

2019-01-09

# Asymptotic prethermalization in periodically driven classical spin chains

---

Owen Howell, Phillip Weinberg, Dries Sels, Anatoli Polkovnikov, Marin Bukov. 2019.

"Asymptotic Prethermalization in Periodically Driven Classical Spin Chains." *Physical Review Letters*, Volume 122, pp. 010602 - 010602. <https://doi.org/10.1103/PhysRevLett.122.010602>

<https://hdl.handle.net/2144/35954>

*"Downloaded from OpenBU. Boston University's institutional repository."*

# Asymptotic Prethermalization in Periodically Driven Classical Spin Chains

Owen Howell,<sup>1,\*</sup> Phillip Weinberg,<sup>1</sup> Dries Sels,<sup>1,2,3</sup> Anatoli Polkovnikov,<sup>1</sup> and Marin Bukov<sup>4,†</sup>

<sup>1</sup>*Department of Physics, Boston University, 590 Commonwealth Ave., Boston, MA 02215, USA*

<sup>2</sup>*Department of Physics, Harvard University, 17 Oxford st., Cambridge, MA 02138, USA*

<sup>3</sup>*Theory of quantum and complex systems, Universiteit Antwerpen, B-2610 Antwerpen, Belgium*

<sup>4</sup>*Department of Physics, University of California, Berkeley, CA 94720, USA*

(Dated: January 21, 2019)

We reveal a continuous dynamical heating transition between a prethermal and an infinite-temperature stage in a clean, chaotic periodically driven classical spin chain. The transition time is a steep exponential function of the drive frequency, showing that the exponentially long-lived prethermal plateau, originally observed in quantum Floquet systems, survives the classical limit. Even though there is no straightforward generalization of Floquet's theorem to nonlinear systems, we present strong evidence that the prethermal physics is well described by the inverse-frequency expansion. We relate the stability and robustness of the prethermal plateau to drive-induced synchronization not captured by the expansion. Our results set the pathway to transfer the ideas of Floquet engineering to classical many-body systems, and are directly relevant for photonic crystals and cold atom experiments in the superfluid regime.

Periodically-driven systems are currently experiencing an unprecedented revival of interest through theoretical and experimental design of novel states of matter. Commonly known as *Floquet engineering*, this approach has enjoyed success in the regime of high driving frequency, where it has been appreciated as a useful tool to ascribe novel properties to otherwise trivial static Hamiltonians [1–3]. Prominent examples include the Kapitza pendulum [4], cold-atom realisations of topological [5–12] and spin-dependent [13] bands, artificial gauge fields [14–22], spin-orbit coupling [23, 24], enhanced magnetic correlations [25], synthetic dimensions [26–28], and photonic topological insulators [29–31].

The applicability of Floquet engineering requires the ability to prepare the periodically driven system in the corresponding Floquet state [32–35], and the stability of the system to detrimental heating [36–41]. Presenting a major bottleneck at the forefront of present-date experimental research, heating processes play an important role in many-body Floquet systems, and understanding the underlying physics is expected to offer significant advances in the field. Unlike single-particle quantum systems, such as the kicked rotor [42] and weakly-interacting bosonic models [43, 44], it is believed that generic isolated clean periodically-driven quantum many-body systems heat up to an infinite-temperature state [45–50], although the debate is not fully settled [51–56]. Heating rates have been shown to be at least exponentially suppressed in the drive frequency [57, 58].

In this paper, we present a numerical study of thermalisation in a clean, globally-driven, isolated *classical* spin chain, reaching times beyond the astronomical  $10^{10}$  driving cycles. We find that, the dynamics falls into four stages, see Fig. 1: an initial transient (i) during which the

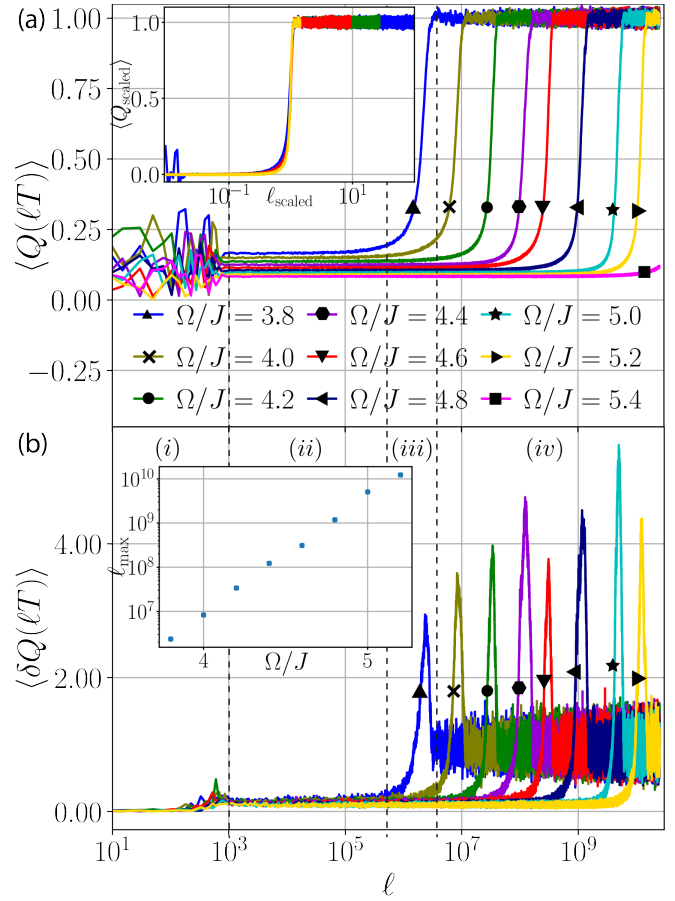


FIG. 1: Noise-averaged energy (a) and energy variance (b) as a function of the number of driving cycles  $\ell$ . Insets: rescaled energy curves (a) by the position of the peak in the variance curves (b), reveal a dynamical heating transition from a prethermal stage to an infinite-temperature stage in the limit  $\Omega, \ell \rightarrow \infty$ . The dashed vertical lines mark the four stages of evolution for  $\Omega/J = 3.8$  [see text]. See Fig. 2 for parameters.

\*Electronic address: [ohl20@bu.edu](mailto:ohl20@bu.edu)

†Electronic address: [mgbukov@berkeley.edu](mailto:mgbukov@berkeley.edu)

system exhibits *constrained* thermalization to a finite energy density set by the initial ensemble, (ii) a frequency-controlled long-lived prethermal plateau [57, 59–61] ideally suited for Floquet engineering, followed by (iii) a late crossover governed by *unconstrained* thermalisation to (iv) a featureless infinite-temperature state. Our analysis reveals the existence of a dynamical heating transition between a prethermal plateau and an infinite-temperature stage at infinite times and infinite drive frequencies. The prethermal plateau is described by the inverse-frequency expansion originally developed for quantum systems, with improving agreement at increasing first few orders of the expansion [62]. This allows to transfer the machinery of Floquet engineering directly to classical many-body models despite the absence of a Floquet theorem for systems governed by nonlinear equations of motion. Focusing on the prethermal plateau, we demonstrate that a key role for its exponentially long duration is played by drive-induced synchronization. In particular, stopping the periodic drive and then restarting it de-synchronizes the system and strongly increases the heating rate.

