. Let $\langle\delta\rangle_E$ denote the average gap at the energy E. MBL gaps approximately obey Poisson statistics [41,46]:

$$P_{\text{MBL}}^{(E)}(\delta) \approx \frac{1}{\langle\delta\rangle_E} e^{-\delta/\langle\delta\rangle_E} . \tag{1}$$

Any given gap has a decent chance of being small: as $\delta \to 0$, $P_{\text{MBL}}^{(E)}(\delta) \to \frac{1}{\langle\delta\rangle_E} > 0$. Neighboring energies have finite probabilities of lying close together: MBL systems' energies do not repel each other, unlike thermal systems' energies. Thermalizing systems governed by real Hamiltonians obey the level statistics of random matrices drawn from the Gaussian orthogonal ensemble (GOE) [41]:

$$P_{\text{GOE}}^{(E)}(\delta) \approx \frac{\pi}{2} \frac{\delta}{\langle\delta\rangle_E^2} e^{-\frac{\pi}{4}\delta^2/\langle\delta\rangle_E^2} . \tag{2}$$
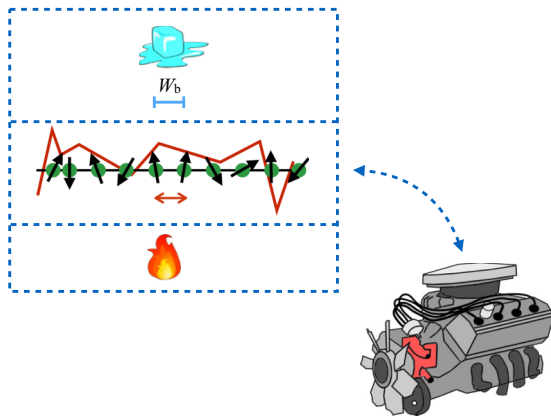
FIG. 1. Schematic of MBL engine. We formulate an Otto engine cycle for a many-body quantum system that exhibits an MBL phase. We illustrate the engine with a spin chain (green dots and black arrows). A random disorder potential (jagged red line) localizes the particles. Particles interact and hop between sites (horizontal red arrows). Consider strengthening the interactions and the hopping frequency. The system transitions from strong localization to a thermal phase or to weak localization. The engine thermalizes with a hot bath (flames) and with a cold bath (ice cube). The cold bath has a small bandwidth W_b, to take advantage of small energy gaps' greater prevalence in the highly localized regime.

Unlike in MBL spectra, small gaps rarely appear; as $\delta \to 0$, $P_{\mathrm{GOE}}^{(E)}(\delta) \to 0$.

MBL's athermal gap statistics should be construed as a thermodynamic resource, we find, as athermal quantum states have been [16–26]. In particular, MBL's athermal gap statistics improve our engine's reliability: the amount W_{tot} of work extracted by our engine fluctuates relatively little from successful trial to successful trial. Athermal statistics also lower the probability of worst-case trials, in which the engine outputs net negative work, $W_{\mathrm{tot}} < 0$. Furthermore, MBL's localization enables the engine to scale robustly; mesoscale "subengines" can run in parallel without disturbing each other much, due to the localization inherent in MBL. Even in the thermodynamic limit, an MBL system behaves like an ensemble of finite, mesoscale quantum systems, due to its *local level correlations* [45,47,48]. Any local operator can probe only a discrete set of sharp energy levels, which emerge from its direct environment.

This paper is organized as follows. Section II contains background about the Otto cycle and about quantum work and heat. In Sec. III, we introduce the mesoscopic MBL engine. In Sec. III A, we introduce the basic idea with a qubit (two-level quantum system). In Sec. III B, we scale the engine up to a mesoscopic chain tuned between MBL and ETH regimes. In Sec. III C, we calculate the engine's work output and efficiency. In Sec. IV, we argue that the mesoscopic segments can be combined into a macroscopic MBL system while operating in parallel. In Sec. V, we discuss limitations on the speed at which the engine can be run and, consequently, the engine's power. This leads us to a more careful consideration of diabatic corrections to the work output, communication amongst subengines, and the cold bath's nature. We test our

analytic calculations in Sec. VI, with numerical simulations of disordered spin chains. In Sec. VII, we provide order-of-magnitude estimates for a localized semiconductor engine's power and power density.

II. THERMODYNAMIC BACKGROUND

The classical Otto engine consists of a gas that expands, cools, contracts, and heats [50]. During the two isentropic (constant-entropy) strokes, the gas volume is tuned between values V_1 and $V_2 < V_1$. The *compression ratio* is defined as $r := \frac{V_1}{V_2}$. The heating and cooling are isochoric (constant-volume). The engine outputs a net amount W_{tot} of work per cycle, absorbing heat $Q_{\mathrm{in}} > 0$ during the heating isochore.

A general engine's thermodynamic efficiency is

$$\eta := \frac{W_{\mathrm{tot}}}{Q_{\mathrm{in}}}. \tag{3}$$

The Otto engine operates at the efficiency

$$\eta_{\mathrm{Otto}} = 1 - \frac{1}{r^{\gamma-1}} < \eta_{\mathrm{Carnot}}. \tag{4}$$

A ratio of the gas's constant-pressure and constant-volume specific heats is denoted by $\gamma := \frac{C_{\mathrm{P}}}{C_{\mathrm{v}}}$. The Carnot efficiency η_{Carnot} upper bounds the efficiency of every thermodynamic engine that involves just two heat baths.

An Otto cycle for quantum harmonic oscillators (QHOs) was discussed in Refs. [28,36,51–56]. The QHO's gap plays the role of the classical Otto engine's volume. Let ω and $\Omega > \omega$ denote the values between which the angular frequency is tuned. The ideal QHO Otto cycle operates at the efficiency

$$\eta_{\mathrm{QHO}} = 1 - \frac{\omega}{\Omega}. \tag{5}$$

This oscillator model resembles the qubit toy model that informs our MBL Otto cycle (Sec. III A). The energy eigenbasis changes in our model, however, and the engine scales robustly to macroscopically many qubits.

Consider tuning an open system, slowly, between times $t = 0$ and $t = \tau$. The heat and work absorbed are defined as

$$W := \int_0^\tau dt \, \mathrm{Tr}\left(\rho \, \frac{dH}{dt}\right) \quad \text{and} \tag{6}$$

$$Q := \int_0^\tau dt \, \mathrm{Tr}\left(\frac{d\rho}{dt} \, H\right) \tag{7}$$

in quantum thermodynamics [56]. This Q definition is narrower than the definition prevalent in the MBL literature [46,57–59]. Here, all energy exchanged during unitary evolution counts as work.

III. A MESOSCALE MBL ENGINE

We aim to formulate an MBL engine cycle for the thermodynamic limit. Our road to that goal runs through a finite-size, or mesoscale, MBL engine. In Sec. III A, we introduce the intuition behind the mesoscale engine via a qubit toy model. Then, we describe (Sec. III B) and quantitatively analyze (Sec. III C) the mesoscale MBL engine. Table I offers a spotter's guide to notation.

TABLE I. Parameters of the mesoscopic and macroscopic MBL engines (introduced in Secs. III and IV). Boltzmann's and Planck's constants are set to one: $k_B = \hbar = 1$.

Symbol	Significance
N	Number of sites per mesoscale engine (in Sec. III) or per mesoscale subengine (in the macroscopic engine, in Sec. IV). Chosen, in the latter case, to equal $\xi_>$.
$\mathcal{N}$	Dimensionality of one mesoscale (sub)engine's Hilbert space.
$\mathcal{E}$	Unit of energy, average energy density per site.
α_t	Hamiltonian parameter tuned from 0 (in the mesoscale engine's ETH regime, or the macroscopic engine's shallowly localized regime) to 1 (in the engine's deeply MBL regime).
$\langle \delta \rangle$	Average gap in the energy spectrum of a length-N MBL system.
W_b	Bandwidth of the cold bath. Small: $W_b \ll \langle \delta \rangle$.
$\beta_H = 1/T_H$	Inverse temperature of the hot bath.
$\beta_C = 1/T_C$	Inverse temperature of the cold bath.
δ_-	Level-repulsion scale of a length-N MBL system. Minimal size reasonably attributable to any energy gap. Smallest gap size at which a Poissonian (1) approximates the MBL gap distribution well.
v	Speed at which the Hamiltonian is tuned: $v := \mathcal{E} \frac{d\alpha_t}{t}$. Has dimensions of 1/time^2, in accordance with part of [64].
$\xi_>$	Localization length of macroscopic MBL engine when shallowly localized. Length of mesoscale subengine.
$\xi_<$	Localization length of macroscopic MBL engine when deeply localized. Satisfies $\xi_< < \xi_>$.
X_{macro}	Characteristic X of the macroscopic MBL engine (e.g., $X = N, \langle \delta \rangle$).
g	Strength of coupling between engine and cold bath.
τ_{cycle}	Time required to implement one cycle.
$\langle \delta \rangle^{(L)}$	Average energy gap of a length-L MBL system.

A. Qubit toy model

At the MBL Otto engine's core lies a qubit Otto engine whose energy eigenbasis transforms during the cycle [60–63]. Consider a two-level system evolving under the time-varying Hamiltonian

$$H_{qubit}(t) := (1 - \alpha_t)h\sigma^x + \alpha_t h'\sigma^z . \qquad (8)$$

σ^x and σ^z denote the Pauli x and z operators. α_t denotes a parameter tuned between 0 and 1.

Figure 2 illustrates the cycle. The engine begins in thermal equilibrium at a high temperature T_H. During stroke 1, the engine is thermally isolated, and α_t is tuned from 0 to 1. During stroke 2, the engine thermalizes to a temperature $T_C \ll T_H$. During stroke 3, the engine is thermally isolated, and α_t returns from 1 to 0. During stroke 4, the engine resets by thermalizing with the hot bath.

Let us make two simplifying assumptions (see Ref. [[49], Appendix C] for a generalization): first, let $T_H = \infty$ and $T_C = 0$. Second, assume that the engine is tuned slowly enough to satisfy the quantum adiabatic theorem. We also choose[1]

$$h = \frac{\delta_{GOE}}{2} , \quad h' = \frac{\delta_{MBL}}{2} ,$$

and $\delta_{GOE} \gg \delta_{MBL}$.

---

[1] The gaps' labels are suggestive; a qubit, having only one gap, obeys neither GOE nor MBL gap statistics. However, when large, the qubit gap apes a typical GOE gap; and, when small, the qubit gap apes a useful MBL gap. This mimicry illustrates how the mesoscopic engine benefits from the greater prevalence of small gaps in MBL spectra than in GOE spectra.

Let us analyze the cycle's energetics. The system begins with $\langle H_{qubit}(t) \rangle = 0$. Stroke 1 preserves the
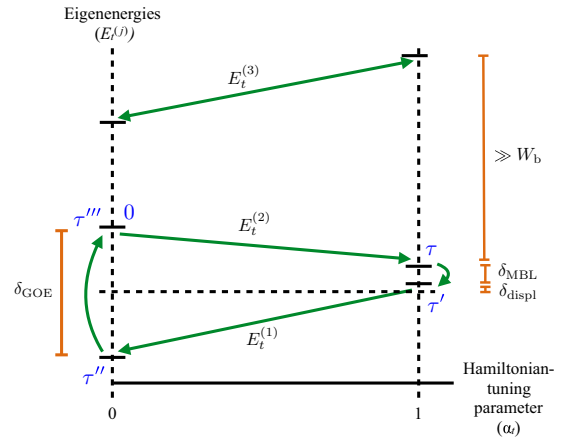

FIG. 2. Qubit toy model for the MBL Otto cycle. A qubit models two "working levels" in the MBL Otto engine's many-body spectrum. The energy eigenstates $|E_t^{(1)}\rangle$ and $|E_t^{(2)}\rangle$ span the "working subspace." The gap $E_t^{(2)} - E_t^{(1)}$ begins at size δ_{GOE} during a successful trial. The gap shrinks to δ_{MBL}, then returns to δ_{GOE}. In addition to changing the gap, each Hamiltonian tuning changes the eigenstates' functional forms. The displacement δ_{displ} is included for generality. The blue text marks the times $t = 0, \tau, \dots, \tau'''$ at which the strokes begin and end during a work-outputting trial. The spectator level $|E_t^{(3)}\rangle$ fails to impact the engine's efficiency. The cold bath has too narrow a bandwidth W_b to couple $|E_t^{(3)}\rangle$ to any other level. If the engine begins any trial on the top green line, the engine remains on that line throughout the trial. Zero net work is outputted.

infinite-temperature state $\mathbb{1}/2$. The energy drops to $-\delta_{\text{MBL}}/2$ during stroke 2 and to $-\delta_{\text{GOE}}/2$ during stroke 3. During stroke 4, the engine resets to zero average energy, absorbing heat $\langle Q_4 \rangle = \frac{\delta_{\text{GOE}}}{2}$, on average.

The energy exchanged during the tunings (strokes 1 and 3) constitutes work [Eq. (6)], while the energy exchanged during the thermalizations (strokes 2 and 4) is heat [Eq. (7)]. The engine outputs the *per-cycle power*, or average work performed per cycle, $\langle W_{\text{tot}} \rangle = \frac{1}{2}(\delta_{\text{GOE}} - \delta_{\text{MBL}})$.

The efficiency is $\eta_{\text{qubit}} = \frac{\langle W_{\text{tot}} \rangle}{\langle Q_4 \rangle} = 1 - \frac{\delta_{\text{MBL}}}{\delta_{\text{GOE}}}$. This result is equivalent to the efficiency η_{Otto} of a thermodynamic Otto engine [Eq. (4)]. The gap ratio $\frac{\delta_{\text{MBL}}}{\delta_{\text{GOE}}}$ plays the role of $\frac{1}{r^{\gamma-1}}$. η_{qubit} equals also η_{QHO} [Eq. (5)] if the frequency ratio ω/Ω is chosen to equal $\delta_{\text{MBL}}/\delta_{\text{GOE}}$. As shown in Secs. III and IV, however, the qubit engine can scale to a large composite engine of densely packed qubit subengines operating in parallel. The dense packing is possible if the qubits are encoded in the MBL system's localized degrees of freedom (l-bits, roughly speaking [1]).

B. Setup for the mesoscale MBL engine

The next step is an interacting finite-size system tuned between MBL and ETH phases. Envision a mesoscale engine as a one-dimensional (1D) system of $N \approx 10$ sites. This engine will ultimately model one region in a thermodynamically large MBL engine. We will analyze the mesoscopic engine's per-trial power $\langle W_{\text{tot}} \rangle$, the efficiency η_{MBL}, and work costs $\langle W_{\text{diab}} \rangle$ of undesirable diabatic transitions.

The mesoscopic engine evolves under the Hamiltonian

$$H_{\text{meso}}(t) := \frac{\mathcal{E}}{Q(\alpha_t)}[(1 - \alpha_t)H_{\text{GOE}} + \alpha_t H_{\text{MBL}}]. \quad (9)$$

The unit of energy, or average energy density per site, is denoted by $\mathcal{E}$. The tuning parameter $\alpha_t \in [0, 1]$. When $\alpha_t = 0$, the system evolves under a random Hamiltonian H_{GOE} whose gaps δ are distributed according to $P_{\text{GOE}}^{(E)}(\delta)$ [Eq. (2)]. When $\alpha_t = 1$, $H_{\text{meso}}(t) = H_{\text{MBL}}$, a Hamiltonian whose gaps are distributed according to $P_{\text{MBL}}^{(E)}(\delta)$ [Eq. (1)]. For a concrete example, take a random-field Heisenberg model whose disorder strength is tuned. H_{GOE} and H_{MBL} have the same bond term, but the disorder strength varies in time. We simulate (a rescaled version of) this model in Sec. VI.

The mesoscale engine's cycle is analogous to the qubit cycle, including initialization at $\alpha_t = 0$, tuning of α_t to one, thermalization with a temperature-T_C bath, tuning of α_t to zero, and thermalization [65–68] with a temperature-T_H bath. To highlight the role of level statistics in the cycle, we hold the average energy gap, $\langle \delta \rangle$, constant.[2] We do so using the

renormalization factor $Q(\alpha_t)$.[3] Section VI details how we define $Q(\alpha_t)$ in numerical simulations.

The key distinction between GOE level statistics (2) and Poisson (MBL) statistics (1) is that small gaps (and large gaps) appear more often in Poisson spectra. A toy model illuminates these level statistics' physical origin: an MBL system can be modeled as a set of noninteracting quasilocal qubits [1]. Let g_j denote the jth qubit's gap. Two qubits, j and j', may have nearly equal gaps: $g_j \approx g_{j'}$. The difference $|g_j - g_{j'}|$ equals a gap in the many-body energy spectrum. Tuning the Hamiltonian from MBL to ETH couples the qubits together, producing matrix elements between the nearly degenerate states. These matrix elements force energies apart.

To take advantage of the phases' distinct level statistics, we use a cold bath that has a small bandwidth W_b. According to Sec. III A, net positive work is extracted from the qubit engine because $\delta_{\text{MBL}} < \delta_{\text{GOE}}$. The mesoscale analog of δ_{GOE} is $\sim \langle \delta \rangle$, the typical gap ascended during hot thermalization. The engine must not emit energy on this scale during cold thermalization. Limiting W_b ensures that cold thermalization relaxes the engine only across gaps $\delta \leqslant W_b \ll \langle \delta \rangle$. Such anomalously small gaps appear more often in MBL energy spectra than in ETH spectra [69–71].

This level-statistics argument holds only within superselection sectors. Suppose, for example, that $H_{\text{meso}}(t)$ conserves the particle number. The level-statistics arguments apply only if the particle number remains constant throughout the cycle [[49], Appendix F]. Our numerical simulations (Sec. VI) take place at half-filling, in a subspace of dimensionality $\mathcal{N}$ of the order of magnitude of the whole space's dimensionality: $\mathcal{N} \sim \frac{2^N}{\sqrt{N}}$.

We are now ready to begin analyzing the mesoscopic-engine Otto cycle. The engine begins in the thermal state $\rho(0) = e^{-\beta_H H_{\text{GOE}}}/Z$, wherein $Z := \text{Tr}(e^{-\beta_H H_{\text{GOE}}})$. The engine can be regarded as starting each trial in some energy eigenstate j drawn according to the Gibbs distribution (Fig. 3). During stroke 1, $H_{\text{meso}}(t)$ is tuned from H_{GOE} to H_{MBL}. We approximate the tuning as quantum-adiabatic (diabatic corrections are modeled in Sec. V). Stroke 2, cold thermalization, depends on the gap δ'_j between the jth and $(j-1)$th MBL levels. δ'_j typically exceeds W_b. If it does, cold thermalization preserves the engine's energy, and the cycle outputs $W_{\text{tot}} = 0$. With probability $\sim \frac{W_b}{\langle \delta \rangle}$, the gap is small enough to thermalize: $\delta'_j < W_b$. In this case, cold thermalization drops the engine to level $j - 1$. Stroke 3 brings the engine to level $j - 1$ of H_{GOE}. The gap δ_j between the $(j-1)$th and jth H_{GOE} levels is $\langle \delta \rangle \gg W_b$, with the high probability $\sim 1 - (W_b/\langle \delta \rangle)^2$. Hence the engine likely outputs $W_{\text{tot}} > 0$. Hot thermalization (stroke 4) returns the engine to $\rho(0)$.