We report that the time scale for a classical system to leave a small corner of phase space around the initial state and heat up to an infinite-temperature state, scales exponentially with the driving frequency [Fig. 1]. A related problem on ergodicity in classical systems was addressed in a pioneering numerical study by Fermi, Pasta, Ulam and Tsingou, and gave first evidence for a parametrically slow thermalisation in a system of coupled classical oscillators [63, 64], puzzling the community for the past few decades. For systems with finite number of degrees of freedom, theorems in Hamiltonian mechanics have been proven showing that the motion of action variables in nearly-integrable systems remains confined to a small region of phase space until exponentially long times, controlled by the integrability breaking parameter [65–68]. This behaviour is accompanied by subdiffusion, as reported for a system of periodically-kicked coupled pendula [69–72]. Chaotic many-body dynamics in periodically-kicked spin chains has been studied using a classical Loschmidt echo approach [73, 74].

*Model.*—Consider a classical Ising chain with periodic boundary conditions, described by the energy function

$$H(t) = \begin{cases} \sum_{j=1}^N JS_j^z S_{j+1}^z + hS_j^z & \text{for } t \in [0, T/2] \bmod T \\ \sum_{j=1}^N gS_j^x & \text{for } t \in [T/2, T] \bmod T \end{cases}$$

where  $J$  denotes the nearest-neighbour interaction strength, while  $h$  and  $g$  are the magnetic field strengths along the  $z$  and  $x$ -directions, respectively. The spin [or rotor] variable  $\vec{S}_j$ ,  $|\vec{S}_j| = 1$ , on site  $j$  satisfies the Poisson bracket relation  $\{S_i^\mu, S_j^\nu\} = \delta_{ij} \varepsilon^{\mu\nu\rho} S_j^\rho$ , with  $\varepsilon^{\mu\nu\rho}$  the fully antisymmetric tensor.

The time dependence arises due to periodic switching of two time-independent Hamilton functions, for a duration of  $T/2$  each, with frequency  $\Omega = 2\pi/T$ . The time evolution of the system is governed by Hamilton's

EOM  $\dot{S}_j^\mu(t) = \{S_j^\mu, H(t)\}$ . Interested in the long-time thermalisation properties, we focus on stroboscopic evolution. Integrating the EOM over one total period  $T$ , the evolved state is obtained from a successive application of a discrete map  $\vec{S}_j(\ell T) = [\tau_2 \circ \tau_1]^\ell (\vec{S}_j(0))$ , with  $\ell \in \mathbb{N}$  counting the driving cycles. During the first half-period, the time evolution follows the *non-linear* rotation  $\tau_1$  about the  $z$ -axis:

$$\tau_1(\vec{S}_j) = \begin{bmatrix} S_j^x \cos(\kappa_j T/2) - S_j^y \sin(\kappa_j T/2) \\ S_j^x \sin(\kappa_j T/2) + S_j^y \cos(\kappa_j T/2) \\ S_j^z \end{bmatrix} \quad (1)$$

with spin-dependent natural frequency of rotation  $\kappa_j = J(S_{j-1}^z + S_{j+1}^z) + h$ . The dynamics in the second half-period follows the rotation  $\tau_2$  about the  $x$ -axis:

$$\tau_2(\vec{S}_j) = \begin{bmatrix} S_j^x \\ S_j^y \cos(gT/2) - S_j^z \sin(gT/2) \\ S_j^y \sin(gT/2) + S_j^z \cos(gT/2) \end{bmatrix} \quad (2)$$

The map  $\tau_2 \circ \tau_1$  is the nonlinear classical analogue of the quantum Floquet unitary.

Motivated by experiments which study Floquet-engineered ordered states at high drive frequencies, we prepare the system at time  $t = 0$  in the lowest-energy state (i.e. the ground state, GS) of the time-averaged Hamiltonian

$$H_{\text{ave}} \equiv H_F^{(0)} = \frac{1}{2} \sum_{j=1}^N JS_j^z S_{j+1}^z + hS_j^z + gS_j^x, \quad (3)$$

with energy density  $E_{\text{GS}}(h/J, g/J)/N \approx -1.235J$  for  $g/J = 0.9045$ ,  $h/J = 0.809$ . Whenever  $J, h, g$  have equal sign and are of the same order of magnitude, the GS features antiferromagnetic (AFM) order w.r.t. a direction in the  $xz$ -plane, parametrized by the azimuthal angle  $\theta$ . Making use of translational invariance, one can determine the value of  $\theta$  which minimizes the energy  $E_{\text{GS}}(\theta)$ . Translational invariance constrains the GS evolution to be uniquely described by two coupled spin degrees of freedom, corresponding to the AFM unit cell. To bring out the many-body character of the model, we add small noise to the azimuthal angle of each spin, independently drawn from a uniform distribution over  $[-\pi/100, \pi/100]$ , which breaks translational symmetry and allows for thermalisation. The quantities we consider are averaged over an ensemble of 100 noisy initial state realizations. We verified that the long-time dynamics is independent of the strength of the noise [62], provided the latter remains small enough to not significantly change the energy of the initial state. In the following, we denote by  $\langle \cdot \rangle$  the average over the ensemble of noise realizations.

*Heating Transition.*—Compared to classical systems, studies on thermalising dynamics in quantum models feature serious deficiencies, due to significant finite-size effects inherent to state-of-the-art exact diagonalization simulations. Since energy absorption is known to happen through Floquet many-body resonances [75–77], (i),

their density depends strongly on the drive frequency at any fixed many-body bandwidth. As the bandwidth scales linearly with the system size  $N$ , this puts an upper bound on  $\Omega$  for the system to be in the many-body regime. This also limits the occurrence of higher-order absorption processes, reducing the overall capacity for energy absorption. (ii) Low-energy initial states, whose energy level spacing does not follow the  $2^{-N}$  scaling of the bulk, further restrict the appearance of resonances. However, these issues are intrinsic to quantum models and none of them is problematic in periodically-driven classical systems. The classical energy manifold is continuous allowing for excitations at all energies, and one can easily reach system sizes of several hundred spins, which mitigates the constraint on the reliable upper bound for the driving frequency by a few orders of magnitude. Nevertheless, studying classical systems comes at a notable price: one cannot access the infinite-time behavior, and is thus limited to finite times.

Often, experiments in Floquet engineering are designed to study the GS of the infinite-frequency Hamiltonian (3). Therefore,  $H_{\text{ave}}$  constitutes a natural observable to measure the excess energy pumped into the system from the drive. Let us define the dimensionless expected energy and energy variance [55, 75], over the initial ensemble of noisy AFM states:

$$\langle Q(\ell T) \rangle \equiv \langle Q^{(0)}(\ell T) \rangle = \frac{\langle H_{\text{ave}}[\{\vec{S}_j(\ell T)\}] \rangle - E_{\text{GS}}}{\langle H_{\text{ave}} \rangle_{\beta=0} - E_{\text{GS}}} \in [0, 1],$$

$$\langle \delta Q(\ell T) \rangle = \sqrt{\frac{\langle H_{\text{ave}}^2[\{\vec{S}_j(\ell T)\}] \rangle - \langle H_{\text{ave}}[\{\vec{S}_j(\ell T)\}] \rangle^2}{\langle H_{\text{ave}}^2 \rangle_{\beta=0} - \langle H_{\text{ave}} \rangle_{\beta=0}^2}}. \quad (4)$$

The normalization is chosen w.r.t. an infinite-temperature ensemble, where each spin points at a random direction, and hence  $\langle H_{\text{ave}} \rangle_{\beta=0} = 0$  and  $\langle H_{\text{ave}}^2 \rangle_{\beta=0} = N/3(J^2/3 + h^2 + g^2)$ . Initializing the system in the ensemble of noisy AFM states, we have  $\langle Q(\ell T) \rangle \approx 0$  if the system does not absorb energy, and  $\langle Q(\ell T) \rangle = 1$  whenever the ensemble is heated to infinite temperature.

There are four stages in the evolution of the system, see Fig. 1. Notice that the time [cf. stage (iii)] between the pre-thermal plateau and the infinite-temperature state, corresponding to maximum energy variance:  $\ell_{\text{max}}(\Omega) = \text{argmax}_{\ell} \langle \delta Q(\ell T) \rangle$ , scales *exponentially* [78] with the driving frequency  $\Omega$ , c.f. Fig. 1b (inset). Our numerical study indicates that for  $h/J = 0.0$  the heating time can be parametrized as  $\ell_{\text{max}}(\Omega) = r(g/J) \exp[-\gamma(g/J) \times \Omega/J]$ , with  $\gamma(g/J)$  a slowly-varying function of  $g/J$ , and  $r(g/J) \propto (g/J)^\alpha$ ,  $\alpha \approx -2.12$ , within the entire range of existence of the pre-thermal stage [62]. Thus, we can rescale the  $\langle Q(\ell T) \rangle$  curves with respect to the energy in the beginning of the pre-thermal plateau (ii):

$$\langle Q_{\text{scaled}}(\ell_{\text{scaled}} T) \rangle = \frac{\langle Q(\ell_{\text{scaled}} T) \rangle - \langle Q \rangle_{\text{prethermal}}}{\langle Q \rangle_{\beta=0} - \langle Q \rangle_{\text{prethermal}}}, \quad (5)$$

where  $\ell_{\text{scaled}} = \ell/\ell_{\text{max}}(\Omega)$ . Figure 1a (inset) shows the collapse in the energy absorption curves with increasing

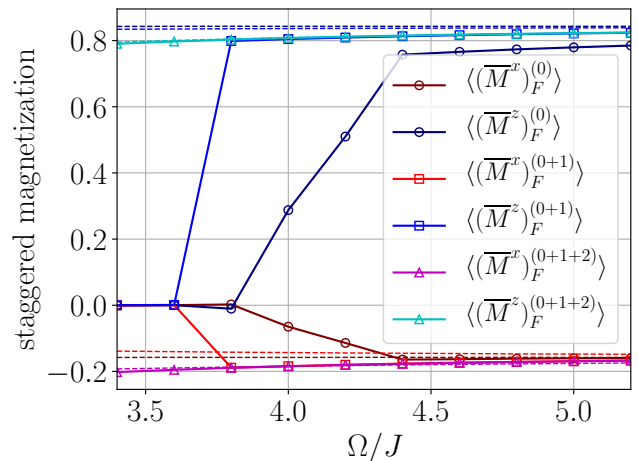


FIG. 2: Staggered magnetization  $\langle \langle \overline{M}^{\alpha} \rangle_F^{(0+1+\dots+m)} \rangle$  in the prethermal plateau to different order  $m$  in the ME (solid lines), compared to its value in the initial GS of the corresponding Floquet Hamiltonian (dashed). The parameters are  $g/J = 0.9045$ ,  $h/J = 0.809$ ,  $N_T = 10^6$ , and  $N = 100$ . Every point is averaged over an ensemble of 100 noise realisations.

frequency, whereas the width of the peak in  $\delta Q(\ell_{\text{scaled}})$  stays constant [62]. This behavior has remarkable similarities with the continuous phase transition observed in the 1d Ising model at zero temperature, where the squared magnetization plays the role of energy absorption, while inverse temperature and system size are analogous to drive frequency and time, respectively [62]; the correlation length corresponds to the heating time. This analogy suggests that heating may happen through a continuous phase transition in the limit  $\Omega, t \rightarrow \infty$ . At moderate frequencies and times, relevant for experiments, we find a sharp crossover instead.

*Prethermal Regime.*—The prethermal plateau plays a crucial role in strongly-interacting Floquet systems because it offers a stable window to experimentally realize novel many-body states of matter. We demonstrate that the inverse-frequency expansion can be used to gain a better understanding of the prethermal plateau, and present compelling evidence that this stage of evolution is captured by a *local* effective Hamilton function, amenable to Floquet engineering [79], even in chaotic classical many-body systems.

Even though Floquet theory does not apply to non-linear EOM, a Magnus expansion (ME) can be formally defined for classical systems by replacing commutators with Poisson brackets [3, 55]. The ME approximates the exact Floquet Hamiltonian  $H_F \approx H_F^{(0+\dots+m)} \sim \mathcal{O}(\Omega^{-m})$ , to a given order  $m$  in the inverse frequency [62]. However, it is an open question if and why the ME should work for classical systems. On one hand stands the notable application of the ME to the Kapitza pendulum [3, 49, 55], on the other – the recent finding that the ME does not capture resonances, which renders its convergence at most

asymptotic [33, 75, 77].

While energy is the most natural observable to study heating, it typically cannot be measured directly in experiments. We now show that the prethermal plateau affects generic local observables. To test Floquet theory, we initialize the system in the GS of  $H_F^{(0+\dots+m)}$  with different values of  $m$ . Consider the staggered magnetization and its long-time average

$$M^\alpha(\ell T) = \frac{1}{N} \sum_{j=1}^N (-1)^j S_j^\alpha(\ell T); \overline{M^\alpha} = \frac{1}{10^3} \sum_{\ell=N_T}^{N_T+10^3} M^\alpha(\ell T),$$

where  $N_T$  denotes a large number of driving cycles. This observable is an order parameter for AFM correlations, and its time dependence measures how well the system retains the initial AFM structure. Figure 2 (circles) shows the components of the time-averaged magnetisation  $\overline{M^\alpha}$  as a function of frequency to order  $m=0$ . For  $\Omega \lesssim \Omega_*$ , where  $\Omega_*$  is the crossover frequency, the system enters quickly the infinite-temperature stage, and all information about the initial state is lost:  $\overline{M^\alpha} = 0$ . However, in the long-lived prethermal plateau  $\Omega \gtrsim \Omega_*$ , much of the AFM correlations are preserved. The corresponding dashed lines show the staggered magnetization of the initial state, which is approached in the limit  $\Omega/J \rightarrow \infty$ . We verified that a similar behaviour is displayed by the spin-spin correlation function.

This raises the question whether one can Floquet-engineer expectation values of observables in the prethermal plateau. Upon increasing the order  $m$  of the ME, we find a significant improvement between the staggered magnetization of the initial ensemble and its time-averaged value in the pre-thermal plateau (squares, triangles), cf. Fig. 2. We also checked that starting from the GS of  $H_F^{(0)}$ , and evolving the system with the time-independent  $H_F^{(0+\dots+m)}$ , results in better agreement of the magnetization dynamics with the exact stroboscopic evolution, upon increasing  $m$  [62]. Since the ME is the main theoretical tool used in Floquet engineering [1–3], this result implies that one should be able to successfully Floquet-engineer the behavior of observables in the prethermal plateau in classical systems [79]. Such a behavior likely originates from the emergent quasiconserved local integrals of motion for  $\Omega \gtrsim \Omega_*$ .