---

[2]$\langle \delta \rangle$ is defined as follows. The density of states at the energy E has the form $\mu(E) \approx \frac{\mathcal{N}}{\sqrt{2\pi N}\mathcal{E}} e^{-E^2/2N\mathcal{E}^2}$ (see Table I for the symbols' meanings). Inverting $\mu(E)$ yields the *local average gap*: $\langle \delta \rangle_E := \frac{1}{\mu(E)}$. Inverting the average of $\mu(E)$ yields the *average gap*,

$$\langle \delta \rangle := \frac{1}{\langle \mu(E) \rangle_{\text{energies}}} = \frac{\mathcal{N}}{\int_{-\infty}^{\infty} dE\, \mu^2(E)} = \frac{2\sqrt{\pi N}}{\mathcal{N}} \mathcal{E}. \quad (10)$$

[3]Imagine removing $Q(\alpha_t)$ from Eq. (9). One could increase α_t— could tune the Hamiltonian from ETH to MBL [43]—by strengthening a disorder potential. This strengthening would expand the energy band; tuning oppositely would compress the band. By expanding and compressing, in accordion fashion, and thermalizing, one could extract work. This engine would benefit little from properties of MBL, whose thermodynamic benefits we wish to highlight. Hence we "zero out" the accordion motion, by fixing $\langle \delta \rangle$ through $Q(\alpha_t)$. For a brief discussion of the accordionlike engine, see Appendix E 1.
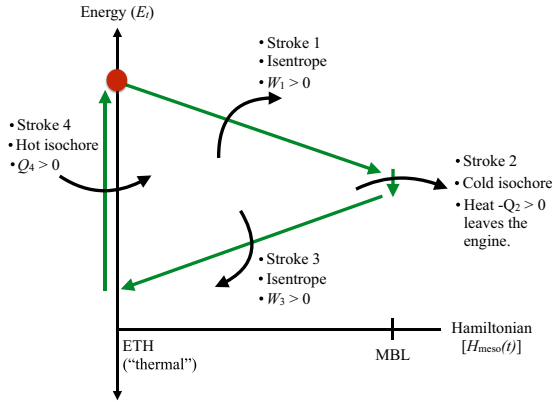
FIG. 3. Otto engine cycle for a mesoscale MBL system. Two energies in the many-body spectrum capture the cycle's basic physics. The engine can be regarded as beginning each trial in an energy eigenstate drawn from a Gibbs distribution. The red dot represents the engine's starting state in some trial of interest. During stroke 1, $H_{\text{meso}}(t)$ is tuned from "thermal" to MBL. During stroke 2, the engine thermalizes with a cold bath. $H_{\text{meso}}(t)$ returns from MBL to thermal during stroke 3. Stroke 4 resets the engine, which thermalizes with a hot bath. The tunings (strokes 1 and 3) map onto the thermodynamic Otto cycle's isentropes. The thermalizations (strokes 2 and 4) map onto isochores. The engine outputs work W_1 and W_3 during the tunings and absorbs heat Q_2 and Q_4 during thermalizations. MBL gap statistics' lack of level repulsion enhances the cycle; the engine "slides down" the lines that represent tunings, losing energy outputted as work.

C. Quantitative analysis of the mesoscale engine in the adiabatic limit

How well does the mesoscale Otto engine perform? We calculate the average work $\langle W_{\text{tot}} \rangle$ outputted per cycle and the efficiency η_{MBL}. Details appear in Appendix A.

We focus on the parameter regime in which the cold bath is very cold, the cold-bath bandwidth W_b is very small, and the hot bath is very hot: $T_C \ll W_b \ll \langle \delta \rangle$ and $\sqrt{N} \beta_H \mathcal{E} \ll 1$. The mesoscale engine resembles a qubit engine whose state and gaps are averaged over. The gaps, δ_j and δ'_j, obey the distributions $P_{\text{GOE}}^{(E)}(\delta_j)$ and $P_{\text{MBL}}^{(E)}(\delta'_j)$ [Eqs. (2) and (1)]. Correlations between the H_{GOE} and H_{MBL} spectra can be neglected.

We make three simplifying assumptions, generalizing later: (i) the engine is assumed to be tuned quantum-adiabatically. Diabatic corrections are estimated in Sec. V. (ii) The hot bath is at $T_H = \infty$. We neglect finite-temperature corrections, which scale as $N(\beta_H \mathcal{E})^2 \frac{(W_b)^2}{\langle \delta \rangle}$ and are calculated numerically in Appendix A. (iii) The gap distributions vary negligibly with energy: $P_{\text{GOE}}^{(E)}(\delta_j) \approx P_{\text{GOE}}(\delta_j)$, and $P_{\text{MBL}}^{(E)}(\delta'_j) \approx P_{\text{MBL}}(\delta'_j)$, wherein $\langle \delta \rangle_E \approx \langle \delta \rangle$.

1. Average work $\langle W_{\text{tot}} \rangle$ per cycle

The key is whether the cold bath relaxes the engine downwards across the MBL-side gap $\delta' \equiv \delta'_j$, distributed as $P_{\text{MBL}}(\delta')$, during a given trial. If $\delta' < W_b$, the engine has a probability $1/(1 + e^{-\beta_C \delta})$ of thermalizing. Hence the overall

probability of relaxation by the cold bath is

$$p_{\text{cold}} \approx \int_0^{W_b} d\delta' \, \frac{1}{\langle \delta \rangle} \frac{e^{-\delta'/\langle \delta \rangle}}{1 + e^{-\beta_C \delta'}} \approx \frac{1}{\langle \delta \rangle}(W_b - T_C \ln 2), \quad (11)$$

wherein we neglected $W_b/\langle \delta \rangle$ by setting $e^{-\delta'/\langle \delta \rangle} \approx 1$.

Alternatively, the cold bath could excite the engine to a level a distance δ' above the initial level. Such an upward hop occurs with a probability

$$\bar{p}_{\text{cold}} \approx \int_0^{W_b} d\delta' \, \frac{e^{-\delta'/\langle \delta \rangle}}{\langle \delta \rangle} \frac{e^{-\beta_C \delta'}}{1 + e^{-\beta_C \delta'}} \approx \frac{T_C \ln 2}{\langle \delta \rangle}. \quad (12)$$

If the engine relaxed downward during stroke 2, then, upon thermalizing with the hot bath during stroke 4, the engine gains heat $\langle Q \rangle_4 \approx \langle \delta \rangle$, on average. If the engine thermalized upward during stroke 2, then the engine loses $\langle \delta \rangle$ during stroke 4, on average. Therefore, the cycle outputs average work

$$\langle W_{\text{tot}} \rangle \approx (p_{\text{cold}} - \bar{p}_{\text{cold}})\langle \delta \rangle + \langle Q_2 \rangle \approx W_b - \frac{2 \ln 2}{\beta_C}. \quad (13)$$

$\langle Q_2 \rangle$ denotes the average heat absorbed by the engine during cold thermalization:

$$\langle Q_2 \rangle \approx -\int_0^{W_b} d\delta' \, \frac{\delta'}{\langle \delta \rangle} \frac{e^{-\delta'/\langle \delta \rangle}}{1 + e^{-\beta_C \delta'}} \approx -\frac{(W_b)^2}{2\langle \delta \rangle}, \quad (14)$$

which is $\ll \langle Q_4 \rangle$. This per-cycle power scales with the system size N as[4] $W_b \ll \langle \delta \rangle \sim \frac{\text{effective bandwidth}}{\text{\# energy eigenstates}} \sim \frac{\mathcal{E}\sqrt{N}}{\mathcal{N}}$.

2. Efficiency η_{MBL}

The efficiency is

$$\eta_{\text{MBL}} = \frac{\langle W_{\text{tot}} \rangle}{\langle Q_4 \rangle} = \frac{\langle Q_4 \rangle + \langle Q_2 \rangle}{\langle Q_4 \rangle} \approx 1 - \frac{W_b}{2\langle \delta \rangle}. \quad (15)$$

The imperfection is small, $\frac{W_b}{2\langle \delta \rangle} \ll 1$, because the cold bath has a small bandwidth. This result mirrors the qubit-engine efficiency η_{qubit}.[5] But our engine is a many-body system of N interacting sites. MBL will allow us to employ segments of the system as independent qubitlike subengines, despite interactions. In the absence of MBL, each subengine's effective $\langle \delta \rangle = 0$. With $\langle \delta \rangle$ vanishes the ability to extract $\langle W_{\text{tot}} \rangle > 0$. Whereas the efficiency is nearly perfect, an effective engine requires also a finite power. The MBL engine's power will depend on dynamics, as discussed below.

---

[4]The *effective bandwidth* is defined as follows. The many-body system has a Gaussian density of states: $\mu(E) \approx \frac{\mathcal{N}}{\sqrt{2\pi N}\mathcal{E}} e^{-E^2/2N\mathcal{E}^2}$. The states within a standard deviation $\mathcal{E}\sqrt{N}$ of the mean obey Eqs. (1) and (2). These states form the effective band, whose width scales as $\mathcal{E}\sqrt{N}$.

[5]η_{MBL} is comparable also to η_{QHO} [Eq. (5)]. Imagine operating an ensemble of independent QHO engines. Let the jth QHO frequency be tuned between Ω_j and ω_j, distributed according to $P_{\text{GOE}}(\Omega_j)$ and $P_{\text{MBL}}(\omega_j)$. The average MBL-like gap ω_j, conditioned on $\omega_j \in [0, W_b]$, is $\langle \omega_j \rangle \sim \frac{1}{W_b/\langle \delta \rangle} \int_0^{W_b} d\omega_j \, \omega_j \, P_{\text{MBL}}(\omega_j) \approx \frac{1}{W_b} \int_0^{W_b} d\omega_j \, \omega_j = \frac{W_b}{2}$. Averaging the efficiency over the QHO ensemble yields $\langle \eta_{\text{QHO}} \rangle := 1 - \frac{\langle \omega \rangle}{\langle \Omega \rangle} \approx 1 - \frac{W_b}{2\langle \delta \rangle} \approx \eta_{\text{MBL}}$. The mesoscale MBL engine operates at the ideal average efficiency of an ensemble of QHO engines. However, MBL enables qubitlike engines to pack together densely in a large composite engine.

IV. MBL ENGINE IN THE THERMODYNAMIC LIMIT

The MBL engine's advantage lies in having a simple thermodynamic limit that does not compromise efficiency or power output. A nonlocalized Otto engine would suffer from a suppression of the average level spacing: $\langle \delta \rangle \sim \frac{\mathcal{E}\sqrt{N}}{2^N}$, which suppresses the average output per cycle $\langle W_{\text{tot}} \rangle \sim W_{\text{b}} \ll \langle \delta \rangle$ exponentially in the system size. Additionally, the tuning speed v must shrink exponentially: $H_{\text{meso}}(t)$ is ideally tuned quantum-adiabatically. The time per tuning stroke must far exceed $\langle \delta \rangle^{-1}$. The mesoscale engine scales poorly, but properties of MBL offer a solution.

A thermodynamically large MBL Otto engine consists of mesoscale subengines that operate mostly independently. This independence hinges on *local level correlations* of the MBL phase [45,47,48]; subsystems separated by a distance L evolve roughly independently until times exponential in L, due to the localization [15].

Particularly important is the scaling of the typical strength of a local operator in an MBL phase. Let O denote a generic local operator that has support on just a size-L region. O can connect only energy eigenstates $|\psi_1\rangle$ and $|\psi_2\rangle$ that differ just in their local integrals of motion in that region. Such states are said to be "close together," or "a distance L apart." Let ξ denote the system's localization length. If the eigenfunctions lie far apart ($L \gtrsim \xi$), the matrix-element size scales as

$$|O_{21}| \sim 2^{-L} e^{-L/\xi}. \tag{16}$$

(All lengths appear in units of the lattice spacing, set to one.) This scaling determines the typical level spacing, since such matrix elements give rise to level repulsion:

$$\delta \sim \mathcal{E} 2^{-L} e^{-L/\xi} \tag{17}$$

(possibly to within a power-law correction). The localization-induced exponential suppresses long-distance communication (see [15,40,45,72] and Appendix B).

Let us apply this principle to a chain of N-site mesoscale engines separated by N-site buffers. The engine is cycled between a shallowly localized (H_{GOE}-like) Hamiltonian, which has a localization length $\xi_>$, and a deeply localized (H_{MBL}-like) Hamiltonian, which has $\xi_< \ll \xi_>$.

The key element in the construction is that the cold bath acts through local operators confined to $< N \sim \xi_>$ sites. This defines the subengines of the thermodynamic MBL Otto engine. Localization guarantees that "what happens in a subengine stays in a subengine": subengines do not interfere much with each other's operation.

This subdivision boosts the engine's power. A length-N mesoscale engine operates at the average per-cycle power $\langle W_{\text{tot}} \rangle_{\text{meso}} \sim W_{\text{b}} \ll \frac{\mathcal{E}\sqrt{N}}{2^N}$ (Sec. III C). A subdivided length-N_{macro} MBL engine outputs average work $\sim \frac{N_{\text{macro}}}{2N} \langle W_{\text{tot}} \rangle_{\text{meso}}$. In contrast, if the length-N_{macro} engine were not subdivided, it would output average work $\sim \frac{\mathcal{E}\sqrt{N_{\text{macro}}}}{2^{N_{\text{macro}}}}$, which vanishes in the thermodynamic limit.

V. TIMESCALE RESTRICTIONS ON THE MBL OTTO ENGINE'S OPERATION

We estimate the restrictions on the speed with which the Hamiltonian must be tuned to avoid undesirable
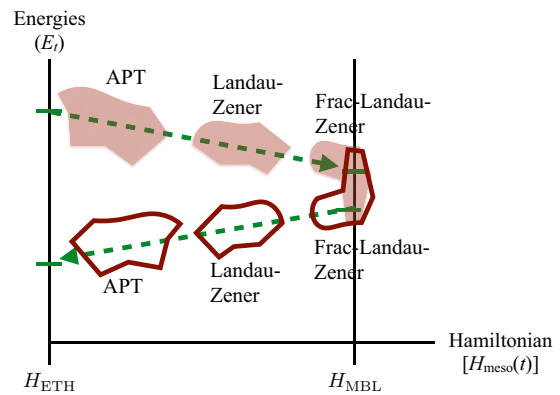

FIG. 4. Three classes of diabatic transitions. Hops to arbitrary energy levels, modeled with general adiabatic perturbation theory (APT), plague the ETH regime. Landau-Zener transitions and fractional-Landau-Zener transitions plague the many-body-localized regime.

diabatic transitions and intersubengine communication. Most importantly, we estimate the time required for cold thermalization (stroke 2).

A. Diabatic corrections

We have modeled the Hamiltonian tuning as quantum-adiabatic, but realistic tuning speeds $v := \mathcal{E}|\frac{d\alpha_t}{dt}|$ are finite. To understand diabatic tuning's effects, we distinguish the time-t density matrix $\rho(t)$ from the corresponding diagonal ensemble,

$$\rho_{\text{diag}}(t) = \sum_j |E_j(t)\rangle \varepsilon_j(t)\langle E_j(t)|, \quad \text{wherein}$$
$$\varepsilon_j(t) = \langle E_j(t)|\rho|E_j(t)\rangle \tag{18}$$

and $|E_j(t)\rangle$ is an instantaneous energy eigenbasis of $H_{\text{meso}}(t) = \sum_j |E_j(t)\rangle E_j(t)\langle E_j(t)|$. The average energy depends on $\rho(t)$ only through $\rho_{\text{diag}}(t)$. (More generally, the state's off-diagonal elements dephase under the dynamics. $\rho_{\text{diag}}(t)$ is "slow" and captures most of the relevant physics [46].)

In the adiabatic limit, $\varepsilon_j(t) = \varepsilon_j(0)$. We seek to understand how this statement breaks down when the tuning proceeds at a finite speed v. It is useful to think of "infinite-temperature thermalization" in the sense of this diagonal ensemble: fast tuning may push the diagonal-ensemble weights $\varepsilon_j(t)$ towards uniformity—even though the process is unitary and the entropy $S = -\rho(t)\ln \rho(t)$ remains constant—thanks to the off-diagonal elements.

The effects of diabatic tuning appear in three distinct regimes, which we label "fractional-Landau-Zener," "Landau-Zener," and "APT" (Fig. 4). We estimate the average per-cycle work costs $\langle W_{\text{diab}} \rangle$ of diabatic jumps, guided by the numerics in Sec. VI. We focus on $T_{\text{H}} = \infty$ and $T_{\text{C}} = 0$, for simplicity. Since $T_{\text{H}} = \infty$, diabatic hops cannot bring $\rho_{\text{diag}}(t)$ closer to $\mathbb{1}/2^N$—cannot change the average energy—during stroke 1. Hence we focus on stroke 3.

1. Fractional-Landau-Zener transitions

At the beginning of stroke 3, nonequilbrium effects could excite the system back across the small gap to energy level

j. The transition would cost work and would prevent the trial from outputting $W_{\text{tot}} > 0$. We dub this excitation a *fractional-Landau-Zener (frac-LZ) transition*. It could be suppressed by a sufficiently slow drive [64]. The effects, and the resultant bound on v, are simple to derive.

Let the gap start stroke 3 at size δ and grow to a size $\Delta > \delta$. Because the two energy levels begin close together, one cannot straightforwardly apply the Landau-Zener formula. One must use the fractional-Landau-Zener result of De Grandi and Polkovnikov [64],

$$p_{\text{frac-LZ}}(\delta) \approx \frac{v^2(\delta_-)^2}{16}\left(\frac{1}{\delta^6} + \frac{1}{\Delta^6}\right) \approx \frac{v^2(\delta_-)^2}{16\delta^6}. \quad (19)$$

δ_- denotes the MBL level-repulsion scale, the characteristic matrix element introduced by a perturbation between eigenstates of an unperturbed Hamiltonian. We suppose that energy-level pairs with $p_{\text{frac-LZ}} \lesssim 1$ are returned to the infinite-temperature state from which the cold bath disturbed them. These pairs do not contribute to $\langle W_{\text{tot}}\rangle$. Pairs that contribute have $p_{\text{frac-LZ}} < 1$, i.e.,

$$\delta > (v\delta_-)^{1/3}. \quad (20)$$

If the rest of the stroke is adiabatic, the average work performed during the cycle is

$$\langle W_{\text{tot}}\rangle \sim \langle Q_4\rangle - \langle Q_2\rangle - (v\delta_-)^{1/3}, \quad (21)$$

which results immediately in the correction

$$\langle W_{\text{diab,frac-LZ}}\rangle \sim (v\delta_-)^{1/3}. \quad (22)$$

This correction is negligible at speeds low enough that

$$v \ll \frac{(W_{\text{b}})^3}{\delta_-}. \quad (23)$$

2. Landau-Zener transitions

While the system is localized, the disturbances induced by the tuning $\frac{dH(t)}{dt}$ can propagate only a short distance l_v. The tuning effectively reduces the mesoscale engine to a length-l_v subengine. To estimate l_v, we compare the minimum gap of a length-l_v subsystem to the speed v:

$$\mathcal{E}2^{-l_v}e^{-l_v/\xi_<} \sim \sqrt{v}. \quad (24)$$

The left-hand side comes from Eq. (17). This minimum gap—the closest that two levels are likely to approach—is given by the smallest level-repulsion scale, δ_-. δ_- characterizes the deeply localized system, whose $\xi = \xi_<$. Consequently,

$$l_v \sim \frac{\ln(\mathcal{E}^2/v)}{2\left(\ln 2 + \frac{1}{\xi_<}\right)}. \quad (25)$$

Suppose that $l_v \leqslant N$, and consider a length-l_v effective subengine. In the adiabatic limit, $\langle W_{\text{tot}}\rangle$ does not depend on the engine's size. ($\langle W_{\text{tot}}\rangle$ depends only on the bath bandwidth $W_{\text{b}} \ll \langle\delta\rangle$.) To estimate how a finite v changes $\langle W_{\text{tot}}\rangle$, we consider the gaps $\delta < W_{\text{b}}$ of the size-l_v subengine. We divide the gaps into two classes.

(1) Gaps connected by flipping l-bits on a region of diameter $l < l_v$. The tuning is adiabatic with respect to these gaps, so they result in work output.

(2) Gaps connected by flipping l-bits on a region of diameter $l = l_v$. The tuning is resonant with these gaps and so thermalizes them, in the sense of the diagonal ensemble [Eq. (18)]: the tuning makes the instantaneous-energy-eigenvector weights ε_j uniform, on average.

Type-1 gaps form a v-independent $O(1)$ fraction θ of the length-l_v subengine's short-length-scale gaps.[6] Type-2 gaps therefore make up a fraction $1 - \theta$. Hence Landau-Zener physics leads to a v-independent $O(1)$ diabatic correction $(1 - \theta)W_{\text{b}}$ to $\langle W_{\text{tot}}\rangle$, provided that v is high enough that $l_v < N$.

3. Adiabatic-perturbation-theory (APT) transitions

When the system is in the ETH phase (or has correlation length $\xi \sim N$), typical minimum gaps (points of closest approach) are still given by the level-repulsion scale, which is now $\langle\delta\rangle$. Hence one expects the tuning to be adiabatic if

$$v \ll \langle\delta\rangle^2. \quad (26)$$

This criterion could be as stringent (depending on the system size and localization lengths) as the requirement (23) that fractional Landau-Zener transitions occur rarely. The numerics in Sec. VI C indicate that fractional-Landau-Zener transitions limit the power more than APT transitions do.

Both fractional Landau-Zener transitions and APT transitions bound the cycle time τ_{cycle} less stringently than thermalization with the cold bath; hence a more detailed analysis of APT transitions would be gratuitous. Such an analysis would rely on the general adiabatic perturbation theory of De Grandi and Polkovnikov [64]; hence the moniker "APT transitions."