Floquet prethermal plateaus are defined with respect to the local approximate extensive Floquet Hamiltonian  $H_F^{(0+\dots+m)}$ . They are often assumed to be featureless states which are stroboscopically equivalent to thermal equilibrium with respect to  $H_F^{(0+\dots+m)}$  for some optimal  $m$ . We now show that this assumption is incomplete: the prethermal plateau is sustained by drive-induced synchronization which is responsible for their exceptionally long stability. To demonstrate this, we compare the exact Floquet evolution, with an evolution where we repeatedly restart the dynamics from the thermal Gaussian ensemble of  $H_F^{(0)}$ , with mean energy and width chosen to match those of the time-evolved initial ensemble into

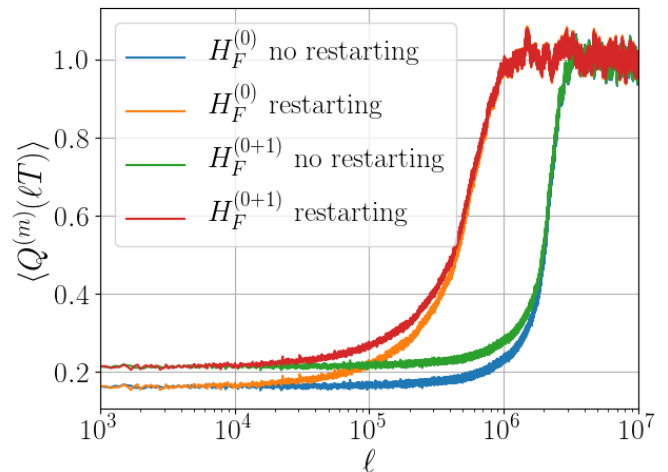


FIG. 3: Heating behavior measuring  $H_F^{(0)}$  and  $H_F^{(0+1)}$  with and without “restarting”, for  $\Omega/J = 3.8$ . The restarting procedure is repeated every  $10^3$  cycles [see text]. We define  $\langle Q^{(m)}(\ell T) \rangle$  to measure the normalized excess energy w.r.t.  $H_F^{(0+\dots+m)}$ , cf. Eq. (4). The difference in energy and its variance between the new and old ensembles was chosen to be less than  $0.005J$ . See Fig. 2 for parameters.

the prethermal plateau. We call this procedure “restarting”. Figure 3 shows a comparison of the expected values of  $H_F^{(0)}$ , following the uninterrupted (blue) and “restarting” (green) evolution. The “restarting” procedure was applied every  $10^3$  cycles to prevent the system from re-synchronizing. As a result, the “restarted” dynamics enters the infinite-temperature stage exponentially earlier. The increase of energy caused by “restarting” suggests that the prethermal plateau contains additional very slow synchronized dynamics, probably related to Arnold diffusion or subdiffusion [70–72], which serves as a glue for the prethermal state. To argue that the prethermal plateau is a property of the time-evolved ensemble and not of the Gaussian energy ensemble used for “restarting”, we fix the initial ensemble of noisy AFM states based on the GS of  $H_{\text{ave}} = H_F^{(0)}$ , and repeat the procedure measuring  $H_F^{(0+1)}$ , c.f. Fig 3. As expected, this affects the energy density of the prethermal plateau but not the duration of the stage. The restarting procedure is carried out w.r.t. a microcanonical distribution; we checked both a microcanonical and canonical distribution and found the same results. This restarting procedure shows that the pre thermal distribution cannot be characterized by energy and energy variance alone. Specifically, if this were the case then the thermalization times for the restarting and true evolutions would be the same. This restarting procedure is analogous to dephasing of density matrix in quantum mechanics.

*Outlook.*—We revealed a dynamical heating transition between a prethermal and an infinite-temperature stage in the limit of infinite times and drive frequen-

cies. Its existence influences strongly the evolution of periodically-driven many-body spin chains even at the experimentally-relevant moderate frequencies and times. This becomes manifest in a long-lived prethermal plateau, which can be modeled in a controllable fashion by an approximate effective nonlinear Floquet-Hamilton function derived within the limitations of the inverse-frequency expansion. Contrary to naïve expectations, the prethermal plateau is fused by drive-induced synchronization, and is not a featureless thermal state.

Even though a detailed comparison of thermalization in classical and quantum Floquet systems would be desirable, our analysis already presents compelling evidence that the prethermal plateau, observed in a variety of quantum models, survives in the classical limit [55, 72]. This suggests that studies in cold atomic Floquet systems aiming to explain the contribution to heating due to higher bands or preparation of states under periodic driving, can be done (semi-)classically to reduce finite-size effects. In fact, our study directly relates to experimental platforms, such as shaken superfluid ultracold gases, where the physics is governed by a classical spin model [17], or photonic topological insulators [29–31], described by the nonlinear wave equation.

*Note added:* While this manuscript was under peer review, a related work proving prethermalization in classical periodically-driven spin systems appeared [80].

*Acknowledgements.*— We thank Pankaj Mehta and Takashi Mori for interesting and insightful discussions. OH was supported by BU UROP student funding and the Simmons Foundation Investigator grant MMLS to Pankaj Mehta. DS acknowledges support from the FWO as post-doctoral fellow of the Research Foundation – Flanders and CMTV. AP and PW were supported by NSF DMR-1813499, AFOSR FA9550-16-1-0334 and ARO W911NF1410540. MB acknowledges support from the Emergent Phenomena in Quantum Systems initiative (EPiQS) of the Gordon and Betty Moore Foundation and ERC synergy grant UQUAM. We used **QuSpin** for simulating the nonlinear EOM in the case of monochromatic drive [81, 82]. The authors are pleased to acknowledge that the computational work reported on in this paper was performed on the Shared Computing Cluster which is administered by **Boston University’s Research Computing Services**. The authors also acknowledge the Research Computing Services group for providing consulting support which has contributed to the results reported within this paper.