B. Precluding communication between subengines

To maintain the MBL engine's advantage, we must approximately isolate subengines. The subengines' (near) independence implies a lower bound on the tuning speed v: the price paid for scalability is the impossibility of adiabaticity. Suppose that $H_{\text{macro}}(t)$ were tuned infinitely slowly. Information would have time to propagate from one subengine to every other. The slow spread of information through MBL [73] lower-bounds v. This consideration, however, does not turn out to be the most restrictive constraint on the cycle time. Therefore we address it only qualitatively.

As explained in Sec. V A 2, v determines the effective size of an MBL subengine. Ideally, v is large enough to prevent adiabatic transitions between configurations extended beyond the mesoscale N. For each stage of the engine's operation, v

---

[6]We can estimate θ crudely. For a given diameter-l_v subset, each gap connected by a diameter-$(l_v - 1)$ operator can be made into a diameter-l_v gap: one flips the last (l_v)th l-bit. Adding a qubit to the system doubles the dimensionality of the system's Hilbert space. The number of levels doubles, so the number of gaps approximately doubles, so $\theta \approx 1/2$. This estimate neglects several combinatorial matters. A more detailed analysis would account for the two different diameter-$(l_v - 1)$ regions of a given length-l_v subengine, gaps connected by l-bit flips in the intersections of those subengines, the number of possible diameter-l_v subengines of an N-site system, etc.

should exceed the speed given in Eq. (24) for the localization length ξ of a length-$(N + 1)$ chain:

$$v \gg [\delta_-(N + 1, \xi)]^2 \sim \mathcal{E}^2 2^{-2(N+1)} e^{-2(N+1)/\xi}. \qquad (27)$$

We have made explicit the dependence of the level-repulsion scale δ_- on the mesoscale-engine size N and on the localization length ξ. During stroke 1, ξ drops, so the RHS of (27) decays quickly. Hence the speed should interpolate between $[\delta_-(N + 1, \xi_>)]^2$ and $\frac{(W_b)^3}{\delta_-(N,\xi_<)}$ [from Eq. (23)].

C. Lower bound on the cycle time τ_{cycle} from cold thermalization

Thermalization with the cold bath (stroke 2) bounds τ_{cycle} more stringently than the Hamiltonian tunings do. The reasons are (i) the slowness with which MBL thermalizes and (ii) the restriction $W_b \ll \langle \delta \rangle$ on the cold-bath bandwidth. We elaborate after introducing our cold-thermalization model (see Ref. [[49], Appendix I] for details).

We envision the cold bath as a bosonic system that couples to the engine locally, as via the Hamiltonian

$$H_{\text{int}} = g \int_{-W_b/\xi_>}^{W_b/\xi_>} d\omega \sum_{j \in \text{subengine}} (c_j^\dagger c_{j+1} + \text{H.c.})(b_\omega + b_\omega^\dagger)$$

$$\times \delta(\langle 0 | c_j H_{\text{macro}}(\tau) c_{j+1}^\dagger | 0 \rangle - \omega). \qquad (28)$$

The sum runs over the sites in the subengines, excluding the sites in the buffers between subengines. The coupling strength is denoted by g. We have switched from spin notation to fermion notation via a Jordan-Wigner transformation. c_j and $c_j^\dagger$ denote the annihilation and creation of a fermion at site j. $H_{\text{macro}}(t)$ denotes the Hamiltonian that would govern the engine at time t in the bath's absence. Cold thermalization lasts from $t = \tau$ to $t = \tau'$ (Fig. 2). b_ω and $b_\omega^\dagger$ represent the annihilation and creation of a frequency-ω boson in the bath. The Dirac delta function is denoted by $\delta(.)$.

The bath couples locally, e.g., to pairs of nearest-neighbor spins. This locality prevents subengines from interacting with each other much through the bath. The bath can, e.g., flip spin j upward while flipping spin $j + 1$ downward. These flips likely change a subengine's energy by an amount E. The bath can effectively absorb only energy quanta of size $\leqslant W_b$ from any subengine. The cap is set by the bath's speed of sound [74], which follows from microscopic parameters in the bath's Hamiltonian [75]. The rest of the energy emitted during the spin flips, $|E - W_b|$, is distributed across the subengine as the intrinsic subengine Hamiltonian flips more spins.

Let τ_{th} denote the time required for stroke 2. We estimate τ_{th} from Fermi's Golden Rule,

$$\Gamma_{fi} = \frac{2\pi}{\hbar} |\langle f | V | i \rangle|^2 \mu_{\text{bath}}. \qquad (29)$$

Cold thermalization transitions the engine from an energy level $|i\rangle$ to a level $|f\rangle$. The bath has a density of states $\mu_{\text{bath}} \sim 1/W_b$. V denotes the operator, defined on the engine's Hilbert space, induced by the coupling to the bath.

We estimate the matrix-element size $|\langle f | V | i \rangle|$ as follows. Cold thermalization transfers energy $E_{if} \sim W_b$ from the subengine to the bath. W_b is very small. Hence the energy change rearranges particles across a large distance $L \gg \xi = \xi_<$, due to local level

correlations (17). V nontrivially transforms just a few subengine sites. Such a local operator rearranges particles across a large distance L at a rate that scales as (17), $\mathcal{E} e^{-L/\xi} 2^{-L} \sim \delta_-$. Whereas $\mathcal{E}$ sets the scale of the level repulsion δ_-, g sets the scale of $|\langle f | V | i \rangle|$. The correlation length $\xi = \xi_<$ during cold thermalization. We approximate L with the subengine length $\xi_>$. Hence $|\langle f | V | i \rangle| \sim \frac{g\delta_-}{\mathcal{E}}$.

We substitute into Eq. (29). The transition rate $\Gamma_{fi} = \frac{1}{\tau_{\text{th}}}$. Inverting yields

$$\tau_{\text{cycle}} \sim \tau_{\text{th}} \sim W_b \left(\frac{\mathcal{E}}{g\delta_-} \right)^2. \qquad (30)$$

To bound τ_{cycle}, we must bound the coupling g. The interaction is assumed to be Markovian: information leaked from the engine dissipates throughout the bath quickly. Bath correlation functions must decay much more quickly than the coupling transfers energy. If τ_{bath} denotes the correlation-decay time, $\tau_{\text{bath}} < \frac{1}{g}$. The small-bandwidth bath's $\tau_{\text{bath}} \sim 1/W_b$, so $g < W_b$. This inequality, with Eq. (30), implies

$$\tau_{\text{cycle}} = \tau_{\text{th}} > \frac{\mathcal{E}^2}{W_b(\delta_-)^2} \sim \frac{10}{\mathcal{E}} e^{2\xi_>/\xi_<} 2^{3\xi_>}. \qquad (31)$$

The final expression follows if $W_b \sim \frac{\langle \delta \rangle}{10}$.

Like Markovianity, higher-order processes bound τ_{th}. Such processes transfer energy $E > W_b$ between the engine and the cold bath. These transfers must be suppressed. g^a, wherein $a > 1$, determine the rates at which these processes occur. The resulting bound on τ_{th} is less stringent than Eq. (31) (Appendix C).

VI. NUMERICAL SIMULATIONS

We use numerical exact diagonalization to check our analytical results. In Sec. VI A, we describe the Hamiltonian used in our numerics. In Sec. VI B, we study engine performance in the adiabatic limit (addressed analytically in Sec. III C). In Sec. VI C, we study diabatic corrections (addressed analytically in Sec. V A). We numerically study the preclusion of communication between mesoscale subengines (addressed analytically in Sec. V B) only insofar as these results follow from diabatic corrections: Limitations on computational power restricted the system size to 12 sites. Details about the simulation appear in Appendix D. Our code is in Ref. [76].

A. Hamiltonian

The engine can be implemented with a disordered Heisenberg model. A similar model's MBL phase has been realized with ultracold atoms [3]. We numerically simulated a 1D mesoscale chain governed by a Hamiltonian

$$H_{\text{sim}}(t) = \frac{\mathcal{E}}{Q(h(\alpha_t))} \left[\sum_{j=1}^{N-1} \boldsymbol{\sigma}_j \cdot \boldsymbol{\sigma}_{j+1} + h(\alpha_t) \sum_{j=1}^{N} h_j \sigma_j^z \right]; \qquad (32)$$

this is a special case of the general mesoscopic Hamiltonian (9) described in Sec. III B. Equation (32) describes spins equivalent to interacting spinless fermions. Energies are expressed in units of $\mathcal{E}$, the average per-site energy density.

For $\gamma = x, y, z$, the γth Pauli operator that operates nontrivially on the jth site is denoted by σ_j^γ. The Heisenberg interaction $\boldsymbol{\sigma}_j \cdot \boldsymbol{\sigma}_{j+1}$ encodes nearest-neighbor hopping and repulsion.

The tuning parameter $\alpha_t \in [0, 1]$ determines the phase occupied by $H_{\text{sim}}(t)$. The site-j disorder potential depends on a random variable h_j distributed uniformly across $[-1, 1]$. The disorder strength $h(\alpha_t)$ varies as $h(\alpha_t) = \alpha_t h_{\text{GOE}} + (1 - \alpha_t)h_{\text{MBL}}$. When $\alpha_t = 0$, the disorder is weak, $h = h_{\text{GOE}}$, and the engine occupies the ETH phase. When $\alpha_t = 1$, the disorder is strong, $h = h_{\text{MBL}} \gg h_{\text{GOE}}$, and the engine occupies the MBL phase.

The normalization factor $Q(h(\alpha_t))$ preserves the width of the density of states (DOS) and so preserves $\langle \delta \rangle$. $Q(h(\alpha_t))$ prevents the work extractable via change of bandwidth from polluting the work extracted with help from level statistics (see Appendix E 1 for a discussion of work extraction from bandwidth change). $Q(h(\alpha_t))$ is defined and calculated in Appendix D 1.

The ETH-side field had a magnitude $h(0) = 2.0$, and the MBL-side field had a magnitude $h(1) = 20.0$. These $h(\alpha_t)$ values fall squarely on opposite sides of the MBL transition at $h \approx 7$.

B. Adiabatic engine

We compare the analytical predictions of of Sec. III C and Appendix A to numerical simulations of a 12-site engine governed by the Hamiltonian (32). During strokes 1 and 3, the state was evolved as though the Hamiltonian were tuned adiabatically. We index the energies $E_j(t)$ from least to greatest at each instant: $E_j(t) < E_k(t) \, \forall j < k$. Let ρ_j denote the state's weight on eigenstate j of the initial Hamiltonian, whose $\alpha_t = 0$. The engine ends stroke 1 with weight ρ_j on eigenstate j of the post-tuning Hamiltonian, whose $\alpha_t = 1$.

The main results appear in Fig. 5. Figure 5(a) shows the average work extracted per cycle, $\langle W_{\text{tot}} \rangle$. Figure 5(b) shows the efficiency, η_{MBL}.

In these simulations, the baths had the extreme temperatures $T_{\text{H}} = \infty$ and $T_{\text{C}} = 0$. This limiting case elucidates the W_{b}-dependence of $\langle W_{\text{tot}} \rangle$ and of η_{MBL}: disregarding finite-temperature corrections, on a first pass, builds intuition. Finite-temperature numerics appear alongside finite-temperature analytical calculations in Appendix A.

Figure 5 shows how the per-cycle power and the efficiency depend on the cold-bath bandwidth W_{b}. As expected, $\langle W_{\text{tot}} \rangle \approx W_{\text{b}}$. The dependence's linearity, and the unit proportionality factor, agree with Eq. (13). Also as expected, the efficiency declines as the cold-bath bandwidth rises: $\eta_{\text{MBL}} \approx 1 - \frac{W_{\text{b}}}{2\langle \delta \rangle}$. The linear dependence and the proportionality factor agree with Eq. (15).

The gray columns in Fig. 5 highlight the regime in which the analytics were performed, where $\frac{W_{\text{b}}}{\langle \delta \rangle} \ll 1$. If the cold-bath bandwidth is small, $W_{\text{b}} < \langle \delta \rangle$, the analytics-numerics agreement is close. But the numerics agree with the analytics even outside this regime. If $W_{\text{b}} \gtrsim \langle \delta \rangle$, the analytics slightly underestimate η_{MBL}: the simulated engine operates more efficiently than predicted. To predict the numerics' overachievement, one would calculate higher-order corrections in Appendix A; one would Taylor-approximate to higher pow-
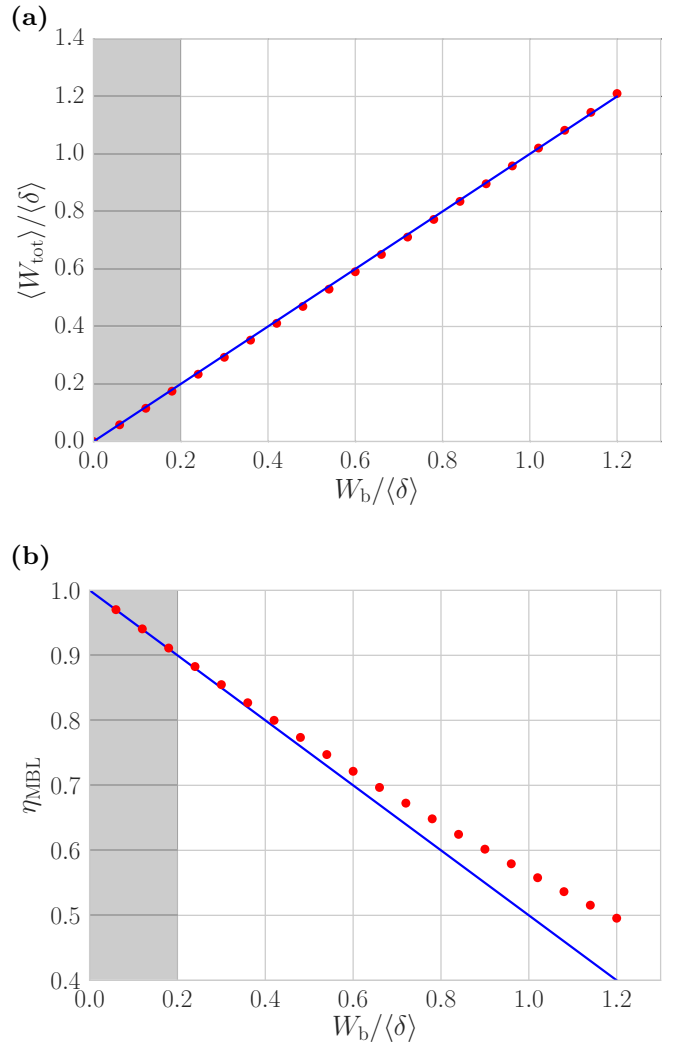

FIG. 5. Average per-cycle power $\langle W_{\text{tot}} \rangle$ (top) and efficiency η_{MBL} (bottom) as functions of the cold-bath bandwidth W_{b}. Each red dot represents an average over 1000 disorder realizations of the random-field Heisenberg Hamiltonian (32). The slanted blue lines represent the analytical predictions (13) and (15) of Sec. III C. When $W_{\text{b}} \ll \langle \delta \rangle$ (in the gray shaded region), the engine operates in the regime of interest. Here, $\langle W_{\text{tot}} \rangle$ and η_{MBL} vary linearly with W_{b}, as predicted. The error bars are smaller than the numerical-data points.

ers, modeling subleading physical processes. Such processes include the engine's dropping across a chain of three small gaps, $\delta_1', \delta_2', \delta_3' < W_{\text{b}}$, during cold thermalization.

The error bars are smaller than the numerical-data points. Each error bar represents the error in the estimate of a mean (of $\langle W_{\text{tot}} \rangle$ or of $\eta_{\text{MBL}} := 1 - \frac{\langle W_{\text{tot}} \rangle}{\langle Q_{\text{in}} \rangle}$) over 1000 disorder realizations. Each error bar extends a distance (sample standard deviation)/$\sqrt{\#\text{ realizations}}$ above and below that mean.

C. Diabatic engine

We then simulated strokes 1 and 3 as though $H_{\text{sim}}(t)$ were tuned at finite speed v. Computational limitations restricted the engine to 8 sites. (That our upper bounds on v scale as powers of $\langle \delta \rangle \sim 2^{-N}$ implies that these simulations quickly
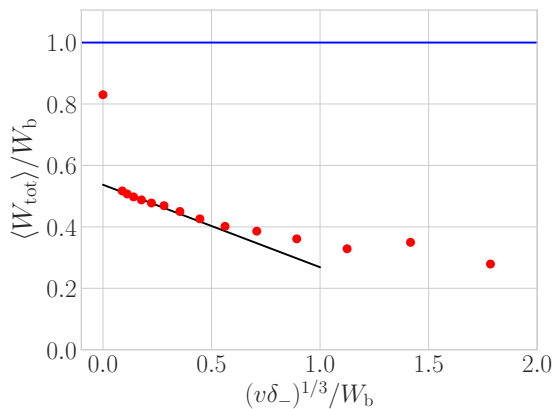
FIG. 6. Average per-cycle work as a function of tuning speed. We numerically simulated 995 disorder realizations of the random-field Heisenberg Hamiltonian (32) for a system of $N = 8$ sites (red dots). The results are compared to the analytical estimate (13) for the adiabatic work output (blue line) and an empirical straight-line fit $W_{tot} = W_0 - W_1 (v \delta_-)^{1/3} / W_b$ (black line). Errors in the estimate of the mean, computed as (sample standard deviation)/$\sqrt{(\# \text{ realizations})}$, lead to error bars smaller than the numerical-data points.

become slow to run.) We simulate a stepwise tuning, taking

$$\alpha_t = \frac{\delta t \lfloor t/\delta t \rfloor}{T} . \tag{33}$$

δt denotes a time-step size, and $T \propto \frac{h_{MBL} - h_{GOE}}{v}$ denotes the time for which one tuning stroke lasts. This protocol is more violent than the protocols treated analytically: v is assumed to remain finite in the diabatic analytics. In the numerics, we tune by sudden jumps (for reasons of numerical convenience). We work at $T_H = \infty$ and $T_C = 0$—again, to capture the essential physics without the complication of finite-temperature corrections.

Figure 6 shows the average work output, $\langle W_{tot} \rangle$, as a function of v. Despite the simulated protocol's violence, both a fractional-Landau-Zener correction $W_{frac-LZ} \sim (v \delta_-)^3$, explained in Sec. V A 1, and a v-independent $O(1)$ Landau-Zener correction, explained in Sec. V A 2, are visible. We believe that the adiabatic numerics ($v = 0$ red dot) differ from the analytics (blue line) due to finite-size effects: for small systems away from the spectrum's center, the average gap estimated from the density of states can vary appreciably over one gap. These numerics confirm the analytics and signal the MBL Otto engine's robustness with respect to changes in the tuning protocol.

VII. ORDER-OF-MAGNITUDE ESTIMATES

How well does the localized engine perform? We estimate the engine's power and power density, in addition to comparing the engine with three competitors.

A. Localized engine

Localization has been achieved in solid-state systems.[7] Consider silicon doped with phosphorus [44]. A distance of

~10 nm may separate phosphorus impurities. Let our engine cycle's shallowly localized regime have a localization length of $\xi_> \sim 10$ sites, or 100 nm. The work-outputting degrees of freedom will be electronic. The localized states will correspond to energies $\mathcal{E} \sim 1$ eV. Each subengine's half-filling Hilbert space has dimensionality $\mathcal{N} = \binom{10}{5} \sim 10^2$. Hence each subengine has an effective average gap $\langle \delta \rangle \sim \frac{\mathcal{E} \sqrt{N}}{\mathcal{N}} \sim \frac{1 \text{ eV}}{10^2} \sim 10$ meV. The cold-bath bandwidth must satisfy $\langle \delta \rangle \gg W_b$. We set W_b to be an order of magnitude down from $\langle \delta \rangle$: $W_b \sim 1$ meV ~ 10 K. The cold-bath bandwidth approximates the work outputted by one subengine per cycle:[8] $\langle W_{tot} \rangle \sim W_b \sim 1$ meV [Eq. (13)].

What volume does a localized subengine fill? Suppose that the engine is three-dimensional (3D).[9] A little room should separate the subengines. Classical-control equipment requires more room. Also, the subengine needs space to connect to the baths. We therefore associate each subengine with a volume of $V \approx (100 \text{ nm})^3$.