- 
- [1] N. Goldman and J. Dalibard, *Phys. Rev. X* **4**, 031027 (2014).
- [2] A. Eckardt, *Rev. Mod. Phys.* **89**, 011004 (2017).
- [3] M. Bukov, L. D’Alessio, and A. Polkovnikov, *Advances in Physics* **64**, 139 (2015).
- [4] P. Kapitza, *Soviet Phys. JETP* **21**, 588 ((1951)).
- [5] T. Oka and H. Aoki, *Phys. Rev. B* **79**, 081406 (2009).
- [6] T. Kitagawa, T. Oka, A. Brataas, L. Fu, and E. Demler, *Phys. Rev. B* **84**, 235108 (2011).
- [7] G. Jotzu, M. Messer, R. Desbuquois, M. Lebrat, T. Uehlinger, D. Greif, and T. Esslinger, *Nature* **515**, 237 (2014).
- [8] M. Aidelsburger, M. Lohse, C. Schweizer, M. Atala, J. T. Barreiro, S. Nascimbène, N. R. Cooper, I. Bloch, and N. Goldman, *Nature Physics* **11**, 162 (2015).
- [9] N. Fläschner, B. S. Rem, M. Tarnowski, D. Vogel, D.-S. Lühmann, K. Sengstock, and C. Weitenberg, *Science* **352**, 1091 (2016).
- [10] M. Tarnowski, F. N. Ünal, N. Fläschner, B. S. Rem, A. Eckardt, K. Sengstock, and C. Weitenberg, *arXiv preprint arXiv:1709.01046* (2017).
- [11] M. Tarnowski, M. Nuske, N. Fläschner, B. Rem, D. Vogel, L. Freystatzky, K. Sengstock, L. Mathey, and C. Weitenberg, *Phys. Rev. Lett.* **118**, 240403 (2017).
- [12] M. Aidelsburger, S. Nascimbene, and N. Goldman, *arXiv preprint arXiv:1710.00851* (2017).
- [13] G. Jotzu, M. Messer, F. Görg, D. Greif, R. Desbuquois, and T. Esslinger, *Phys. Rev. Lett.* **115**, 073002 (2015).
- [14] J. Struck, C. Ölschläger, R. Le Targatn, P. Soltan-Panahi, A. Eckardt, M. Lewenstein, P. Windpassinger, and K. Sengstock, *Science* **333** (6045), 996 (2011).
- [15] J. Struck, C. Ölschläger, M. Weinberg, P. Hauke, J. Simonet, A. Eckardt, M. Lewenstein, K. Sengstock, and P. Windpassinger, *Phys. Rev. Lett.* **108**, 225304 (2012).
- [16] P. Hauke, O. Tieleman, A. Celi, C. Ölschläger, J. Simonet, J. Struck, M. Weinberg, P. Windpassinger, K. Sengstock, M. Lewenstein, and A. Eckardt, *Phys. Rev. Lett.* **109**, 145301 (2012).
- [17] J. Struck, M. Weinberg, C. Ölschläger, P. Windpassinger, J. Simonet, K. Sengstock, R. Höppner, P. Hauke, A. Eckardt, M. Lewenstein, and L. Mathey, *Nature Physics* **9**, 738 (2013).
- [18] M. Aidelsburger, M. Atala, M. Lohse, J. T. Barreiro, B. Paredes, and I. Bloch, *Phys. Rev. Lett.* **111**, 185301 (2013).
- [19] H. Miyake, G. A. Siviloglou, C. J. Kennedy, W. C. Burton, and W. Ketterle, *Phys. Rev. Lett.* **111**, 185302 (2013).
- [20] M. Atala, M. Aidelsburger, M. Lohse, J. T. Barreiro, B. Paredes, and I. Bloch, *Nature Physics* **10**, 588 (2014).
- [21] C. J. Kennedy, W. C. Burton, W. C. Chung, and W. Ketterle, *Nature Physics* **11**, 859 (2015).
- [22] M. Bukov, M. Kolodrubetz, and A. Polkovnikov, *Phys. Rev. Lett.* **116**, 125301 (2016).
- [23] V. Galitski and I. B. Spielman, *Nature* **494**, 49 (2013).
- [24] K. Jiménez-García, L. J. LeBlanc, R. A. Williams, M. C. Beeler, C. Qu, M. Gong, C. Zhang, and I. B. Spielman, *Phys. Rev. Lett.* **114**, 125301 (2015).
- [25] F. Görg, M. Messer, K. Sandholzer, G. Jotzu, R. Desbuquois, and T. Esslinger, *arXiv preprint arXiv:1708.06751* (2017).
- [26] A. Celi, P. Massignan, J. Ruseckas, N. Goldman, I. B. Spielman, G. Juzeliūnas, and M. Lewenstein, *Phys. Rev. Lett.* **112**, 043001 (2014).
- [27] B. Stuhl, H.-I. Lu, L. Ayccock, D. Genkina, and I. Spielman, *Science* **349**, 1514 (2015).

- [28] M. Mancini, G. Pagano, G. Cappellini, L. Livi, M. Rider, J. Catani, C. Sias, P. Zoller, M. Inguscio, M. Dalmonte, *et al.*, *Science* **349**, 1510 (2015).
- [29] M. C. Rechtsman, J. M. Zeuner, Y. Plotnik, Y. Lumer, D. Podolsky, F. Dreisow, S. Nolte, M. Segev, and A. Szameit, *Nature* **496**, 196 (2013).
- [30] M. Hafezi, *Phys. Rev. Lett.* **112**, 210405 (2014).
- [31] S. Mittal, J. Fan, S. Faez, A. Migdall, J. M. Taylor, and M. Hafezi, *Phys. Rev. Lett.* **113**, 087403 (2014).
- [32] R. Desbuquois, M. Messer, F. Görg, K. Sandholzer, G. Jotzu, and T. Esslinger, *Phys. Rev. A* **96**, 053602 (2017).
- [33] P. Weinberg, M. Bukov, L. DAlessio, A. Polkovnikov, S. Vajna, and M. Kolodrubetz, *Physics Reports* (2017).
- [34] V. Novičenko, E. Anisimovas, and G. Juzeliūnas, *Phys. Rev. A* **95**, 023615 (2017).
- [35] M. Bukov, *arXiv preprint arXiv:1808.08910* (2018).
- [36] T. Bilitewski and N. R. Cooper, *Phys. Rev. A* **91**, 033601 (2015).
- [37] T. Bilitewski and N. R. Cooper, *Phys. Rev. A* **91**, 063611 (2015).
- [38] M. Weinberg, C. Ölschläger, C. Sträter, S. Prella, A. Eckardt, K. Sengstock, and J. Simonet, *Phys. Rev. A* **92**, 043621 (2015).
- [39] M. Reitter, J. Näger, K. Wintersperger, C. Sträter, I. Bloch, A. Eckardt, and U. Schneider, *Phys. Rev. Lett.* **119**, 200402 (2017).
- [40] J. Näger, K. Wintersperger, M. Bukov, S. Lellouch, E. Demler, U. Schneider, I. Bloch, N. Goldman, and M. Aidelsburger, *arXiv preprint arXiv:1808.07462* (2018).
- [41] T. Boulier, J. Maslek, M. Bukov, C. Bracamontes, E. Magnan, S. Lellouch, E. Demler, N. Goldman, and J. Porto, *arXiv preprint arXiv:1808.07637* (2018).
- [42] S. Fishman, D. R. Grempel, and R. E. Prange, *Phys. Rev. Lett.* **49**, 509 (1982).
- [43] S. Lellouch, M. Bukov, E. Demler, and N. Goldman, *Phys. Rev. X* **7**, 021015 (2017).
- [44] S. Lellouch and N. Goldman, “Parametric instabilities in resonantly-driven bose-einstein condensates,” (2017), *arXiv:1711.08832*.
- [45] L. D’Alessio and M. Rigol, *Phys. Rev. X* **4**, 041048 (2014).
- [46] A. Lazarides, A. Das, and R. Moessner, *Phys. Rev. E* **90**, 012110 (2014).
- [47] Y. Bar Lev, D. J. Luitz, and A. Lazarides, *SciPost Physics* **3**, 029 (2017).
- [48] R. Moessner and S. Sondhi, *Nature Physics* **13**, 424 (2017).
- [49] R. Citro, E. G. Dalla Torre, L. DAlessio, A. Polkovnikov, M. Babadi, T. Oka, and E. Demler, *Annals of Physics* **360**, 694 (2015).
- [50] K. Seetharam, P. Titum, M. Kolodrubetz, and G. Refael, *Phys. Rev. B* **97**, 014311 (2018).
- [51] T. Prosen, *Phys. Rev. Lett.* **80**, 1808 (1998).
- [52] T. Prosen, *J. Phys. A: Math. Gen.* **31**, L645 (1998).
- [53] T. Prosen, *Phys. Rev. E* **60**, 3949 (1999).
- [54] T. Prosen, *Phys. Rev. E* **65**, 036208 (2002).
- [55] L. DAlessio and A. Polkovnikov, *Annals of Physics* **333**, 19 (2013).
- [56] A. Haldar, R. Moessner, and A. Das, *Phys. Rev. B* **97**, 245122 (2018).
- [57] D. A. Abanin, W. De Roeck, and F. Huveneers, *Phys. Rev. Lett.* **115**, 256803 (2015).
- [58] T. Mori, T. Kuwahara, and K. Saito, *Phys. Rev. Lett.* **116**, 120401 (2016).
- [59] M. Bukov, S. Gopalakrishnan, M. Knap, and E. Demler, *Phys. Rev. Lett.* **115**, 205301 (2015).
- [60] S. A. Weidinger and M. Knap, *Scientific reports* **7**, 45382 (2017).
- [61] F. Peronaci, M. Schiró, and O. Parcollet, *arXiv preprint arXiv:1711.07889* (2017).
- [62] See Supplemental Material.
- [63] E. Fermi, J. Pasta, S. Ulam, and M. Tsingou, *Los Alamos National Laboratory Document LA-1940* (1955).
- [64] S. F. C. Danieli, DK Campbell, “Intermittent fpu dynamics at equilibrium,” (2016), *arXiv:1611.00434*.
- [65] N. N. Nekhoroshev, *Functional Analysis and Its Applications* **5**, 338 (1971).
- [66] J. Moser, *Math. Phys. K1 Ila nr.6* (1955), 87120 (1955).
- [67] J. Littlewood, *Proceedings of the London Mathematical Society* **3**, 343 (1959).
- [68] J. Pöschel, *Mathematische Zeitschrift* **213**, 187 (1993).
- [69] M. Pettini and M. Landolfi, *Phys. Rev. A* **41**, 768 (1990).
- [70] T. Konishi and K. Kaneko, *Journal of Physics A: Mathematical and General* **23**, L715 (1990).
- [71] S. Notarnicola, F. Iemini, D. Rossini, R. Fazio, A. Silva, and A. Russomanno, *arXiv preprint arXiv:1709.05657* (2017).
- [72] A. Rajak, R. Citro, and E. G. D. Torre, *arXiv preprint arXiv:1801.01142* (2018).
- [73] G. Veble and T. Prosen, *Phys. Rev. Lett.* **92**, 034101 (2004).
- [74] G. Veble and T. Prosen, *Phys. Rev. E* **72**, 025202 (2005).
- [75] M. Bukov, M. Heyl, D. A. Huse, and A. Polkovnikov, *Phys. Rev. B* **93**, 155132 (2016).
- [76] P. W. Claeys and J.-S. Caux, *arXiv preprint arXiv:1708.07324* (2017).
- [77] P. W. Claeys, S. De Baerdemacker, O. E. Araby, and J.-S. Caux, *arXiv preprint arXiv:1712.03117* (2017).
- [78] F. Machado, G. D. Meyer, D. V. Else, C. Nayak, and N. Y. Yao, *arXiv preprint arXiv:1708.01620* (2017).
- [79] F. H. Higashikawa, Sho and M. Sato, *Arxiv preprint arXiv:1810.01103* (2018).
- [80] T. Mori, *Phys. Rev. B* **98**, 104303 (2018).
- [81] P. Weinberg and M. Bukov, *SciPost Phys.* **2**, 003 (2017).
- [82] P. Weinberg and M. Bukov, *arXiv preprint arXiv:1804.06782* (2018).