The last element needed is the cycle time, τ_{cycle}. We choose for δ_- to be a little smaller than W_b—of the same order: $\delta_- \sim W_b \sim 1$ meV. In the extreme case allowed by Eq. (31), $\tau_{cycle} \sim \frac{\hbar \mathcal{E}^2}{W_b (\delta_-)^2} \sim \frac{\hbar \mathcal{E}^2}{(W_b)^3} \sim \frac{(10^{-15} \text{ eV s})(1 \text{ eV})^2}{(1 \text{ meV})^3} \sim 1 \mu s$.

The localized engine therefore operates with a power $\mathscr{P} \sim \frac{W_b}{\tau_{cycle}} \sim \frac{1 \text{ meV}}{1 \mu s} \approx 10^{-16}$ W. Interestingly, this $\mathscr{P}$ is one order of magnitude greater than a flagellar motor's [77] power, according to our estimates.

We can assess the engine by calculating not only its power, but also its power density. The localized engine packs a punch at $\frac{\mathscr{P}}{V} \sim \frac{10^{-16} \text{ W}}{(10^{-7} \text{ m})^3} = 100$ kW/m^3.

B. Car engine

The quintessential Otto engine powers cars. A typical car engine outputs $\mathscr{P} \sim 100$ horsepower ~ 100 kW . A car's power density is $\frac{\mathscr{P}}{V} \sim \frac{100 \text{ kW}}{100 \text{ L}} = 1$ MW/ m^3 (wherein L represents liters). The car engine's $\frac{\mathscr{P}}{V}$ exceeds the MBL engine's by only an order of magnitude, according to these rough estimates.

C. Array of quantum dots

MBL has been modeled with quasilocal bits [1,78]. A string of ideally independent bits or qubits, such as quantum dots, forms a natural competitor. Each quantum dot would form a qubit Otto engine whose gap is shrunk, widened, and shrunk [79–83].

---

[7]This localization is single-particle, or Anderson [72], rather than many-body. Appendix E 4 extends the MBL Otto engine to an Anderson-localized Otto engine.

[8]The use of semiconductors would require corrections to our results. (Dipolar interactions would couple the impurities' spins. Energy eigenfunctions would decay as power laws with distance.) However, we aim for just a rough estimate.

[9]Until now, we have supposed that the engine is 1D. Anderson localization, which has been realized in semiconductors, exists in all dimensionalities. Yet whether MBL exists in dimensionalities $D > 1$ remains an open question. Some evidence suggests that MBL exists in $D \geqslant 2$ [6,8,10]. However, attributing a 3D volume to the engine facilitates comparisons with competitors. We imagine 10-nm-long 1D strings of sites. Strings are arrayed in a plane, separated by 10 nm. Planes are stacked atop each other, separated by another 10 nm.

A realization could consist of double quantum dots [84,85]. The scales in Refs. [84,85] suggest that a quantum-dot engine could output an amount $W_{\text{tot}} \sim 10$ meV of work per cycle per dot. We approximate the cycle time τ_{cycle} with the spin relaxation time: $\tau_{\text{cycle}} \sim 1$ μs. (The energy eigenbasis need not rotate, unlike for the MBL engine. Hence diabatic hops do not lower-bound the ideal-quantum-dot τ_{cycle}.) The power would be $\mathscr{P} \sim \frac{W_{\text{tot}}}{\tau_{\text{cycle}}} \sim \frac{10 \text{ meV}}{1 \text{ }\mu s} \sim 10^{-15}$ W. The quantum-dot engine's power exceeds the MBL engine's by an order of magnitude.

However, the quantum dots must be separated widely. Otherwise, they will interact, as an ETH system. (See Ref. [62] for disadvantages of interactions in another quantum thermal machine. Spin-spin couplings cause "quantum friction," limiting the temperatures to which a refrigerator can cool.) We compensate by attributing a volume $V \sim (1 \text{ }\mu\text{m})^3$ to each dot. The power density becomes $\frac{\mathscr{P}}{V} \sim 1$ kW/m^3, two orders of magnitude less than the localized engine's. Localization naturally implies near independence of the subengines.

In Appendix E, we compare the MBL Otto engine to four competitors: a bandwidth engine, a variant of the MBL engine that is tuned between two disorder strengths, an engine of quantum dots (analyzed partially above), and an Anderson-localized engine. We argue that the MBL Otto engine is more robust against perturbations than the bandwidth, Anderson, and quantum-dot engines. We also argue that our MBL engine is more reliable than the equal-disorder-strength engine: Our MBL engine's W_{tot} varies less from trial to trial and suppresses worst-case trials, in which $W_{\text{tot}} < 0$. This paper's arguments go through almost unchanged for an Anderson-localized medium. Such a medium would lack robustness against interactions, though; even if the interactions do not delocalize the medium—which would destroy the engine—they would turn the Anderson engine into an MBL engine. One can view our MBL engine as an easy generalization of the Anderson engine.

VIII. OUTLOOK

The realization of thermodynamic cycles with quantum many-body systems was proposed very recently [35,37,38,86–90]. MBL offers a natural platform, due to its "athermality" and to athermality's resourcefulness in thermodynamics. We designed an Otto engine that benefits from the discrepancy between many-body-localized and "thermal" level statistics. The engine illustrates how MBL can be used for thermodynamic advantage.

Realizing the engine may provide a near-term challenge for existing experimental setups. Possible platforms include ultracold atoms [3,4,6,7,10], nitrogen-vacancy centers [8], trapped ions [9], and doped semiconductors [44], for which we provided order-of-magnitude estimates. Realizations will require platform-dependent corrections due to, e.g., variable-range hopping induced by particle-phonon interactions. As another example, semiconductors' impurities suffer from dipolar interactions. The interactions extend particles' wave functions from decaying exponentially across space to decaying as power laws.

Reversing the engine should pump heat from the cold bath to the hot, lowering the cold bath's temperature. Low temperatures facilitate quantum computation and low-temperature experiments. An MBL engine cycle might therefore facilitate state preparation and coherence preservation in quantum many-body experiments: a quantum many-body engine would cool quantum many-body systems.

We have defined as work the energy outputted during Hamiltonian tunings. Some battery must store this energy. We have refrained from specifying the battery's physical form, using an *implicit battery model*. An equivalent *explicit battery model* could depend on the experimental platform. Quantum-thermodynamics batteries have been modeled abstractly with ladderlike Hamiltonians [91]. An oscillator battery for our engine could manifest as the mode of an electromagnetic field in cavity quantum electrodynamics.

MBL is expected to have thermodynamic applications beyond this Otto engine. A localized ratchet, for example, could leverage information to transform heat into work. Additionally, the paucity of transport in MBL may have technological applications beyond thermodynamics. Dielectrics, for example, prevent particles from flowing in undesirable directions. However, dielectrics break down in strong fields. To survive, a dielectric must insulate well—as does MBL.

In addition to suggesting applications of MBL, this work identifies an opportunity within quantum thermodynamics. Athermal quantum states (e.g., $\rho \neq e^{-H/T}/Z$) are usually regarded as resources in quantum thermodynamics [16,17,19,20,22–26,92–95]. Not only athermal states, we have argued, but also athermal energy-level statistics, offer thermodynamic advantages. Generalizing the quantum-thermodynamics definition of "resource" may expand the set of goals that thermodynamic agents can achieve.

Optimization offers another theoretical opportunity. We have shown that the engine works, but better protocols could be designed. For example, we prescribe nearly quantum-adiabatic tunings. Shortcuts to adiabaticity (STA) avoid both diabatic transitions and exponentially slow tunings [28,53,62,96–98]. STA have been used to reduce other quantum engines' cycle times [28,53,98]. STA might be applied to the many-body Otto cycle, after being incorporated into MBL generally.

APPENDIX A: ANALYSIS OF THE MESOSCOPIC MBL OTTO ENGINE

In this appendix, we assess the mesoscopic engine introduced in Sec. III. Appendix A 1 reviews and introduces notation. Appendix A 2 introduces small expansion parameters. Appendix A 3 reviews the partial swap [99,100] used to model cold thermalization (stroke 2). The average heat $\langle Q_2 \rangle$ absorbed during stroke 2 is calculated in Appendix A 4; the average heat $\langle Q_4 \rangle$ absorbed during stroke 4, in Appendix A 5; the average per-trial power $\langle W_{\text{tot}} \rangle$, in Appendix A 6; and the efficiency η_{MBL}, in Appendix A 7. These calculations rely on adiabatic tuning of the Hamiltonian.

1. Notation and definitions for the mesoscopic engine

We focus on one mesoscopic engine of N sites. The engine corresponds to a Hilbert space of dimensionality $\mathcal{N} \sim \frac{2^N}{\sqrt{N}}$. The Hamiltonian, $H(t) \equiv H_{\text{meso}}(t)$, is tuned between H_{GOE}, which obeys the ETH, and H_{MBL}, which governs an MBL system. Though the energies form a discrete set, they can be approximated as continuous. ETH and MBL Hamiltonians have Gaussian DOSs:

$$\mu(E) = \frac{\mathcal{N}}{\sqrt{2\pi N}\,\mathcal{E}} e^{-E^2/(2N\mathcal{E}^2)}, \qquad \text{(A1)}$$

normalized to $\int_{-\infty}^{\infty} dE\, \mu(E) = \mathcal{N}$. The unit of energy, or energy density per site, is $\mathcal{E}$. We often extend energy integrals' limits to $\pm\infty$, as the Gaussian peaks sharply about $E = 0$.

The local average gap is $\langle \delta \rangle_E = \frac{1}{\mu(E)}$, and the average gap is $\langle \delta \rangle := \frac{\mathcal{N}}{\int_{-\infty}^{\infty} dE\, \mu^2(E)} = \frac{2\sqrt{\pi N}\,\mathcal{E}}{\mathcal{N}}$ (footnote 2). The average H_{GOE} gap, $\langle \delta \rangle$, equals the average H_{MBL} gap, by construction. $\langle \delta \rangle$ sets the scale for work and heat quantities. Hence we cast Q's and W's as (number)(function of small parameters)$\langle \delta \rangle$.

The system begins the cycle in the state $\rho(0) = e^{-\beta_H H_{\text{GOE}}}/Z$, wherein $Z := \text{Tr}(e^{-\beta_H H_{\text{GOE}}})$ denotes the partition function. W_{b} denotes the cold bath's bandwidth. We set $\hbar = k_B = 1$.

$H(t)$ is tuned at a speed $v := \mathcal{E}|\frac{d\alpha_t}{dt}|$, wherein α_t denotes the dimensionless tuning parameter. v has dimensions of energy^2, as in Ref. [101]. Though our v is not defined identically to the v in Ref. [101], ours is expected to behave similarly.

2. Small parameters of the mesoscopic engine

We estimate low-order contributions to $\langle W_{\text{tot}} \rangle$ and to η_{MBL} in terms of small parameters: (1) the cold bath has a small bandwidth: $\frac{W_{\text{b}}}{\langle \delta \rangle} \ll 1$. (2) The cold bath is cold: $\beta_C W_{\text{b}} \gg 1$. Therefore, also $1 \gg e^{-\beta_C W_{\text{b}}} \approx 0$, and $\beta_C \langle \delta \rangle \gg 1$. (3) The hot bath is hot: $\sqrt{N}\,\beta_H \mathcal{E} \ll 1$. The final assumption lets us neglect β_H from leading-order contributions to heat and work quantities. (β_H dependence manifests in factors of $e^{-N(\beta_H \mathcal{E})^2/4}$.) Since $\beta_H \mathcal{E} \ll \frac{1}{\sqrt{N}}$ and $\frac{\langle \delta \rangle}{\mathcal{E}} \ll 1$, $\beta_H \langle \delta \rangle \ll \frac{1}{\sqrt{N}}$.

We focus on the parameter regime in which

$$T_C \ll W_{\text{b}} \ll \langle \delta \rangle \quad \text{and} \quad \sqrt{N}\,\beta_H \mathcal{E} \ll 1, \qquad \text{(A2)}$$

the regime explored in the numerical simulations of Sec. VI.

3. Partial-swap model of thermalization

Classical thermalization can be modeled with a *probabilistic swap*, or *partial swap*, or *p-SWAP* [99,100]. Let a column vector $\vec{v}$ represent the state. The thermalization is broken into time steps. At each step, a doubly stochastic matrix M_p operates on $\vec{v}$. The matrix's fixed point is a Gibbs state $\vec{g}$.

M_p models a probabilistic swapping out of $\vec{v}$ for $\vec{g}$: at each time step, the system's state has a probability $1 - p$ of being preserved and a probability $p \in [0, 1]$ of being replaced by $\vec{g}$. This algorithm gives M_p the form $M_p = (1 - p)\mathbb{1} + p\vec{g}(1, 1)$.

We illustrate with thermalization across two levels. Let 0 and Δ label the levels, such that $\vec{g} = (\frac{e^{-\beta\Delta}}{1+e^{-\beta\Delta}}, \frac{1}{1+e^{-\beta\Delta}})$:

$$M_p = \begin{bmatrix} 1 - p\,\frac{1}{1+e^{-\beta\Delta}} & p\,\frac{e^{-\beta\Delta}}{1+e^{-\beta\Delta}} \\ p\,\frac{1}{1+e^{-\beta\Delta}} & 1 - p\,\frac{e^{-\beta\Delta}}{1+e^{-\beta\Delta}} \end{bmatrix}. \qquad \text{(A3)}$$

The off-diagonal elements, or transition probabilities, obey detailed balance [102,103]: $\frac{P(0 \to \Delta)}{P(\Delta \to 0)} = e^{-\beta\Delta}$.

Repeated application of M_p maps every state to $\vec{g}$ [102]: $\lim_{n \to \infty} (M_p)^n \vec{v} = \vec{g}$. The parameter p reflects the system-bath-coupling strength. We choose $p = 1$: the system thermalizes completely at each time step. (If $p \neq 1$, a more sophisticated model may be needed for thermalization across >2 levels.)

4. Average heat $\langle Q_2 \rangle$ absorbed during stroke 2

Let j denote the H_{GOE} level in which the engine begins the trial of interest. We denote by $Q_2^{(j)}$ the average heat absorbed during stroke 2, from the cold bath. ($Q_2^{(j)}$ will be negative and, provided that j is around the energy band's center, independent of j.)

The heat absorbed can be calculated easily from the following observation. Stroke 1 (adiabatic tuning) preserves the occupied level index. The level closest to j lies a distance δ away when stroke 3 begins. δ can have either sign, can lie above or below j. Heat is exchanged only if $|\delta| < W_b$. Let us initially neglect the possibility that two nearby consecutive gaps are very small, that $|E_{j\pm 2} - E_j| \leqslant W_b$. We can write the average (over trials begun in level j) heat absorbed as

$$Q_2^{(j)} = \int_{-W_{\text{b}}}^{W_{\text{b}}} d\delta\, \delta\, \frac{e^{-\beta_C \delta}}{1 + e^{-\beta_C \delta}}\, P_{\text{MBL}}^{(E)}(\delta) + O\left(W_{\text{b}}^3/\langle \delta \rangle^2\right). \qquad \text{(A4)}$$

This equation assumes a Sommerfeld-expansion form, as the Boltzmann factor is $\frac{e^{-\beta_C \delta}}{1+e^{-\beta_C \delta}} = \Theta(-\delta) + \text{sgn}(\delta)\,\frac{e^{-\beta_C|\delta|}}{1+e^{-\beta_C|\delta|}}$. Hence

$$Q_2^{(j)} = -\frac{W_{\text{b}}^2}{2}\mu(E) + \frac{\pi^2}{6}\mu(E)(T_C)^2$$
$$+ O([W_{\text{b}}]^3/\langle \delta \rangle^2) + O(\mu(E)^2[T_C]^3). \qquad \text{(A5)}$$

The first correction accounts for our not considering two levels within W_b of level j.

(a) $|\langle Q_2 \rangle|$ vs. W_b at $T_C = 0$ and $T_H = \infty$

(b) $|\langle Q_2 \rangle|$ vs. T_C at $T_H = \infty$ and $W_b = 2^{-4} \langle \delta \rangle$

(c) $|\langle Q_2 \rangle|$ vs. β_H at $T_C = 0$ and $W_b = 2^{-4} \langle \delta \rangle$
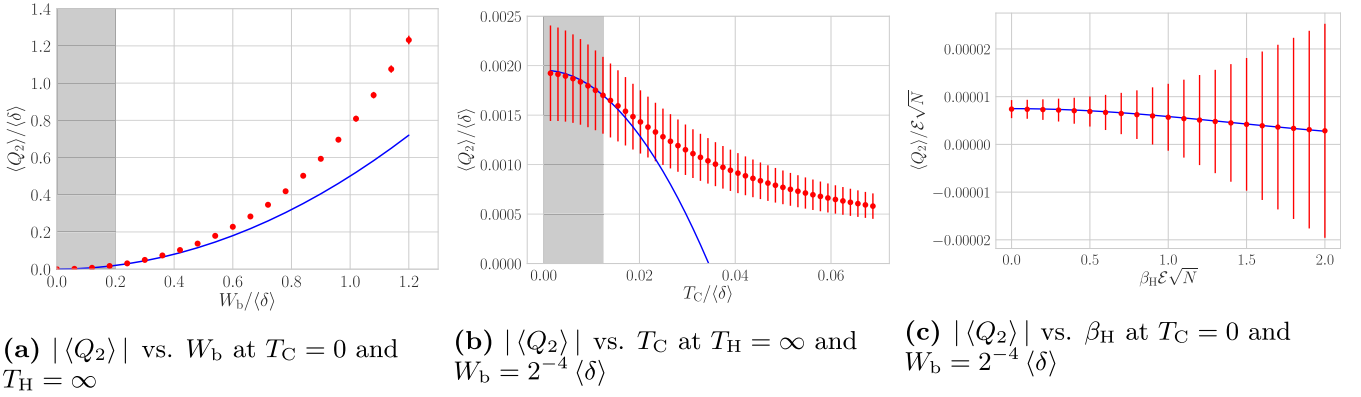
FIG. 7. Magnitude $|\langle Q_2 \rangle|$ of the average heat absorbed during cold thermalization (stroke 2) as a function of the cold-bath bandwidth W_b (a), the cold-bath temperature T_C (b), and the hot-bath temperature $T_H = 1/\beta_H$ (c). The blue lines represent the magnitude of the analytical prediction (A10). See Sec. VI for other parameters and definitions. The analytics match the numerics' shapes, and the agreement is fairly close, in the appropriate limits (where $\frac{W_b}{\langle \delta \rangle} \ll 1$ and $T_C/\langle \delta \rangle \ll 1$, in the gray shaded regions). The analytics systematically underestimate $|\langle Q_2 \rangle|$ at fixed W_b, due to the small level repulsion at finite N. The analytical prediction (A10) substantially underestimates $|\langle Q_2 \rangle|$ when the cold-bath bandwidth is large, $W_b \gtrsim \langle \delta \rangle$. Such disagreement is expected: the analytics rely on $\frac{W_b}{\langle \delta \rangle} \ll 1$, neglecting chains of small gaps: $\delta'_j, \delta'_{j+1}, \cdots < W_b$. Such chains proliferate as W_b grows. A similar reason accounts for the curve's crossing the origin in (b): we analytically compute $\langle Q_2 \rangle$ only to second order in $T_C/\langle \delta \rangle$.

Next, we need to average this result over all initial states j, assuming the initial density operator, $\rho(0) = e^{-\beta_H H_{GOE}}/Z$:

$$\langle Q_2 \rangle := \left\langle \left\langle \langle Q_2(E) \rangle_{\substack{\text{cold} \\ \text{therm.}}} \right\rangle_{\text{gaps}} \right\rangle_{\rho(0)} \quad \text{(A6)}$$

$$= \left(-\frac{(W_b)^2}{2} + \frac{\pi^2}{6} \frac{1}{(\beta_C)^2} \right) \int_{-\infty}^{\infty} dE \, \mu^2(E) \frac{e^{-\beta_H E}}{Z}$$

$$+ \langle \delta \rangle \left\{ O\left(\left[\frac{W_b}{\langle \delta \rangle} \right]^3 \right) + O\left(\frac{W_b}{\langle \delta \rangle} e^{-\beta_C W_b} \right) \right.$$

$$\left. + O\left(\left[\frac{\mu(E)}{\beta_C} \right]^3 \right) \right\}. \quad \text{(A7)}$$

We substitute in for the DOS from Eq. (A1):

$$\langle Q_2 \rangle = \frac{\mathcal{N}^2}{2\pi N \mathcal{E}^2} \frac{1}{Z} \left(-\frac{(W_b)^2}{2} + \frac{\pi^2}{6} \frac{1}{(\beta_C)^2} \right)$$

$$\times \int_{-\infty}^{\infty} dE \, e^{-E^2/N\mathcal{E}^2} e^{-\beta_H E} + O(.), \quad \text{(A8)}$$

wherein the correction terms are abbreviated. The integral evaluates to $\sqrt{\pi N} \, \mathcal{E} \, e^{N(\beta_H \mathcal{E})^2/4}$. The partition function is

$$Z = \int_{-\infty}^{\infty} dE \, \mu(E) e^{-\beta_H E} = \mathcal{N} e^{N(\beta_H \mathcal{E})^2/2}. \quad \text{(A9)}$$

Substituting into Eq. (A8) yields

$$\langle Q_2 \rangle = \left(-\frac{(W_b)^2}{2\langle \delta \rangle} + \frac{\pi^2}{6} \frac{1}{(\beta_C)^2 \langle \delta \rangle} \right) e^{-N(\beta_H \mathcal{E})^2/4}$$

$$+ \langle \delta \rangle \left\{ O\left(\left[\frac{W_b}{\langle \delta \rangle} \right]^3 \right) + O\left([\mu(E) W_b] \frac{\mu(E)}{\beta_C} e^{-\beta_C W_b} \right) \right.$$

$$\left. + O\left(\left[\frac{\mu(E)}{\beta_C} \right]^3 \right) + O([\sqrt{N} \beta_H \mathcal{E}]^4) \right\}. \quad \text{(A10)}$$

We have replaced the prefactor with $1/\langle \delta \rangle$, using Eq. (10).

Equation (A10) is compared with numerical simulations in Fig. 7. In the appropriate regime (wherein $W_b \ll \langle \delta \rangle$ and $T_C \ll W_b$), the analytics agree well with the numerics, to within finite-size effects.

In terms of small dimensionless parameters,

$$\langle Q_2 \rangle = \langle \delta \rangle \left[-\frac{1}{2} \left(\frac{W_b}{\langle \delta \rangle} \right)^2 + \frac{\pi^2}{6} \frac{1}{(\beta_C \langle \delta \rangle)^2} \right]$$

$$\times \left[1 - \frac{N}{4} (\beta_H \mathcal{E})^2 \right] + O(.). \quad \text{(A11)}$$

The leading-order term is second-order. So is the β_C correction; but $\frac{1}{(\beta_C \langle \delta \rangle)^2} \ll \left(\frac{W_b}{\langle \delta \rangle} \right)^2$, by assumption [Eq. (A2)]. The β_H correction is fourth-order—too small to include. To lowest order,

$$\boxed{\langle Q_2 \rangle \approx -\frac{(W_b)^2}{2\langle \delta \rangle}.} \quad \text{(A12)}$$

5. Average heat $\langle Q_4 \rangle$ absorbed during stroke 4

The $\langle Q_4 \rangle$ calculation proceeds similarly to the $\langle Q_2 \rangle$ calculation. When calculating $\langle Q_2 \rangle$, however, we neglected contributions from the engine's cold-thermalizing down two small gaps. Two successive gaps have a joint probability $\sim (\frac{W_b}{\langle \delta \rangle})^2$ of being $< W_b$ each. Thermalizing across each gap, the engine absorbs heat $\leqslant W_b$. Each such pair therefore contributes negligibly to $\langle Q_2 \rangle$, as $\langle \delta \rangle O([\frac{W_b}{\langle \delta \rangle}]^3)$.

We cannot neglect these pairs when calculating $\langle Q_4 \rangle$. Each typical small gap widens, during stroke 3, to a size $\sim \langle \delta \rangle$. These larger gaps are thermalized across during stroke 4, contributing at the nonnegligible second order, as $\sim \langle \delta \rangle O([\frac{W_b}{\langle \delta \rangle}]^2)$ to $\langle Q_4 \rangle$. Chains of $\geqslant 3$ small MBL gaps contribute negligibly.

(a) $\langle Q_4 \rangle$ vs. W_b at $T_C = 0$ and $T_H = \infty$

(b) $\langle Q_4 \rangle$ vs. T_C at $T_H = \infty$ and $W_b = 2^{-4} \langle \delta \rangle$

(c) $\langle Q_4 \rangle$ vs. β_H at $T_C = 0$ and $W_b = 2^{-4} \langle \delta \rangle$
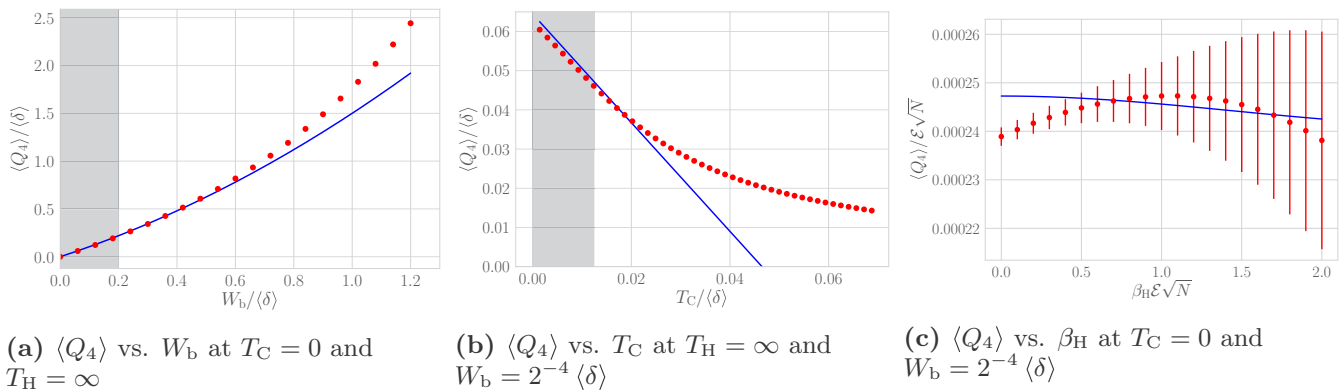
FIG. 8. Average heat $\langle Q_4 \rangle$ absorbed during hot thermalization (stroke 4) as a function of the cold-bath bandwidth W_b (a), the cold-bath temperature T_C (b), and the hot-bath temperature $T_H = 1/\beta_H$ (c). The blue lines represent the analytical prediction (A13), to lowest order in T_C, with the β_H dependence of $\langle Q_4 \rangle$, too small a correction to include in Eq. (A13): $\langle Q_4 \rangle \approx W_b - \frac{2 \ln 2}{\beta_C} + \frac{(W_b)^2}{2 \langle \delta \rangle} e^{-N(\beta_H \mathcal{E})^2/4}$. See Sec. VI for other parameters and definitions. The analytics' shapes agree with the numerics, and the fit is fairly close, in the appropriate limits (where $e^{-\beta_C W_b} \ll 1$, $\frac{1}{\beta_C \langle \delta \rangle} \ll 1$, and $\frac{W_b}{\langle \delta \rangle} \ll 1$, in the gray shaded regions). The predictions underestimate $\langle Q_4 \rangle$; see the Fig. 7 caption. (c) suggests that the numerics deviate significantly from the analytics: the numerics appear to depend on β_H via a linear term absent from the $\langle Q_4 \rangle$ prediction. This seeming mismatch appears symptomatic of finite sample and system sizes.

The calculation is tedious, appears in Ref. [[49], Appendix G 5], and yields

$$\langle Q_4 \rangle \approx W_b - \frac{2 \ln 2}{\beta_C} + \frac{(W_b)^2}{2 \langle \delta \rangle} + 4 \ln 2 \frac{W_b}{\beta_C \langle \delta \rangle}. \quad (A13)$$

The leading-order terms are explained heuristically below Eq. (13) in the main text.

The leading-order β_C correction, $-\frac{2 \ln 2}{\beta_C}$, shows that a warm cold bath lowers the heat required to reset the engine. Suppose that the cold bath is maximally cold: $T_C = 0$. Consider any trial that the engine begins just above a working gap (an ETH gap $\delta > W_b$ that narrows to an MBL gap $\delta' < W_b$). Cold thermalization drops the engine deterministically to the lower level. During stroke 4, the engine must absorb $Q_4 > 0$ to return to its start-of-trial state. Now, suppose that the cold bath is only cool: $T_C \gtrsim 0$. Cold thermalization might leave the engine in the upper level. The engine needs less heat, on average, to reset than if $T_C = 0$. A finite T_C, therefore, detracts from $\langle Q_4 \rangle$. The $+4 \ln 2 \frac{W_b}{\beta_C \langle \delta \rangle}$ offsets the detracting. However, the positive correction is smaller than the negative correction, as $\frac{W_b}{\langle \delta \rangle} \ll 1$.

A similar argument concerns $T_H < \infty$. However, the β_H correction is too small to include in Eq. (A13): $\langle Q_4 \rangle \approx W_b - \frac{2 \ln 2}{\beta_C} + \frac{(W_b)^2}{2 \langle \delta \rangle} e^{-N(\beta_H \mathcal{E})^2/4}$.

Figure 8 shows Eq. (A13), to lowest order in T_C, as well as the β_H dependence of $\langle Q_4 \rangle$. The analytical prediction is compared with numerical simulations. The agreement is close, up to finite-size effects, in the appropriate regime ($T_C \ll W_b \ll \langle \delta \rangle$).

6. Average per-cycle power $\langle W_{tot} \rangle$

By the first law of thermodynamics, the net work outputted by the engine equals the net heat absorbed. Summing Eqs. (A13) and (A12) yields the per-trial power, or average

work outputted per engine cycle:

$$\langle W_{tot} \rangle = \langle Q_2 \rangle + \langle Q_4 \rangle \approx W_b - \frac{2 \ln 2}{\beta_C} + 4 \ln 2 \frac{W_b}{\beta_C \langle \delta \rangle}. \quad (A14)$$

The leading-order β_H correction is negative and too small to include—of order $\langle \delta \rangle \left(\frac{W_b}{\langle \delta \rangle} \right)^2 N (\beta_H \mathcal{E})^2$. Equation (A14) agrees well with the numerics in the appropriate limits ($T_C \ll W_b \ll \langle \delta \rangle$) and beyond, as shown in Fig. 9. The main text contains the primary analysis of Eq. (A14). Here, we discuss the $\langle Q_2 \rangle$ correction, limiting behaviors, and scaling.

The negative $\langle Q_2 \rangle = -\frac{(W_b)^2}{\langle \delta \rangle}$ detracts little from the leading term W_b of $\langle Q_4 \rangle$: $\frac{(W_b)^2}{\langle \delta \rangle} \ll W_b$, since $\frac{W_b}{\langle \delta \rangle} \ll 1$. The $\langle Q_2 \rangle$ cuts down on the per-trial power little.

The limiting behavior of Eq. (A14) makes sense: consider the limit as $W_b \to 0$. The cold bath has too small a bandwidth to thermalize the engine, so the engine should output no work, on average. Indeed, the first and third terms in Eq. (A14) vanish, being proportional to W_b. The second term vanishes because $\beta_C \to \infty$ more quickly than $W_b \to 0$, by Eq. (A2): the cold bath is very cold.

Equation (A14) scales with the system size N no more quickly than $\sqrt{N}/2^N$, by the assumption $W_b \ll \langle \delta \rangle \sim \sqrt{N}/2^N$. This scaling makes sense: the engine outputs work because the energy eigenvalues meander upward and downward in Fig. 3 as $H(t)$ is tuned. In the thermodynamic limit, levels squeeze together. Energy eigenvalues have little room in which to wander, and S outputs little work. Hence our parallelization of fixed-length mesoscopic subengines in the thermodynamic limit (Sec. IV).

7. Efficiency η_{MBL} in the adiabatic approximation

The efficiency is defined as

$$\eta_{MBL} := \frac{\langle W_{tot} \rangle}{\langle Q_{in} \rangle}. \quad (A15)$$
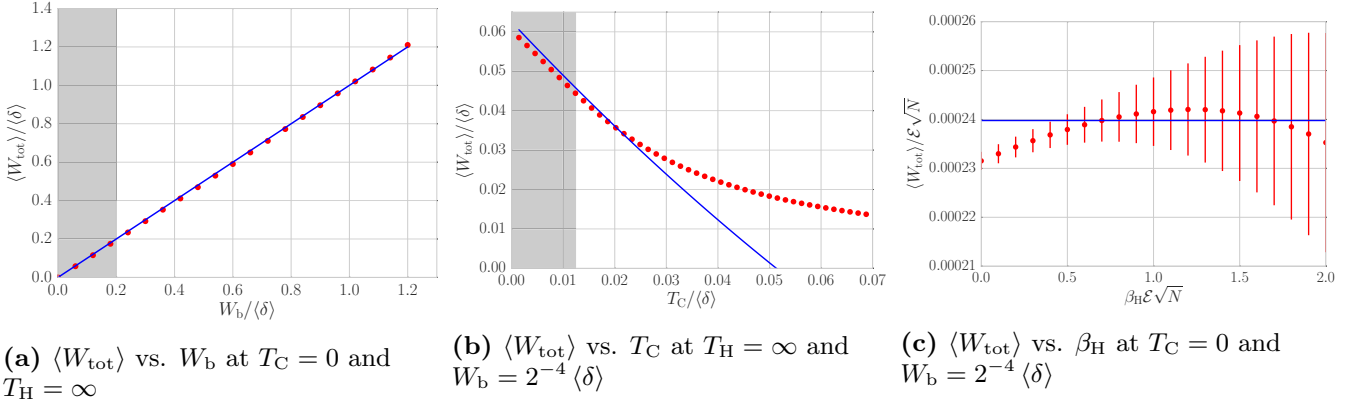
(a) $\langle W_{\mathrm{tot}} \rangle$ vs. W_{b} at $T_{\mathrm{C}} = 0$ and $T_{\mathrm{H}} = \infty$

(b) $\langle W_{\mathrm{tot}} \rangle$ vs. T_{C} at $T_{\mathrm{H}} = \infty$ and $W_{\mathrm{b}} = 2^{-4} \langle \delta \rangle$

(c) $\langle W_{\mathrm{tot}} \rangle$ vs. β_{H} at $T_{\mathrm{C}} = 0$ and $W_{\mathrm{b}} = 2^{-4} \langle \delta \rangle$

FIG. 9. Per-cycle power $\langle W_{\mathrm{tot}} \rangle$ as a function of the cold-bath bandwidth W_{b} (a), the cold-bath temperature T_{C} (b), and the hot-bath temperature $T_{\mathrm{H}} = 1/\beta_{\mathrm{H}}$ (c). The blue lines represent the analytical prediction $\langle W_{\mathrm{tot}} \rangle \approx W_{\mathrm{b}} - \frac{2 \ln 2}{\beta_{\mathrm{C}}}$: Eq. (A14), to first order in $\frac{W_{\mathrm{b}}}{\langle \delta \rangle}$ and in $\frac{1}{\beta_{\mathrm{C}} \langle \delta \rangle}$. The analytics largely agree with the numerics in the appropriate regime: $\frac{W_{\mathrm{b}}}{\langle \delta \rangle} \ll 1$ and $\frac{T_{\mathrm{C}}}{\langle \delta \rangle} \ll 1$ (in the gray shaded region). Outside that regime, the analytics underestimate $\langle W_{\mathrm{tot}} \rangle$; see Fig. 7 for an analysis. (c) suggests that the numerics depend on β_{H} via a linear term absent from the analytical prediction; see the caption of Fig. 8(c).

The numerator is averaged separately from the denominator because averaging W_{tot} over runs of one mesoscopic engine is roughly equivalent to averaging over simultaneous runs of parallel subengines in one macroscopic engine. $\frac{\langle W_{\mathrm{tot}} \rangle}{\langle Q_{\mathrm{in}} \rangle}$ may therefore be regarded as the $\frac{W_{\mathrm{tot}}}{Q_{\mathrm{in}}}$ of one macroscopic-engine trial.

The positive-heat-absorbing-stroke is stroke 4, in the average trial:

$$\langle Q_{\mathrm{in}} \rangle = \langle Q_4 \rangle = \langle W_{\mathrm{tot}} \rangle - \langle Q_2 \rangle$$
$$= \langle W_{\mathrm{tot}} \rangle \left(1 - \frac{\langle Q_2 \rangle}{\langle W_{\mathrm{tot}} \rangle} \right) = \langle W_{\mathrm{tot}} \rangle (1 + \phi), \quad \text{(A16)}$$

wherein

$$\phi := -\frac{\langle Q_2 \rangle}{\langle W_{\mathrm{tot}} \rangle} \approx \frac{W_{\mathrm{b}}}{2 \langle \delta \rangle}. \quad \text{(A17)}$$

Substituting from Eq. (A16) into Eq. (A15) yields

$$\eta_{\mathrm{MBL}} \approx \frac{\langle W_{\mathrm{tot}} \rangle}{\langle W_{\mathrm{tot}} \rangle (1 + \phi)} \approx 1 - \phi = 1 - \frac{W_{\mathrm{b}}}{2 \langle \delta \rangle}. \quad \text{(A18)}$$

Using suboptimal baths diminishes the efficiency. Adding β_{C}-dependent terms from Eq. (A14) to $\langle W_{\mathrm{tot}} \rangle$ yields

$$\phi' = \frac{W_{\mathrm{b}}}{2 \langle \delta \rangle} + \frac{\ln 2}{\beta_{\mathrm{C}} \langle \delta \rangle} - 2 \ln 2 \frac{W_{\mathrm{b}}}{\langle \delta \rangle} \frac{1}{\beta_{\mathrm{C}} \langle \delta \rangle}. \quad \text{(A19)}$$

The β_{H} correction, $1 - \frac{W_{\mathrm{b}}}{2 \langle \delta \rangle} e^{-N(\beta_{\mathrm{H}} \mathcal{E})^2/4}$, is too small to include. The correction shares the sign of β_{H}: a lukewarm hot bath lowers the efficiency.

Expressions (A18) and (A19) are compared with results from numerical simulations in Fig. 10. The analytics agree with the numerics in the appropriate regime ($T_{\mathrm{C}} \ll W_{\mathrm{b}} \ll \langle \delta \rangle$).

APPENDIX B: PHENOMENOLOGICAL MODEL FOR THE MACROSCOPIC MBL OTTO ENGINE

The macroscopic MBL Otto engine benefits from properties of MBL (Sec. IV), localization and local level repulsion.

We understand these properties from Anderson insulators [72] and perturbation theory. Anderson insulators are reviewed in Appendix B 1. Local level repulsion in Anderson insulators [45] in the strong-disorder limit is reviewed in Appendix B 2. Appendix B 3 extends local level repulsion to MBL. Local level repulsion's application to the MBL engine is discussed in Appendix B 4. Throughout this section, N denotes the whole system's length.

1. Anderson localization

Consider a 1D spin chain or, equivalently, a lattice of spinless fermions. An Anderson-localized Hamiltonian H_{And} has almost the form of Eq. (32), but three elements are removed: the t dependence, $Q(h(\alpha_t))$, and the interaction ($\boldsymbol{\sigma}_j \cdot \boldsymbol{\sigma}_{j+1}$ is replaced with $\sigma_j^+ \sigma_{j+1}^- + \text{H.c.}$).

Let $|0\rangle$ denote some reference state in which all the spins point downward (all the fermionic orbitals are empty). In this section, we focus, for concreteness, on the properties of single-spin excitations relative to $|0\rangle$ [45,72]. The ℓth excitation is represented, in fermionic notation, as $\sum_x \psi_\ell(x) \sigma_x^+ |0\rangle$. The single-excitation wave functions $\psi_\ell(x)$ are localized: x_ℓ denotes the point at which the probability density $|\psi_\ell(x)|^2$ peaks. The wave function decays exponentially with the distance $|x - x_\ell|$ from the peak:

$$\psi_\ell(x) \approx \sqrt{\frac{2}{\xi_{\mathrm{And}}}} \, e^{-|x - x_\ell|/\xi_{\mathrm{And}}}. \quad \text{(B1)}$$

The localization length varies with the Hamiltonian parameters as

$$\xi_{\mathrm{And}} \sim \frac{1}{\ln h} \quad \text{(B2)}$$

at large disorder, whose overall strength is h.