## Supplemental Material

### I. ANALOGY OF THE PHYSICS OF THE DYNAMICAL HEATING TRANSITION WITH THE 1D ISING MODEL CLOSE TO ZERO TEMPERATURE

The thermalization dynamics of the dimensionless energy  $\langle Q(\ell T) \rangle$  and energy variance  $\langle \delta Q(\ell T) \rangle$  studied in this work are very similar in nature to the behavior of the squared magnetization in the one-dimensional Ising model close to zero temperature. Consider the 1d Ising energy function

$$\mathcal{H} = -J \sum_{j=1}^N \sigma_{j+1} \sigma_j, \quad (6)$$

where  $\sigma_j \in \{\pm 1\}$  are Ising variables and  $J$  is the interaction strength. This model is a textbook example for a continuous phase transition between a disordered phase at finite temperature and a magnetized point at zero temperature. Below, we discuss how this critical point affects the small but finite temperature physics, and draw an analogy to the dynamical heating phase transition reported on in the main text.

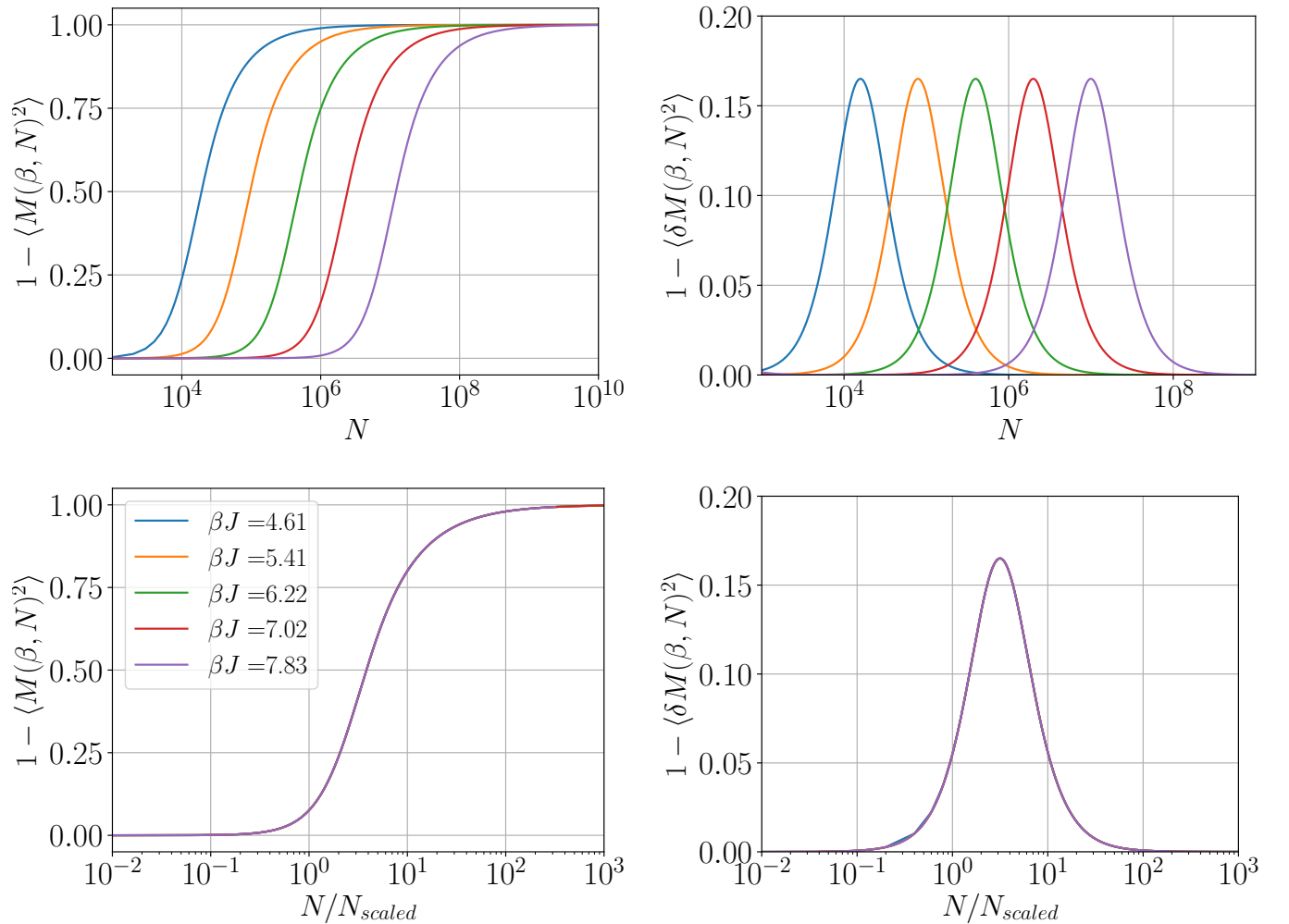


FIG. 4: The left plots show the behavior of the squared magnetization  $\langle M^2(\beta, N) \rangle$  for different inverse temperatures  $\beta$  as a function of the system size  $N$ . The right plots show the behavior of the variance of the squared magnetization  $\langle \delta M(\beta, N)^2 \rangle$ . The top panel on both sides shows the exact behavior while the bottom panel is rescaled with respect to the correlation length  $N_{\text{scaled}} = \xi(\beta)$  which is an exponential function of the inverse temperature. To be compared to Fig. 5.