2. Local level repulsion in Anderson insulators

We begin with the infinitely localized limit, $h \to \infty$. We take $\mathcal{E} \to 0$ to keep the Hamiltonian's energy scale finite. The

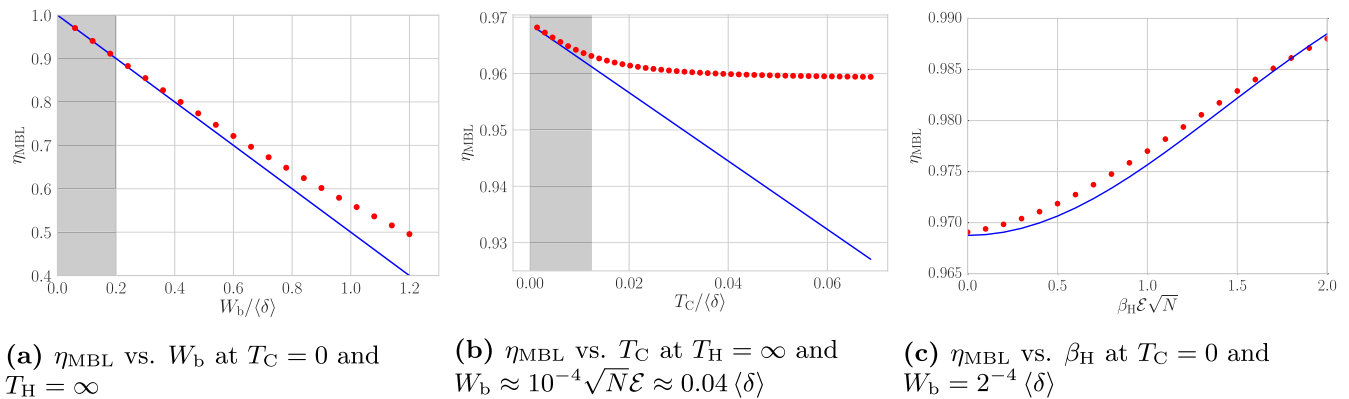**(a)** η_{MBL} vs. W_{b} at $T_{\text{C}} = 0$ and $T_{\text{H}} = \infty$

(b) η_{MBL} vs. T_{C} at $T_{\text{H}} = \infty$ and $W_{\text{b}} \approx 10^{-4}\sqrt{N}\mathcal{E} \approx 0.04\langle\delta\rangle$

(c) η_{MBL} vs. β_{H} at $T_{\text{C}} = 0$ and $W_{\text{b}} = 2^{-4}\langle\delta\rangle$

FIG. 10. Efficiency η_{MBL} as a function of the cold-bath bandwidth W_{b} (a), the cold-bath temperature T_{C} (b), and the hot-bath temperature $T_{\text{H}} = 1/\beta_{\text{H}}$ (c). The blue lines represent the analytical predictions (A18) and (A19). (c) shows the leading-order β_{H} dependence of η_{MBL}, a correction too small to include in Eq. (A19): $1 - \frac{W_{\text{b}}}{2\langle\delta\rangle}e^{-N(\beta_{\text{H}}\mathcal{E})^2/4}$. See Sec. VI for other parameters and definitions. The analytics agree with the numerics fairly well in the appropriate regime ($\frac{W_{\text{b}}}{\langle\delta\rangle} \ll 1$, $\frac{T_{\text{C}}}{\langle\delta\rangle} \ll 1$, and $\sqrt{N}\,T_{\text{H}}\mathcal{E} \ll 1$). The analytics underestimate η_{MBL}; see the Fig. 7 caption.

hopping terms can be neglected, and particles on different sites do not repel. Single-particle excitations are localized on single sites. The site-i excitation corresponds to an energy $2\mathcal{E}hh_i$. Since the on-site potentials $h \cdot h_i$ are uncorrelated, neighboring-site excitations' energies are uncorrelated.

Let us turn to large but finite h. Recall that $h \cdot h_i$ is drawn uniformly at random from $[-h, h]$. The uniform distribution has a standard deviation of $\frac{h}{\sqrt{3}} \gg 1$. Therefore $h|h_i - h_{i+1}| \gg 1$ for most pairs of neighboring sites. The hopping affects these sites' wave functions and energies weakly. However, with a probability $\sim\frac{1}{h}$, neighboring sites have local fields $h \cdot h_i$ and $h \cdot h_{i+1}$ such that $h|h_i - h_{i+1}| \lesssim 1$. The hopping hybridizes such sites. The hybridization splits the sites' eigenvalues by an amount $\sim\sqrt{h^2(h_i - h_{i+1})^2 + \mathcal{E}^2} \geqslant \mathcal{E}$.

Consider, more generally, two sites separated by a distance L. Suppose that the sites' disorder-field strengths are separated by $< 1/h^L$. (The upper bound approximates the probability amplitude associated with a particle's hopping the L intervening sites.) The sites' excitation energies and energy eigenfunctions are estimated perturbatively. The expansion parameter is $1/h$. To zeroth order, the energies are uncorrelated and (because $h|h_i - h_{i+L}| < 1/h^L$) are split by $< \mathcal{E}/h^L$. The eigenfunctions are hybridized at order L. The perturbed energies are split by $\geqslant \mathcal{E}/h^L \sim \mathcal{E}e^{-L/\xi_{\text{And}}}$. [Recall that $\xi_{\text{And}} \sim 1/\ln h$, by Eq. (B2).]

Hence eigenstates localized on nearby sites have correlated energies: *the closer together sites lie in real space, the lower the probability that they correspond to similar energies.* This conclusion agrees with global Poisson statistics: consider a large system of $N \gg 1$ sites. Two randomly chosen single-particle excitations are typically localized a distance $\sim N$ apart. The argument above implies only that the energies lie $> \mathcal{E}e^{-N/\xi_{\text{And}}}$ apart. This scale is exponentially smaller (in N) than the average level spacing $\sim\frac{\mathcal{E}h}{N}$ between single-particle excitations.[10]

We can quantify more formally the influence of hybridization on two energies separated by ω and associated with eigenfunctions localized a distance L apart. The *level correlation function* is defined as

$$R(L, \omega) := \frac{1}{N^2}\sum_{i,n,n'}|\langle 0|\sigma_i^-|n\rangle|^2\,|\langle 0|\sigma_{i+L}^-|n'\rangle|^2$$
$$\times\,\delta(E_n - E_{n'} - \omega) - \tilde{\mu}(\omega)^2. \quad \text{(B3)}$$

The spatially averaged density of states at frequency ω is denoted by $\tilde{\mu}(\omega) := \frac{1}{N}\sum_n |\langle 0|\sigma_i^-|n\rangle|^2\,\delta(E_n - \omega)$. $|n\rangle$ and $|n'\rangle$ denote eigenstates, corresponding to single-particle excitations relative to $|0\rangle$, associated with energies E_n and $E_{n'}$. In the Anderson insulator, $R(L, \omega) \approx 0$ when $\omega \gg \mathcal{E}e^{-L/\xi_{\text{And}}}$: Levels are uncorrelated when far apart in space and/or energy. When energies are close ($\omega \lesssim \mathcal{E}e^{-L/\xi_{\text{And}}}$), $R(L, \omega)$ is negative. These levels repel (in energy space).

3. Generalization to many-body localization

The estimates above can be extended from single-particle Anderson-localized systems to MBL systems initialized in arbitrary energy eigenstates (or in position-basis product states). $R(L, \omega)$ is formulated in terms of matrix elements $\langle 0|\sigma_i^-|n\rangle$ of local operators σ_i^-. The local operators relevant to Anderson insulators have the forms of the local operators relevant to MBL systems. Hence $R(L, \omega)$ is defined for MBL as for Anderson insulators. However, $|0\rangle$ now denotes a generic many-body state.

Let us estimate the scale $\mathcal{J}_L$ of the level repulsion between MBL energies, focusing on exponential behaviors. The MBL energy eigenstates result from perturbative expansions

---

[10]The average level spacing between single-particle excitations scales as $\sim 1/N$ for the following reason. The reference state $|0\rangle$

consists of N downward-pointing spins. Flipping one spin upward yields a single-particle excitation. N single-particle-excitation states exist, as the chain contains N sites. Each site has an energy $\sim \pm\mathcal{E}h$, to zeroth order, as explained three paragraphs ago. The excitation energies therefore fill a band of width $\sim\mathcal{E}h$. An interval $\sim\frac{\mathcal{E}h}{N}$ therefore separates single-particle-excitation energies, on average.

about Anderson energy eigenstates. Consider representing the Hamiltonian as a matrix $\mathcal{M}$ with respect to the true MBL energy eigenbasis. Off-diagonal matrix elements couple together unperturbed states. These couplings hybridize the unperturbed states, forming corrections. The couplings may be envisioned as rearranging particles throughout a distance L.

MBL dynamics is unlikely to rearrange particles across considerable distances, due to localization. Such a rearrangement is encoded in an off-diagonal element $\mathcal{M}_{ij}$ of $\mathcal{M}$. This $\mathcal{M}_{ij}$ must be small—suppressed exponentially in L. $\mathcal{M}_{ij}$ also forces the eigenstates' energies apart, contributing to level repulsion [[49], Appendix F]. Hence the level-repulsion scale is suppressed exponentially in L:

$$\mathcal{J}_L \sim \mathcal{E}e^{-L/\zeta}, \tag{B4}$$

for some ζ. At infinite temperature, ζ must $< \frac{1}{\ln 2}$ for the MBL phase to remain stable [104]. Substituting into Eq. (B4) yields $\mathcal{J}_L < \frac{\mathcal{E}}{2^L}$. The level-repulsion scale is smaller than the average gap.

The size and significance of $\mathcal{J}_L$ depend on the size of L. At the crossover distance ξ, the repulsion $\mathcal{J}_L$ (between energy eigenfunctions localized a distance ξ apart) becomes comparable to the average gap $\sim \frac{\mathcal{E}}{2^\xi}$ between the eigenfunctions in the same length-ξ interval: $\mathcal{E}e^{-\xi/\zeta} \sim \frac{1}{e} \frac{\mathcal{E}}{2^\xi}$. Solving for the crossover distance yields

$$\xi \sim \frac{1}{\frac{1}{\zeta} - \ln 2}. \tag{B5}$$

Relation (B5) provides a definition of the MBL localization length ξ. [This ξ differs from the Anderson localization length ξ_{And}, Eq. (B2).] Solving for ζ yields

$$\zeta \sim \frac{1}{\frac{1}{\xi} + \ln 2}. \tag{B6}$$

The MBL Otto cycle involves two localization lengths in the thermodynamic limit. In the shallowly localized regime, $\xi = \xi_>$. Each eigenfunction has significant weight on $\xi_> \approx 10$ sites, in an illustrative example. In the highly localized regime, $\xi = \xi_<$. Eigenfunctions peak tightly: $\xi_< \approx 1$.

Suppose that the particles are rearranged across a large distance $L \gg \xi$. The level-repulsion scale

$$\mathcal{J}_{L \gg \xi} \sim \mathcal{E}e^{-L/\xi} 2^{-L}. \tag{B7}$$

In the MBL engine's very localized regime, in which $\xi = \xi_<$, if $L = \xi_>$ equals one subengine's length, $\mathcal{J}_{L \gg \xi} = \delta_-$.

Now, suppose that particles are rearranged across a short distance $L \lesssim \xi$. Random-matrix theory approximates this scenario reasonably (while slightly overestimating the level repulsion). We can approximate the repulsion between nearby-eigenfunction energies with the average gap $\langle \delta \rangle^{(L)}$ in the energy spectrum of a length-L system:

$$\mathcal{J}_{L \lesssim \xi} \sim \langle \delta \rangle^{(L)} \sim \frac{\mathcal{E}}{2^L}. \tag{B8}$$

4. Application of local level repulsion to the MBL Otto engine in the thermodynamic limit

Consider perturbing an MBL system locally. In the Heisenberg picture, the perturbing operator spreads across

a distance $L(t) \sim \zeta \ln(\mathcal{E}t)$ [15]. (See also Ref. [73].) The longer the time t for which the perturbation lasts, the farther the influence spreads.

Consider tuning the Hamiltonian infinitely slowly, to preclude diabatic transitions: $t \to \infty$. Even if the Hamiltonian consists of spatially local terms, the perturbation to each term spreads across the lattice. The global system cannot be subdivided into independent subengines. The global system's average gap vanishes in the thermodynamic limit: $\langle \delta \rangle \to 0$. Since $\langle W_{\text{tot}} \rangle \sim W_b \ll \langle \delta \rangle$, the per-cycle power seems to vanish in the thermodynamic limit: $W_b \to 0$.

Now, consider tuning the Hamiltonian at a finite speed v. Dimensional analysis suggests that the relevant time scale is $t \sim \frac{\mathcal{E}}{v}$. Local perturbations affect a region of length $\sim L(\mathcal{E}/v) \sim \zeta \ln(\mathcal{E}^2/v)$. On a length scale $L(\mathcal{E}/v)$, global level correlations govern the engine's performance less than local level correlations do, i.e., less than $R(L(\mathcal{E}/v), \omega)$ does. This correlator registers level repulsion at a scale independent of N. Finite-speed tuning renders finite the average gap accessible to independent subengines, the $\langle \delta \rangle$ that would otherwise close in the thermodynamic limit. Each mesoscale subengine therefore outputs $\langle W_{\text{tot}} \rangle > 0$.

We can explain the gap's finiteness differently: suppose that the engine's state starts some trial with weight on the jth energy level. The eigenenergies wiggle up and down during stroke 1. The jth energy may approach the $(j-1)$th. Such close-together energies likely correspond to far-apart subengines. If the levels narrowly avoided crossing, particles would be rearranged across a large distance. Particles must not be, as subengines must function independently. Hence the engine must undergo a diabatic transition: The engine's state must retain its configuration. The engine must behave as though the approaching energy level did not exist. Effectively removing the approaching level from the available spectrum effectively creates a gap in the spectrum. One can create such an effective gap (can promote such diabatic transitions) by tuning the Hamiltonian at a finite v.

APPENDIX C: CONSTRAINT 2 ON COLD THERMALIZATION: SUPPRESSION OF HIGH-ORDER-IN-THE-COUPLING ENERGY EXCHANGES

Section V introduces the dominant mechanism by which the bath changes a subengine's energy. The energy changes by an amount $\sim W_b$, at a rate $\sim g$. Higher-order processes can change the subengine energy by amounts $> W_b$ and operate at rates $O(g^\ell)$, wherein $\ell \geqslant 2$. The subengine should thermalize across just small gaps $\delta \leqslant W_b$. Hence the rate-g^ℓ processes must operate much more slowly than the rate-g processes: g must be small. We describe the higher-order processes, upper-bound g, and lower-bound τ_{th}.

The higher-order processes can be understood as follows. Let $H_{\text{tot}} = H_{\text{macro}}(t) + H_{\text{bath}} + H_{\text{int}}$ denote the Hamiltonian that governs the engine-and-bath composite. H_{tot} generates the time-evolution operator $U(t) := e^{-iH_{\text{tot}}t}$. Consider Taylor-expanding $U(t)$. The ℓth term is suppressed in g^ℓ, contains 2ℓ fermion operators c_j and $c_{j'}^\dagger$, and contains ℓ boson operators b_ω and $b_{\omega'}^\dagger$. This term encodes the absorption, by the bath, of ℓ energy quanta of sizes $\leqslant W_b$. The subengine gives the bath a total amount $\sim \ell W_b$ of heat. The subengine should not lose so much heat. Hence higher-order processes should occur much

more slowly than the rate-g processes:

$$\tau_{\text{high-ord.}} \gg \tau_{\text{th}}. \tag{C1}$$

Let us construct an expression for the left-hand side. Which processes most urgently require suppressing? Processes that change the subengine's energy by $\gtrsim \langle \delta \rangle$. Figure 3 illustrates why. If the right-hand leg has length $\gtrsim \langle \delta \rangle$, the right-hand leg could be longer than the left-hand leg. If it were, the trial would yield net negative work, $W_{\text{tot}} < 0$. The bath would absorb energy $\langle \delta \rangle$ from a subengine by absorbing $\sim \frac{\langle \delta \rangle}{W_{\text{b}}}$ packets of energy $\sim W_{\text{b}}$ each. Hence the bath would appear to need to flip $\sim L = \frac{\langle \delta \rangle}{W_{\text{b}}}$ spins to absorb energy $\sim \langle \delta \rangle$. (We switch from fermion language to spin language for convenience.) However, the length-L spin subchain has a discrete effective energy spectrum. The spectrum might lack a level associated with the amount (initial energy) $- \langle \delta \rangle$ of energy. If so, the bath must flip more than $\frac{\langle \delta \rangle}{W_{\text{b}}}$ spins—local level correlations suggest $\sim \xi_>$ spins (Appendix B). Hence $L = \max\{\frac{\langle \delta \rangle}{W_{\text{b}}}, \xi_>\}$. Energy is rearranged across the distance L at a rate $\propto g^L$.

Having described the undesirable system-bath interactions, we will bound g via Fermi's Golden Rule, Eq. (29). Let $\Gamma_{fi} \sim 1/\tau_{\text{high-ord.}}$ now denote the rate at which order-g^L interactions occur. The bath DOS remains $\mu_{\text{bath}}(E_{if}) \sim \frac{1}{W_{\text{b}}}$. Let us estimate the matrix-element size $|\langle f|V|i\rangle|$. The bath flips each spin at a rate g (modulo a contribution from the bath's DOS). Flipping one spin costs an amount $\sim \mathcal{E}$ of energy, on average. [$\mathcal{E}$ denotes the per-site energy density, as illustrated in Eq. (32).] Hence L spins are flipped at a rate $\sim \mathcal{E}(\frac{g}{\mathcal{E}})^L$. The initial $\mathcal{E}$ is included for dimensionality. We substitute into Fermi's Golden Rule [Eq. (29)], then solve for the time:

$$\tau_{\text{high-ord.}} \sim \frac{W_{\text{b}} \mathcal{E}^{2(L-1)}}{g^{2L}} \quad \text{wherein} \quad L = \max\left\{\frac{\langle \delta \rangle}{W_{\text{b}}}, \xi_>\right\}. \tag{C2}$$

We substitute from Eqs. (C2) and (30) into Eq. (C1). Solving for the coupling yields

$$g \ll \mathcal{E} \cdot \left(\frac{\delta_-}{\mathcal{E}}\right)^{1/(L-1)} \quad \text{wherein} \quad L = \max\left\{\frac{\langle \delta \rangle}{W_{\text{b}}}, \xi_>\right\}. \tag{C3}$$

Substituting back into Eq. (30) yields a second bound on τ_{th}:

$$\tau_{\text{th}} \gg \frac{W_{\text{b}}}{\delta_-^2}\left(\frac{\mathcal{E}}{\delta_-}\right)^{1/(L-1)}, \quad \text{wherein} \quad L = \max\left\{\frac{\langle \delta \rangle}{W_{\text{b}}}, \xi_>\right\}. \tag{C4}$$

Let us express the bound in terms of localization lengths. We set $W_{\text{b}} \sim \frac{\langle \delta \rangle}{10}$ and approximate $L \pm 1 \sim L \sim \xi_>$. We substitute in for $\langle \delta \rangle$ from Eq. (B8) and for δ_- from Eq. (B7):

$$\tau_{\text{th}} \gg \frac{1}{10\mathcal{E}} e^{2\xi_>/\xi_<} 2^{2\xi_>}. \tag{C5}$$

This inequality is looser than Eq. (31): The no-higher-order-processes condition is less demanding than Markovianity.

APPENDIX D: NUMERICAL SIMULATIONS OF THE MBL OTTO ENGINE

We simulated one 12-site mesoscale engine at half-filling. (We also studied other system sizes, to gauge finite-size effects.) Our code is available in in Ref. [76]. The random-field Heisenberg Hamiltonian (32) governed the system. We will drop the subscript from $H_{\text{sim}}(t)$.

Call the times at which the strokes end $t = \tau, \tau', \tau''$, and τ''' (see Fig. 2). For each of $N_{\text{reals}} \approx 1000$ disorder realizations, we computed the whole density matrix $\rho(t)$ at $t = 0, \tau, \tau', \tau'', \tau'''$. (See Appendixes D 3 a and D 4 for an explanation of how.) The engine's time-t internal energy is $E(t) = \text{Tr}(H(t)\rho(t))$. The quantities of interest are straightforwardly

$$\langle W_1 \rangle = E(0) - E(\tau), \quad \langle W_3 \rangle = E(\tau''') - E(\tau''), \tag{D1}$$

$$\langle Q_2 \rangle = E(\tau'') - E(\tau'), \quad \text{and} \quad \langle Q_4 \rangle = E(0) - E(\tau'''). \tag{D2}$$

We disorder-averaged these quantities before dividing to compute the efficiency, $\eta_{\text{MBL}} = 1 - \frac{\langle W_1 \rangle + \langle W_3 \rangle}{\langle Q_2 \rangle}$.

1. Scaling factor

We wish to keep the DOS constant through the cycle. To fix $\mu(E)$, we rescale the Hamiltonian by a factor $Q(h(\alpha_t))$. We define $Q^2(h(\alpha_t))$ as the disorder average of the variance of the unrescaled DOS:

$$Q^2(h(\alpha_t)) := \left\langle \left(\frac{1}{\mathcal{N}}\sum_{j=1}^{\mathcal{N}} E_j^2\right) - \left(\frac{1}{\mathcal{N}}\sum_{j=1}^{\mathcal{N}} E_j\right)^2\right\rangle_{\text{disorder}} = \left\langle \frac{1}{\mathcal{N}}\text{Tr}(\tilde{H}^2(t)) - \left(\frac{1}{\mathcal{N}}\text{Tr}(\tilde{H}(t))\right)^2\right\rangle_{\text{disorder}}. \tag{D3}$$

The $\tilde{H}(t)$ denotes an unrescaled variation on the random-field Heisenberg Hamiltonian $H(t)$ of Eq. (32):

$$\tilde{H}(t) := \mathcal{E}\left[\sum_{j=1}^{N-1} \boldsymbol{\sigma}_j \cdot \boldsymbol{\sigma}_{j+1} + h(\alpha_t)\sum_{j=1}^{N} h_j \sigma_j^z\right]. \tag{D4}$$

To compute $Q^2(h(\alpha_t))$, we rewrite the unrescaled Hamiltonian as

$$\tilde{H}(t) = \mathcal{E}\left[2\sum_{j=1}^{N-1}(\sigma_j^+\sigma_{j+1}^- + \text{H.c.}) + \sum_{j=1}^{N-1}\sigma_j^z\sigma_{j+1}^z + h(\alpha_t)\sum_{j=1}^{N} h_j\sigma_j^z\right]. \tag{D5}$$

We assume that N is even, and we work at half-filling. The $\frac{N}{2}$-particle subspace has dimensionality $\mathcal{N} = \binom{N}{N/2}$.

Let us calculate some operator traces that we will invoke later. Let $X := \prod_{j=1}^{N} \sigma^x$ denote the global spin-flip operator. For any operator A such that $X^\dagger A X = -A$,

$$\mathrm{Tr}(A) = \mathrm{Tr}(X^\dagger A X) = -\mathrm{Tr}(A). \tag{D6}$$

We have used the evenness of N, which implies the invariance of the half-filling subspace under X. Also, $\mathrm{Tr}(A) = 0$. In particular, $0 = \mathrm{Tr}(\sigma_j^z) = \mathrm{Tr}(\sigma_j^z \sigma_{j'}^z \sigma_{j''}^z)$, if $j \neq j' \neq j''$.

Traces of products of even numbers of σ^z factors require more thought:

$$\mathrm{Tr}\left(\sigma_j^z \sigma_{j+1}^z\right) = (\text{\# states } j, j+1 = \uparrow\uparrow) + (\text{\# states } j, j+1 = \downarrow\downarrow) - 2(\text{\# states } j, j+1 = \uparrow\downarrow)$$

$$= \binom{N-2}{N/2-2} + \binom{N-2}{N/2} - 2\binom{N-2}{N/2-1}$$

$$= -\mathcal{N} \frac{1}{N-1}. \tag{D7}$$

Similarly,

$$\mathrm{Tr}([\sigma_j^+\sigma_j^-][\sigma_{j+1}^-\sigma_{j+1}^+]) = \mathrm{Tr}([\sigma_j^-\sigma_j^+][\sigma_{j+1}^+\sigma_{j+1}^-]) = (\text{\# states } j, j+1 = \uparrow\downarrow) = \binom{N-2}{N/2-1} \tag{D8}$$

$$= \mathcal{N}\frac{N}{4(L-1)}, \tag{D9}$$

and

$$\mathrm{Tr}\left(\sigma_j^z \sigma_{j+1}^z \sigma_{j'}^z \sigma_{j'+1}^z\right) = (\text{\# states } j, j+1, j', j'+1 = \uparrow\uparrow\uparrow\uparrow) + \binom{4}{2}(\text{\# states } j, j+1, j', j'+1 = \uparrow\uparrow\downarrow\downarrow)$$

$$+ (\text{\# states } j, j+1, j', j'+1 = \downarrow\downarrow\downarrow\downarrow)$$

$$- \binom{4}{1}(\text{\# states } j, j+1, j', j'+1 = \uparrow\uparrow\uparrow\downarrow) - \binom{4}{1}(\text{\# states } j, j+1, j', j'+1 = \uparrow\downarrow\downarrow\downarrow)$$

$$= \binom{N-4}{N/2-4} + 6\binom{N-4}{N/2-2} + \binom{N-4}{N/2} - 6\binom{N-4}{N/2-3} - 6\binom{N-4}{N/2-1}$$

$$= \mathcal{N}\frac{3}{(N-1)(N-3)}, \tag{D10}$$

wherein the first equality's combinatorial factors come from permutations on sites j, $j+1$, j', and $j'+1$.

Assembling these pieces, we find $\mathrm{Tr}(\tilde{H}(t)) = \mathcal{E} \sum_{j=1}^{N-1} \mathrm{Tr}(\sigma_j^z \sigma_j^z) = -\mathcal{E}\mathcal{N}$. Next, we compute $\mathrm{Tr}(\tilde{H}^2(t))$:

$$\tilde{H}^2(t) = \mathcal{E}^2 \left[4 \sum_{j}^{N-1} (\sigma_j^+\sigma_j^-)(\sigma_{j+1}^-\sigma_{j+1}^+) + 4 \sum_{j}^{N-1} (\sigma_j^-\sigma_j^+)(\sigma_{j+1}^+\sigma_{j+1}^-) \right.$$

$$\left. + \sum_{j,j'=1}^{N-1} \sigma_j^z \sigma_{j+1}^z \sigma_{j'}^z \sigma_{j'+1}^z + h^2(\alpha_t) \sum_{j=1}^{N} h_j^2 + (\text{traceless terms}) \right] \tag{D11}$$

$$= \mathcal{E}^2 \left[4 \sum_{j}^{N-1} (\sigma_j^+\sigma_j^-)(\sigma_{j+1}^-\sigma_{j+1}^+) + 4 \sum_{j}^{N-1} (\sigma_j^-\sigma_j^+)(\sigma_{j+1}^+\sigma_{j+1}^-) + \sum_{j=1}^{N-1} \mathbb{1} + \sum_{j=1}^{N-2} \sigma_j^z \sigma_{j+2}^z \right.$$

$$\left. + \sum_{j=1}^{N-3} \sum_{j'=j+2}^{N-1} \sigma_j^z \sigma_{j+1}^z \sigma_{j'}^z \sigma_{j'+1}^z + h(\alpha_t)^2(\alpha_t) \sum_{j=1}^{N} h_j^2 + (\text{traceless terms}) \right]. \tag{D12}$$

We take the trace, using Eqs. (D7), (D8), and (D10):

$$\mathrm{Tr}(\tilde{H}^2(t)) = \mathcal{N}\left[3N - 1 + \frac{N-2}{N-1} + h^2 \sum_{j=1}^{N} h_j^2 \right]. \tag{D13}$$

We disorder-average by taking $h_j^2 \mapsto \int_0^1 dh_j h_j^2 = \frac{1}{3}$:

$$\langle \text{Tr}(H^2(t)) \rangle_{\text{disorder}} = \mathcal{N} \left[3N - 1 + \frac{N-2}{N-1} + N \frac{h^2}{3} \right]. \tag{D14}$$

Substituting into Eq. (D3), we infer the rescaling factor's square:

$$Q^2(h(\alpha_t)) = 3N - 2 + \frac{N-2}{N-1} + N \frac{h^2}{3}. \tag{D15}$$

Our results are insensitive to the details of Q. The width of the DOS in one disorder realization will differ from the disorder average (D15). Moreover, that difference will vary as we tune $h(\alpha_t)$, because the disorder affects only one term. The agreement between the analytics, in which $\mu(E)$ is assumed to remain constant in t, and the numerics is therefore comforting: The engine is robust against small variations in the rescaling.

2. Representing states and Hamiltonians

We structured our software to facilitate a possible extension: the cold bath might be modeled more realistically, as coupling to the engine only locally.

We represent the state of one mesoscopic MBL Otto engine with a density matrix $\rho \in \mathbb{C}^{\mathcal{N} \times \mathcal{N}}$, and the Hamiltonian with a matrix $H \in \mathbb{C}^{\mathcal{N} \times \mathcal{N}}$, relative to the basis $\{|s_1\rangle, \ldots, |s_{\mathcal{N}}\rangle\} = \{|\uparrow \ldots \uparrow\rangle, \ldots, |\downarrow \ldots \downarrow\rangle\}$ of products of σ^z eigenstates. We track the whole density matrix, rather than just the energy-diagonal elements, with an eye toward the coherent superpositions that diabatic corrections create. For an N-site chain at half-filling, $\mathcal{N} = \binom{N}{N/2} \simeq \sqrt{\frac{2}{\pi N}} \, 2^N$.

3. Strokes 1 and 3: tuning

Simulating diabatic evolution requires a different strategy from simulating adiabatic evolution. We describe the latter in Appendix D 3 a and the former in Appendix D 3 b.

a. Adiabatic evolution

The (l, m) entry of the initial-state density matrix is

$$\rho(0)_{lm} = \langle s_l | \frac{1}{Z} e^{-\beta_H H(0)} | s_m \rangle$$
$$= \frac{1}{Z} \sum_j e^{-\beta_H E_j(0)} \langle s_l | E_j(0) \rangle \langle E_j(0) | s_m \rangle. \tag{D16}$$

The jth eigenstate of $H(0)$, associated with energy $E_j(0)$, is denoted by $|E_j(0)\rangle$. We approximate the time evolution from 0 to τ (during stroke 1) as adiabatic. The evolution therefore does not move weight between levels:

$$\rho(\tau)_{lm} = \frac{1}{Z} \sum_j e^{-\beta_H E_j(0)} \langle s_l | E_j(\tau) \rangle \langle E_j(\tau) | s_m \rangle. \tag{D17}$$

If we represented our density matrix relative to an instantaneous energy eigenbasis, simulating the time evolution would be trivial: we would reinterpret the diagonal matrix ρ as being diagonal, with the same elements in a new basis. However, we wish to represent $\rho(t)$ relative to the σ_j^z product basis. This representation enhances the code's flexibility, facilitating the inclusion of diabatic evolutions and a more detailed model of cold thermalization. To represent $\rho(t)$ relative to the σ_j^z

product basis, we note that

$$\rho(\tau)_{lm} = \sum_j \langle s_l | E_j(\tau) \rangle \langle E_j(0) | \rho(0) | E_j(0) \rangle \langle E_j(\tau) | s_m \rangle$$
$$= [U(\tau, 0) \rho(0) U(\tau, 0)^\dagger]_{lm}. \tag{D18}$$

We have defined a time-evolution matrix $U(\tau, 0) \in \mathbb{C}^{\mathcal{N} \times \mathcal{N}}$ by $U(\tau, 0)_{lm} = \sum_j \langle s_l | E_j(\tau) \rangle \langle E_j(0) | s_m \rangle$. This matrix is easily computed via exact diagonalization of $H(0)$ and $H(\tau)$.

We can compute the density matrix $\rho(\tau'')$ at the end of stroke 3 (the tuning from MBL to GOE) from the density matrix $\rho(\tau')$ at the end of stroke 2 (the cold-bath thermalization) similarly: $\rho(\tau'') = U(\tau'', \tau') \rho(\tau') U(\tau'', \tau')^\dagger$. The time-evolution matrix $U(\tau'', \tau') \in \mathbb{C}^{\mathcal{N} \times \mathcal{N}}$ is given by $U(\tau'', \tau')_{lm} = \sum_j \langle s_l | E_j(0) \rangle \langle E_j(\tau) | s_m \rangle$. [Recall that $H(\tau'') = H(0)$ and $H(\tau') = H(\tau)$.]

b. Diabatic (finite-time) evolution

We simulate a stepwise tuning, taking

$$\alpha(t) = \frac{\delta t \lfloor t/\delta t \rfloor}{T}, \tag{D19}$$

wherein δt denotes a time-step size and $T \propto (h_{\text{MBL}} - h_{\text{GOE}})/v$ denotes the total tuning time. To do this, we compute a time-evolution unitary for the whole stroke by chaining together the unitaries for each time step. For stroke 1,

$$U(\tau, 0; v, \delta t) = e^{-iH(\tau - \delta t)\delta t} e^{-iH(\tau - 2\delta t)\delta t} \ldots e^{-iH(0)\delta t}, \tag{D20}$$

with the number of time steps set by the speed. We use the time step $\delta t = 0.405 \langle \delta \rangle$, but our results are not sensitive to the time-step size.

In judging the engine's effectiveness at a finite v, we must estimate the level-repulsion scale δ_-. We do this by diagonalizing 10^6 disorder realizations at the relevant disorder width, $h = 20$, for $N = 8$ sites. A histogram of the gaps is plotted in Fig. 11. We then visually estimate the point at which the distribution turns over. Our results are not sensitive to this value.

4. Stroke 2: Thermalization with the cold bath

During stroke 2, the system thermalizes with a bandwidth-W_b cold bath. We make three assumptions. First, the bandwidth cutoff is hard: The bath can transfer only amounts $< W_b$
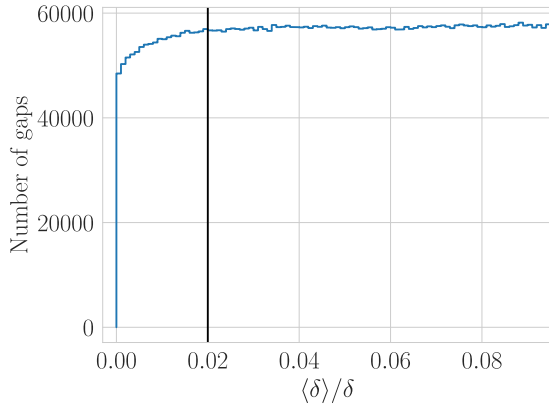
FIG. 11. Level-spacing distribution for 10^6 disorder realizations of the random-field Heisenberg model at disorder width $h = 20$ and system size $N = 8$ (blue line). The vertical black line shows the estimate of the level-repulsion parameter δ_-.

of energy at a time. Therefore, the cold bath cannot move probability mass between adjacent levels separated by just one gap $\delta' > W_b$. Second, the bath is Markovian. Third, the system thermalizes for a long time. The bath has time to move weight across sequences of small gaps $\delta'_j, \delta'_{j+1}, \ldots < W_b$.

We can implement thermalization as follows. First, we identify sequences of levels connected by small gaps. Second, we reapportion weight amongst the levels according to a Gibbs distribution.

Suppose, for example, that the MBL Hamiltonian $H(\tau)$ contains the following chain of six energies, $E_1, \ldots, E_6$,
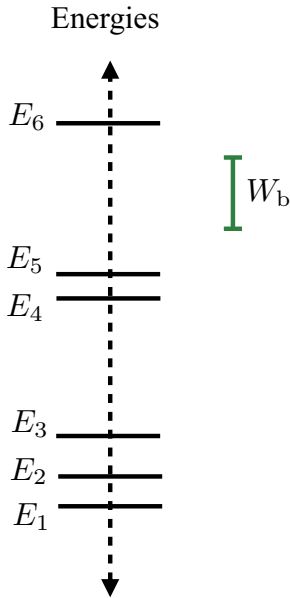

FIG. 12. Energies of a cold-thermalized many-body-localized system. We illustrate our implementation of cold thermalization with this example chain of six energies. The cold bath has a bandwidth of size W_b, depicted in green.

separated from its surrounding levels by large gaps (Fig. 12):

$$(E_2 - E_1), (E_3 - E_2) < W_b, \quad (E_5 - E_4) < W_b,$$

and $(E_4 - E_3), (E_6 - E_5) > W_b. \qquad (D21)$

We suppress the time arguments to simplify notation. Before thermalization, the density operator is diagonal with respect to the energy basis: $\rho(\tau) = \sum_j \rho_j |E_j\rangle\langle E_j|$. The weight on level j is denoted by ρ_j. Thermalization maps

$$\rho(\tau) \mapsto \rho(\tau')$$
$$= \frac{\rho_1 + \rho_2 + \rho_3}{e^{-\beta_c E_1} + e^{-\beta_c E_2} + e^{-\beta_c E_3}}$$
$$\times (e^{-\beta_c E_1}|E_1\rangle\langle E_1| + e^{-\beta_c E_2}|E_2\rangle\langle E_2| + e^{-\beta_c E_3}|E_3\rangle\langle E_3|)$$
$$+ \frac{\rho_4 + \rho_5}{e^{-\beta_c E_4} + e^{-\beta_c E_5}}(e^{-\beta_c E_4}|E_4\rangle\langle E_4| + e^{-\beta_c E_5}|E_5\rangle\langle E_5|)$$
$$+ \rho_6|E_6\rangle\langle E_6|. \qquad (D22)$$

APPENDIX E: COMPARISONS WITH COMPETITOR OTTO ENGINES

This appendix contains more analysis of the bandwidth engine (Appendix E 1) and of an MBL engine tuned between equal-strength disorder realizations (Appendix E 2). Appendix E 2 compares with an MBL engine thermalized with an ordinary-bandwidth cold bath. The quantum-dot and Anderson-localized engines are elaborated on in Appendices E 3 and E 4.

1. Comparison with bandwidth engine

Imagine eliminating the scaling factor $Q(h(\alpha_t))$ from the Hamiltonian (32). The energy band is compressed and expanded as the disorder strength $h(\alpha_t)$ is ramped down and up. The whole band, rather than a gap, contracts and widens as in Fig. 3, between a size $\sim \mathcal{E} N_{\text{macro}} h(\alpha_0)$ and a size $\sim \mathcal{E} N_{\text{macro}} h(\alpha_1) \gg \mathcal{E} N_{\text{macro}} h(\alpha_0)$. The engine can remain in one phase throughout the cycle. The cycle does not benefit from the "athermality" of local level correlations.

Furthermore, this accordionlike motion requires no change of the energy eigenbasis's form. Tuning may proceed quantum-adiabatically: $v \approx 0$. The ideal engine suffers no diabatic jumps, losing $\langle W_{\text{diab}} \rangle_{\text{macro}} = 0$.

However, this engine is impractical: Consider any perturbation V that fails to commute with the ideal Hamiltonian $H(t)$: $[V, H(t)] \neq 0$. Stray fields, for example, can taint an environment. As another example, consider ultracold atoms in an optical lattice. The disorder strength is ideally $\mathcal{E} h(\alpha_t)$. One can strengthen the disorder by strengthening the lattice potential U_{lattice}. Similarly, one can raise the hopping frequency (ideally $\mathcal{E}$) by raising the pressure p. Strengthening U_{lattice} and p while achieving the ideal disorder-to-hopping ratio $\frac{\mathcal{E} h(\alpha_t)}{\mathcal{E}} = h(\alpha_t)$ requires fine control. If the ratio changes from $h(\alpha_t)$, the Hamiltonian $H(t)$ acquires a perturbation V that fails to commute with other terms.

This V can cause diabatic jumps that cost work $\langle W_{\text{diab}}\rangle_{\text{macro}}$. Can the bandwidth engine not withstand several hops—say, through $0.02\mathcal{N}_{\text{macro}}$ levels?

No, because the ground state pulls away from the rest of the spectrum as N_{macro} grows. Suppose, for simplicity, that $T_{\text{C}} = 0$ and $T_{\text{H}} = \infty$. The bandwidth engine starts stroke 1 in $\rho(0) = \mathbb{1}/\mathcal{N}_{\text{macro}}$. Diabatic hops preserve $\rho(t)$ during stroke 1, on average: the engine as likely hops upward as drops. Cold thermalization drops the engine to the ground state (plus an exponentially small dusting of higher-level states). The ground-state energy is generically extensive. Hence the engine absorbs $\langle Q_2\rangle_{\text{macro}} \sim -N_{\text{macro}}$, on average. Suppose that, during stroke 3, the engine jumps up through 2% of the levels. The engine ends about two standard deviations below the spectrum's center, with average energy $\sim\sqrt{N_{\text{macro}}}$. While returning to $T_{\text{H}} = 0$ during the average stroke 4, the bandwidth engine absorbs $\langle Q_4\rangle_{\text{macro}} \sim \sqrt{N_{\text{macro}}}$. The average outputted work $\langle W_{\text{tot}}\rangle_{\text{macro}} = \langle Q_4\rangle_{\text{macro}} + \langle Q_2\rangle_{\text{macro}} \sim \sqrt{N_{\text{macro}}} - N_{\text{macro}}$. As N_{macro} grows, $\langle W_{\text{tot}}\rangle_{\text{macro}}$ dips farther below zero. A few diabatic jumps threaten the bandwidth engine's ability to output $\langle W_{\text{tot}}\rangle > 0$.

The bandwidth engine's v must decline as N_{macro} grows also because the typical whole-system gap $\langle\delta\rangle_{\text{macro}} \sim \frac{\mathcal{E}}{\mathcal{N}_{\text{macro}}}$ shrinks. The smaller the gaps, the greater the likelihood that a given v induces hops. As $\langle\delta\rangle_{\text{macro}} \to 0$, v must $\to 0$. The MBL Otto cycle proceeds more quickly, due to subengines' parallelization.

2. Comparison with MBL engine tuned between same-strength disorder realizations

Take our MBL Otto cycle, and vary not the disorder strength, but the disorder realization during each cycle. The disorder strength $h(\alpha_t)$ in Eq. (32) would remain $\gg 1$ and constant in t, while the random variables h_j would change. Let $\tilde{S}$ denote this constant-$h(\alpha_t)$ engine, and let S denote the MBL engine. $\tilde{S}$ takes less advantage of MBL's "athermality," as $\tilde{S}$ is not tuned between level-repelling and level-repulsion-free regimes.

Yet $\tilde{S}$ outputs the amount $\langle W_{\text{tot}}\rangle$ of work outputted by S per cycle, on average. Because W_{b} is small, cold thermalization drops $\tilde{S}$ across only small gaps $\delta' \ll \langle\delta\rangle$. $\tilde{S}$ traverses a trapezoid, as in Fig. 3, in each trial. However, the MBL engine has two advantages: greater reliability and fewer worst-case (negative-work-outputted) trials.

Both the left-hand gap δ and the right-hand gap δ' traversed by $\tilde{S}$ are Poisson-distributed. Poisson-distributed gaps more likely assume extreme values than GOE-distributed gaps: $P_{\text{MBL}}^{(E)}(\delta) > P_{\text{GOE}}^{(E)}(\delta)$ if $\delta \sim 0$ or $\delta \gg \langle\delta\rangle$ [46]. The left-hand gap δ traversed by S is GOE-distributed. Hence the W_{tot} outputted by $\tilde{S}$ more likely assumes extreme values than the W_{tot} outputted by S. The greater reliability of S may suit S better to "one-shot statistical mechanics" [17,18,20,21,23,24,105–110]. In one-shot theory, predictability of the work W_{tot} extractable in any given trial serves as a resource.

Additionally, S suffers fewer worst-case trials than $\tilde{S}$. We define as *worst-case* a trial in which the engine outputs net negative work, $W_{\text{tot}} < 0$. Consider again Fig. 3. Consider a similar figure that depicts the trapezoid traversed by $\tilde{S}$ in some trial. The left-hand gap, δ, is distributed as the right-hand gap, δ', is, according to $P_{\text{MBL}}^{(E)}(\delta)$. Hence δ has a decent chance of being smaller than δ': $\delta < \delta'$. $\tilde{S}$ would output $W_{\text{tot}} < 0$ in such a trial.

Suppose, for simplicity, that $T_{\text{H}} = \infty$ and $T_{\text{C}} = 0$. The probability that any given S trial outputs $W_{\text{tot}} < 0$ is

$$p_{\text{worst}} \approx (\text{Prob. that the left-hand gap} < \text{the right-hand gap})$$

$$\times (\text{Prob. that the right-hand gap is small enough to be cold-thermalized}) \tag{E1}$$

$$\approx (\text{Prob. that the left-hand gap} < W_{\text{b}}) \times \frac{W_{\text{b}}}{\langle\delta\rangle}. \tag{E2}$$

The initial factor is modeled by the area of a region under the $P_{\text{GOE}}^{(E)}(\delta)$ curve. The region stretches from $\delta = 0$ to $\delta = W_{\text{b}}$. We approximate the region as a triangle of length W_{b} and height $\frac{\pi}{2} \frac{W_{\text{b}}}{\langle\delta\rangle^2} e^{-\frac{\pi}{4}(W_{\text{b}})^2/\langle\delta\rangle^2} \sim \frac{W_{\text{b}}}{\langle\delta\rangle^2}$, [$\delta \approx W_{\text{b}}$, Eq. (2), and $\frac{W_{\text{b}}}{\langle\delta\rangle} \ll 1$]. The triangle has an area of $\frac{1}{2} \times W_{\text{b}} \times \frac{\pi}{2} \frac{W_{\text{b}}}{\langle\delta\rangle^2} \sim (\frac{W_{\text{b}}}{\langle\delta\rangle})^2$. Substituting into Eq. (E2) yields

$$p_{\text{worst}} \sim \left(\frac{W_{\text{b}}}{\langle\delta\rangle}\right)^3. \tag{E3}$$

Let $\tilde{p}_{\text{worst}}$ denote the probability that any given $\tilde{S}$ trial outputs $W_{\text{tot}} < 0$. $\tilde{p}_{\text{worst}}$ shares the form of Eq. (E2). The initial factor approximates to the area of a region under the $P_{\text{MBL}}^{(E)}(\delta)$ curve. The region extends from $\delta = 0$ to $\delta = W_{\text{b}}$. The region resembles a rectangle of height $P_{\text{MBL}}^{(E)}(0) \approx \frac{1}{\langle\delta\rangle}$.

Combining the rectangle's area, $\frac{W_{\text{b}}}{\langle\delta\rangle}$, with Eq. (E2) yields

$$\tilde{p}_{\text{worst}} \sim \left(\frac{W_{\text{b}}}{\langle\delta\rangle}\right)^2. \tag{E4}$$

Since $\frac{W_{\text{b}}}{\langle\delta\rangle} \ll 1$, $p_{\text{worst}} \ll \tilde{p}_{\text{worst}}$.[11]

3. Quantum-dot engine

Section VII introduced the quantum-dot engine, an array of ideally independent bits or qubits. We add to the order-of-magnitude analysis two points about implementations' practicality. First, the MBL potential's generic nature offers an advantage. MBL requires a random disorder potential $\{h(\alpha_t)h_j\}$,

---

[11]The discrepancy is exaggerated if the exponent in Eq. (E3) rises, if the left-hand S Hamiltonian is modeled with a Gaussian ensemble other than the GOE. The Gaussian unitary ensemble (GUE) contains an exponent of 4; the Gaussian symplectic ensemble (GSE), an exponent of 6. Different ensembles model different symmetries.

e.g., a "dirty sample," a defect-riddled crystal. This "generic" potential contrasts with the pristine background required by quantum dots. Imposing random MBL disorder is expected to be simpler. On the other hand, a quantum-dot engine does not necessarily need a small-bandwidth cold bath, $W_{\mathrm{b}} \ll \langle \delta \rangle$.

4. Anderson-localized engine

Anderson localization follows from removing the interactions from MBL (Appendix B). One could implement our Otto cycle with an Anderson insulator because Anderson Hamiltonians exhibit Poissonian level statistics (1). But strokes 1 and 3 would require the switching off and on of interactions. Tuning the interaction, as well as the disorder-to-interaction ratio, requires more effort than tuning just the latter.

Also, particles typically interact in many-body systems. MBL particles interact; Anderson-localized particles do not. Hence one might eventually expect less difficulty in engineering MBL engines than in engineering Anderson-localized engines.

[1] D. A. Huse, R. Nandkishore, and V. Oganesyan, Phys. Rev. B **90**, 174202 (2014).

[2] J. A. Kjäll, J. H. Bardarson, and F. Pollmann, Phys. Rev. Lett. **113**, 107204 (2014).

[3] M. Schreiber *et al.*, Science **349**, 842 (2015).

[4] S. S. Kondov, W. R. McGehee, W. Xu, and B. DeMarco, Phys. Rev. Lett. **114**, 083002 (2015).

[5] M. Ovadia *et al.*, Sci. Rep. **5**, 13503 (2015).

[6] J.-Y. Choi *et al.*, Science **352**, 1547 (2016).

[7] H. P. Lüschen, P. Bordia, S. S. Hodgman, M. Schreiber, S. Sarkar, A. J. Daley, M. H. Fischer, E. Altman, I. Bloch, and U. Schneider, Phys. Rev. X **7**, 011034 (2017).

[8] G. Kucsko *et al.*, Phys. Rev. Lett. **121**, 023601 (2018).

[9] J. Smith *et al.*, Nat. Phys. **12**, 907 (2016).

[10] P. Bordia, H. Lüschen, S. Scherg, S. Gopalakrishnan, M. Knap, U. Schneider, and I. Bloch, Phys. Rev. X **7**, 041047 (2017).

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

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

[13] M. Rigol, V. Dunjko, V. Yurovsky, and M. Olshanii, Phys. Rev. Lett. **98**, 050405 (2007).

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

[15] R. Nandkishore and D. A. Huse, Ann. Rev. Condens. Matter Phys. **6**, 15 (2015).

[16] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, Int. J. Theor. Phys. **39**, 2717 (2000).

[17] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, New J. Phys. **13**, 053015 (2011).

[18] J. Åberg, Nat. Commun. **4**, 1925 (2013).

[19] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Phys. Rev. Lett. **111**, 250404 (2013).

[20] M. Horodecki and J. Oppenheim, Nat. Commun. **4**, 2059 (2013).

[21] D. Egloff, O. C. O. Dahlsten, R. Renner, and V. Vedral, New J. Phys. **17**, 073001 (2015).

[22] J. Goold, M. Huber, A. Riera, L. del Río, and P. Skrzypczyk, J. Phys. A: Math. Theore. **49**, 143001 (2016).

[23] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, Phys. Rep. **583**, 1 (2015).

[24] N. Y. Halpern, J. Phys. A: Math. Theore. **51**, 094001 (2018).

[25] S. Deffner, J. P. Paz, and W. H. Zurek, Phys. Rev. E **94**, 010103 (2016).

[26] H. Wilming and R. Gallego, Phys. Rev. X **7**, 041033 (2017).

[27] J. E. Geusic, E. O. Schulz-DuBios, and H. E. D. Scovil, Phys. Rev. **156**, 343 (1967).

[28] A. del Campo, J. Goold, and M. Paternostro, Sci. Rep. **4**, 6208 (2014).

[29] N. Brunner, M. Huber, N. Linden, S. Popescu, R. Silva, and P. Skrzypczyk, Phys. Rev. E **89**, 032115 (2014).

[30] F. C. Binder, S. Vinjanampathy, K. Modi, and J. Goold, New J. Phys. **17**, 075015 (2015).

[31] M. P. Woods, N. Ng, and S. Wehner, arXiv:1506.02322.

[32] D. Gelbwaser-Klimovsky and A. Aspuru-Guzik, J. Phys. Chem. Lett. **6**, 3477 (2015).

[33] Q. Song, S. Singh, K. Zhang, W. Zhang, and P. Meystre, Phys. Rev. A **94**, 063852 (2016).

[34] H. Terças, S. Ribeiro, M. Pezzutto, and Y. Omar, Phys. Rev. E **95**, 022135 (2017).

[35] M. Perarnau-Llobet, A. Riera, R. Gallego, H. Wilming, and J. Eisert, New J. Phys. **18**, 123035 (2016).

[36] R. Kosloff and Y. Rezek, Entropy **19**, 136 (2017).

[37] J. Lekscha, H. Wilming, J. Eisert, and R. Gallego, Phys. Rev. E **97**, 022142 (2018).

[38] J. Jaramillo, M. Beau, and A. del Campo, New J. Phys. **18**, 075019 (2016).

[39] D. Gelbwaser-Klimovsky, A. Bylinskii, D. Gangloff, R. Islam, A. Aspuru-Guzik, and V. Vuletic, Phys. Rev. Lett. **120**, 170601 (2018).

[40] D. Basko, I. Aleiner, and B. Altshuler, Ann. Phys. **321**, 1126 (2006).

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

[42] A. Pal and D. A. Huse, Phys. Rev. B **82**, 174411 (2010).

[43] M. Serbyn and J. E. Moore, Phys. Rev. B **93**, 041424 (2016).

[44] B. Kramer and A. MacKinnon, Rep. Prog. Phys. **56**, 1469 (1993).

[45] U. Sivan and Y. Imry, Phys. Rev. B **35**, 6074 (1987).

[46] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, Adv. Phys. **65**, 239 (2016).

[47] Y. Imry and S.-K. Ma, Phys. Rev. Lett. **35**, 1399 (1975).

[48] S. V. Syzranov, A. V. Gorshkov, and V. Galitski, Phys. Rev. B **97**, 161114(R) (2018).

[49] N. Yunger Halpern, C. D. White, S. Gopalakrishnan, and G. Refael, arXiv:1707.07008v1.

[50] D. Quattrochi, The internal combustion engine (Otto cycle), 2006, http://web.mit.edu/16.unified/www/FALL/thermodynamics/notes/node26.html.

[51] M. O. Scully, Phys. Rev. Lett. **88**, 050602 (2002).

[52] O. Abah, J. Roßnagel, G. Jacob, S. Deffner, F. Schmidt-Kaler, K. Singer, and E. Lutz, Phys. Rev. Lett. **109**, 203006 (2012).

[53] J. Deng, Q.-H. Wang, Z. Liu, P. Hänggi, and J. Gong, Phys. Rev. E **88**, 062122 (2013).

[54] Y. Zheng and D. Poletti, Phys. Rev. E **90**, 012145 (2014).

[55] B. Karimi and J. P. Pekola, Phys. Rev. B **94**, 184503 (2016).

[56] S. Vinjanampathy and J. Anders, Contemp. Phys. **57**, 545 (2016).

[57] S.-Z. Lin and S. Hayami, Phys. Rev. B **93**, 064430 (2016).

[58] P. Corboz, Phys. Rev. B **94**, 035133 (2016).

[59] S. Gopalakrishnan, M. Knap, and E. Demler, Phys. Rev. B **94**, 094201 (2016).

[60] R. Kosloff and T. Feldmann, Phys. Rev. E **65**, 055102 (2002).

[61] T. D. Kieu, Phys. Rev. Lett. **93**, 140403 (2004).

[62] R. Kosloff and T. Feldmann, Phys. Rev. E **82**, 011134 (2010).

[63] S. Çakmak, F. Altintas, A. Gençten, and Ö. E. Müstecaplıoğlu, Eur. Phys. J. D **71**, 75 (2017).

[64] C. De Grandi and A. Polkovnikov, in *Adiabatic Perturbation Theory: From Landau-Zener Problem to Quenching Through a Quantum Critical Point*, Lecture Notes in Physics, edited by A. K. K. Chandra, A. Das, and B. K. K. Chakrabarti (Berlin, Springer, Verlag, 2010), Vol. 802, p. 75.

[65] D. A. Huse, R. Nandkishore, F. Pietracaprina, V. Ros, and A. Scardicchio, Phys. Rev. B **92**, 014203 (2015).

[66] A. De Luca and A. Rosso, Phys. Rev. Lett. **115**, 080401 (2015).

[67] E. Levi, M. Heyl, I. Lesanovsky, and J. P. Garrahan, Phys. Rev. Lett. **116**, 237203 (2016).

[68] M. H. Fischer, M. Maksymenko, and E. Altman, Phys. Rev. Lett. **116**, 160401 (2016).

[69] A. V. Khaetskii, D. Loss, and L. Glazman, Phys. Rev. Lett. **88**, 186802 (2002).

[70] S. Gopalakrishnan and R. Nandkishore, Phys. Rev. B **90**, 224203 (2014).

[71] S. A. Parameswaran and S. Gopalakrishnan, Phys. Rev. B **95**, 024201 (2017).

[72] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).

[73] V. Khemani, R. Nandkishore, and S. L. Sondhi, Nat. Phys. **11**, 560 (2015).

[74] H. Kim and D. A. Huse, Phys. Rev. Lett. **111**, 127205 (2013).

[75] E. Lieb and D. Robinson, Commun. Math. Phys. **28**, 251 (1972).

[76] https://github.com/christopherdavidwhite/MBL-mobile.

[77] M. T. Brown, Bacterial flagellar motor: Biophysical studies, in *Encyclopedia of Biophysics*, edited by G. C. K. Roberts (Springer, Berlin, Heidelberg, 2013), pp. 155–155.

[78] A. Chandran, I. H. Kim, G. Vidal, and D. A. Abanin, Phys. Rev. B **91**, 085425 (2015).

[79] E. Geva and R. Kosloff, J. Chem. Phys. **96**, 3054 (1992).

[80] E. Geva and R. Kosloff, J. Chem. Phys. **97**, 4398 (1992).

[81] T. Feldmann, E. Geva, R. Kosloff, and P. Salamon, Am. J. Phys. **64**, 485 (1996).

[82] J. He, J. Chen, and B. Hua, Phys. Rev. E **65**, 036145 (2002).

[83] G. Alvarado Barrios, F. Albarrán-Arriagada, F. A. Cárdenas-López, G. Romero, and J. C. Retamal, Phys. Rev. A **96**, 052119 (2017).

[84] J. R. Petta *et al.*, Science **309**, 2180 (2005).

[85] J. R. Petta *et al.*, Physica E **34**, 42 (2006), Proceedings of the 16th International Conference on Electronic Properties of Two-Dimensional Systems (EP2DS-16).

[86] M. Campisi and R. Fazio, Nat. Commun. **7**, 11895 (2016).

[87] R. Modak and M. Rigol, Phys. Rev. E **95**, 062145 (2017).

[88] W. Verstraelen, D. Sels, and M. Wouters, Phys. Rev. A **96**, 023605 (2017).

[89] D. Ferraro, M. Campisi, G. M. Andolina, V. Pellegrini, and M. Polini, Phys. Rev. Lett. **120**, 117702 (2018).

[90] Y.-H. Ma, S.-H. Su, and C.-P. Sun, Phys. Rev. E **96**, 022143 (2017).

[91] P. Skrzypczyk, A. J. Short, and S. Popescu, arXiv:1302.2811.

[92] M. Lostaglio, D. Jennings, and T. Rudolph, Nat. Commun. **6**, 6383 (2015).

[93] M. Lostaglio, D. Jennings, and T. Rudolph, New J. Phys. **19**, 043008 (2017).

[94] N. Y. Halpern, P. Faist, J. Oppenheim, and A. Winter, Nat. Commun. **7**, 12051 (2016).

[95] Y. Guryanova, S. Popescu, A. J. Short, R. Silva, and P. Skrzypczyk, Nat. Commun. **7**, 12049 (2016).

[96] X. Chen, A. Ruschhaupt, S. Schmidt, A. del Campo, D. Guéry-Odelin, and J. G. Muga, Phys. Rev. Lett. **104**, 063002 (2010).

[97] E. Torrontegui *et al.*, Adv. At. Mol. Opt. Phys. **62**, 117 (2013).

[98] O. Abah and E. Lutz, Phys. Rev. E **98**, 032121 (2018).

[99] M. Ziman *et al.*, arXiv:quant-ph/0110164.

[100] V. Scarani, M. Ziman, P. Štelmachovič, N. Gisin, and V. Bužek, Phys. Rev. Lett. **88**, 097905 (2002).

[101] S. Shevchenko, S. Ashhab, and F. Nori, Phys. Rep. **492**, 1 (2010).

[102] N. Yunger Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, New J. Phys. **17**, 095003 (2015).

[103] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[104] S. Gopalakrishnan, M. Müller, V. Khemani, M. Knap, E. Demler, and D. A. Huse, Phys. Rev. B **92**, 104202 (2015).

[105] L. del Río, J. Aberg, R. Renner, O. Dahlsten, and V. Vedral, Nature (London) **474**, 61 (2011).

[106] O. C. O. Dahlsten, Entropy **15**, 5346 (2013).

[107] F. Brandão, M. Horodecki, Woods, N. Ng, J. Oppenheim, and S. Wehner, PNAS **112**, 3275 (2015).

[108] G. Gour, Phys. Rev. A **95**, 062314 (2017).

[109] K. Ito and M. Hayashi, Phys. Rev. E **97**, 012129 (2018).

[110] R. van der Meer, N. H. Y. Ng, and S. Wehner, Phys. Rev. A **96**, 062135 (2017).