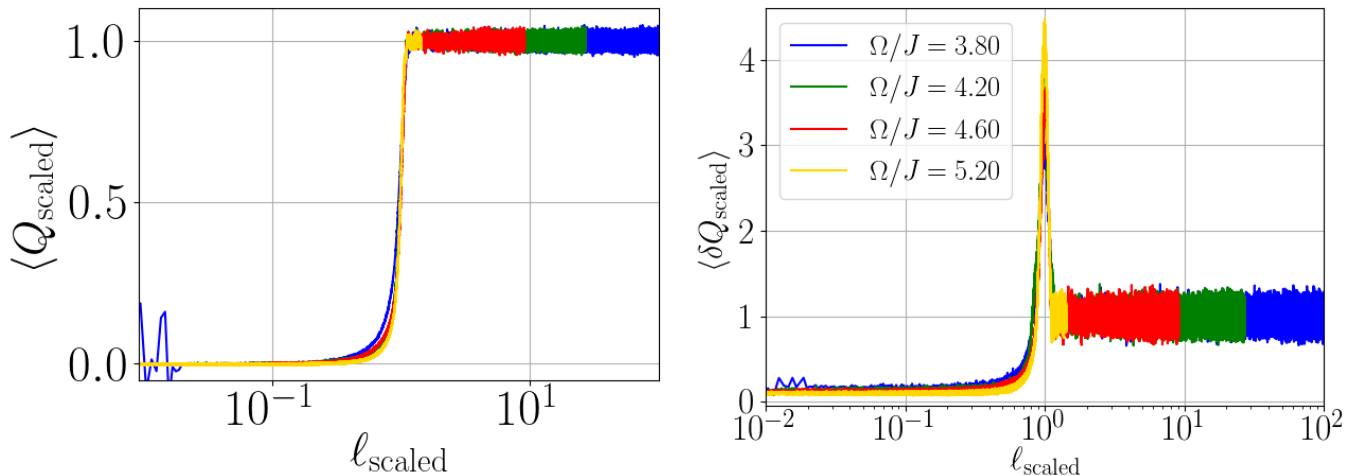


FIG. 5: Noise-averaged energy  $\langle Q(\ell T) \rangle$  and energy variance  $\langle \delta Q(\ell T) \rangle$  as a function of the rescaled number of driving cycles  $\ell_{\text{scaled}}$ . The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ . To be compared to Fig. 4.

Figure 4 shows the thermal expectation of the squared magnetization:  $M^2 = (N^{-1} \sum_i \sigma_i)^2$  and its variance as a function of the length of the chain  $N$  for different inverse temperatures  $\beta$ . The curves show a crossover when the system size becomes of order the correlation length which goes as  $\xi(\beta) \sim \exp(2J\beta)$ .

The scaling analysis for the heating transition described in the main text clearly exhibits the following scaling behaviour:

$$Q(\ell T, \Omega) = f_Q(\ell/l_{\text{max}}(\Omega)), \quad (7)$$

with  $l_{\text{max}}(\Omega) \sim \exp(c\Omega)$ , see Fig. 5. This bears a striking resemblance to the equilibrium finite size scaling in the 1d Ising chain:

$$\langle M^2(\beta, N) \rangle = f_{M^2}(N/\xi(\beta)). \quad (8)$$

Similar to defining a scaling time we can define a scaling length as  $N_{\text{scaled}} = \xi(\beta)$ . This suggests that for the heating transition,  $\Omega$  plays the role of  $\beta$ . At the same time, the correlation length,  $\xi(\beta) = N_{\text{scaled}}$ , is analogous to the time needed to begin the unconstrained thermalization process. Furthermore this analogy gives a very strong evidence for a dynamical heating transition at  $\Omega, t \rightarrow \infty$  as both the time dependent model and the equilibrium model share the same scaling behavior.

## II. TIME EVOLUTION OF THE STAGGERED MAGNETIZATION

Figure 6 shows the time-evolution of the magnetization for different frequencies, starting from the ground state of the period-averaged Hamiltonian  $H_{\text{ave}}$ . Each data point is averaged over 50 noise realizations of the initial state. For low frequencies the mean magnetization exhibits prethermal plateaus that eventually decay to zero. This decay originates from the breakdown of the anti-ferromagnetic structure of the initial state.

Figure 7 shows the behavior of the staggered magnetization for a single realization. In this case the magnetization exhibits spin flips which appear sharp even on a linear time scale. These processes reveal the physical mechanism behind the onset of heating, which is associated with the proliferation of spin flips towards the end of the exponentially long-lived prethermal plateau. For moderate-to-high frequencies these magnetization plateaus persist for the entire duration of the prethermal plateau.

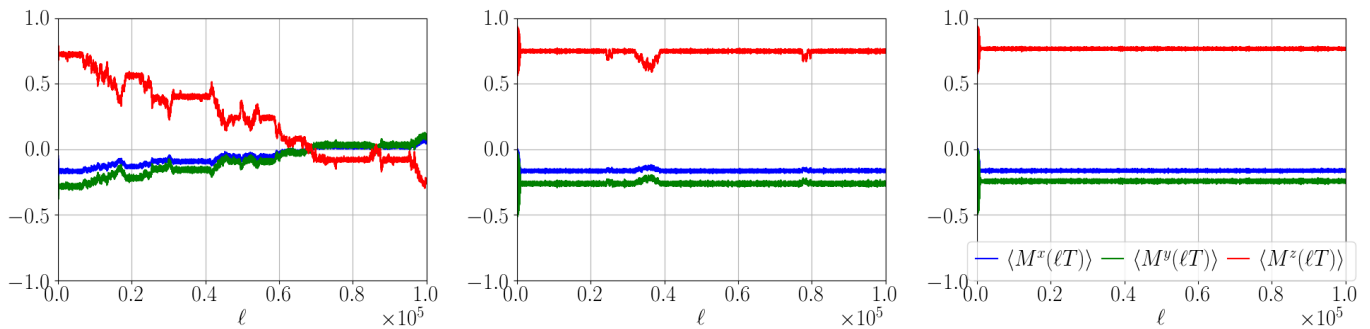


FIG. 6: The components of the staggered magnetization  $\langle M^\alpha(\ell T) \rangle$  as a function of the number of driving cycles  $\ell$ , for  $\Omega/J = 3.8$  (left),  $\Omega/J = 4.2$  (middle) and  $\Omega/J = 4.6$  (right). Each data point is averaged over 50 noise-realizations. The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ .

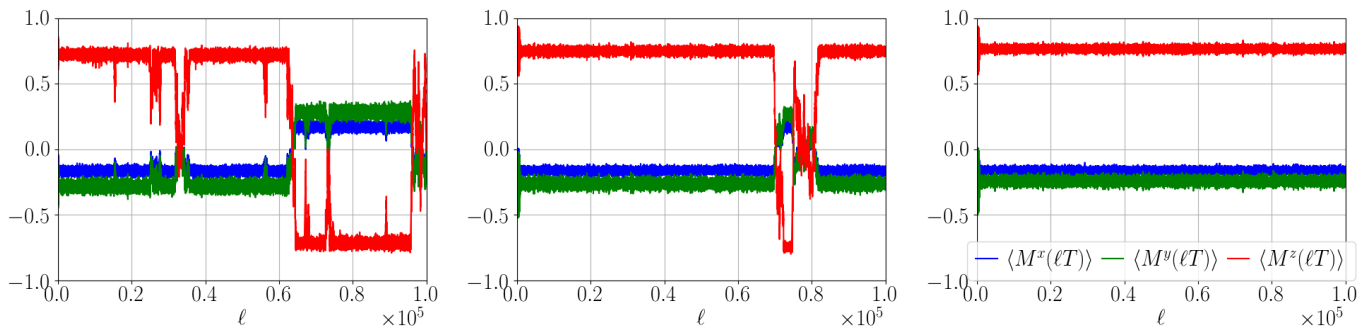


FIG. 7: A single noise realization showing the components of the staggered magnetization  $\langle M^\alpha(\ell T) \rangle$  as a function of the number of driving cycles  $\ell$ , for  $\Omega/J = 3.8$  (left),  $\Omega/J = 4.2$  (middle) and  $\Omega/J = 4.6$  (right). The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ .

### III. FREQUENCY DEPENDENCE OF ENERGY ABSORPTION

Often times one is interested in the average amount of energy absorbed by the system from the drive. To measure this, let us define the following time-averaged energy absorption, averaged at a given decade over  $10^3$  periods

$$\langle \bar{Q}^{N_T}(\Omega) \rangle = \frac{1}{10^3} \sum_{\ell=N_T}^{N_T+1000} \langle \bar{Q}(\ell T; \Omega) \rangle, \quad (9)$$

where  $N_T$  denotes the first driving cycle of the period used for the long-time average. Theoretically, one would like to do the period averaging at infinite time,  $N_T \rightarrow \infty$ . However, in classical systems this is not possible. Therefore, we study the heating dependence as a function of frequency, for a sequence of exponentially growing time spans.

The existence of a heating transition at  $\Omega \rightarrow \infty$  and  $\ell \rightarrow \infty$  suggests that the infinite-temperature plateau of  $\langle \bar{Q}^{N_T}(\Omega) \rangle$  shifts to higher frequencies with increasing waiting time  $N_T$ . Figure 8 shows  $\langle \bar{Q}^{N_T}(\Omega) \rangle$  over seven consecutive decades in  $N_T$ . In agreement with theory, as  $N_T$  is increased, the energy curves shift gradually to the right. Interestingly, the crossover region between the infinite-temperature and the prethermal stages sharpens very slowly which suggests the existence of logarithmic corrections to the finite time scaling.

### IV. DERIVATION OF THE EFFECTIVE FLOQUET-HAMILTON FUNCTION USING THE INVERSE-FREQUENCY EXPANSION

The cornerstone of Floquet theory in physics is Floquet's theorem. However, the theorem applies to *linear* ordinary differential equations, and hence its consequences become invalid for systems obeying nonlinear dynamics. On the other hand, chaos and thermalization are intimately tied to the nonlinearities in the Hamilton equations of motion.

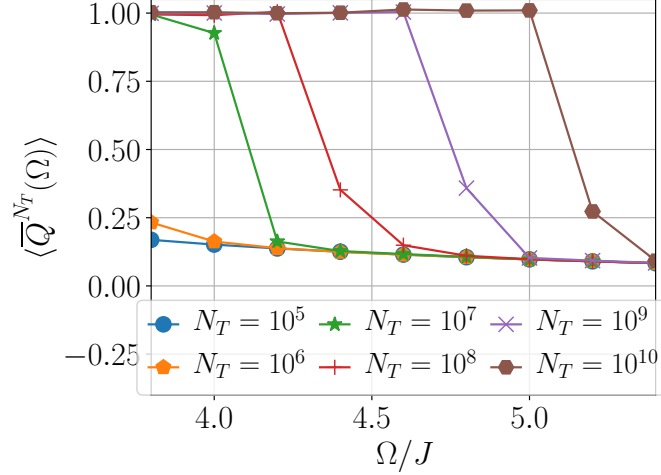


FIG. 8: Decade averaged energy absorption  $\langle \bar{Q}^{N_T}(\Omega) \rangle$  for six different decades as a function of the driving frequency  $\Omega/J$ . Each average is taken over 1000 periods. The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ .

Therefore, finding effective descriptions for classical periodically-driven many-body systems is a challenging and difficult problem, and there might not exist a simple universal solution.

In periodically-driven quantum systems, the linearity of the Schrödinger equation fulfils the criteria for the applicability of Floquet's theorem. According to the latter, the evolution operator  $U(t, 0)$  factorizes as

$$\hat{U}(t, 0) = \mathcal{T}_t \exp \left( -i \int_0^t dt' \hat{H}(t') \right) = \hat{P}(t) e^{-it\hat{H}_F}, \quad (10)$$

with the so-called time-periodic micromotion operator  $\hat{P}(t) = \hat{P}(t+T)$ , which governs the fast intraperiod dynamics, and the effective time-independent Floquet Hamiltonian  $\hat{H}_F$ . Stroboscopically, the time evolution of the system is therefore described by the Floquet Hamiltonian  $\hat{H}_F$ .

The starting point in quantum Floquet theory is the assumption that the Floquet Hamiltonian can be expanded in a Taylor series in the inverse frequency

$$H_F = \sum_{n=0}^{\infty} H_F^{(n)}, \quad H_F^{(n)} \sim \mathcal{O}(\Omega^{-n}), \quad H_F^{(0+\dots+m)} = \sum_{n=0}^m H_F^{(n)}. \quad (11)$$

Combining this with Eq. (10), one arrives at the inverse-frequency expansion for the Floquet Hamiltonian, for more details see Refs. [1–3]. A notable advantage of this expansion is that it remains valid beyond the linear response regime, and allows to study strongly driven systems.

Let us apply the Magnus expansion – one of the variants of the inverse-frequency expansion – to the step-driven spin chain. Note that if we write the Hamiltonian as

$$\hat{H}(t) = \begin{cases} \hat{A} & \text{for } t \in [0, T/2] \text{ mod } T \\ \hat{B} & \text{for } t \in [T/2, T] \text{ mod } T \end{cases}$$

where

$$\hat{A} = \sum_{j=1}^N J \hat{S}_j^z \hat{S}_{j+1}^z + h \hat{S}_j^z, \quad \hat{B} = \sum_{j=1}^N g \hat{S}_j^x, \quad (12)$$

the time evolution operator takes the form

$$\hat{U}(T, 0) = \exp(-i\hat{H}_F T) = \exp\left(-i\hat{B}\frac{T}{2}\right) \exp\left(-i\hat{A}\frac{T}{2}\right). \quad (13)$$

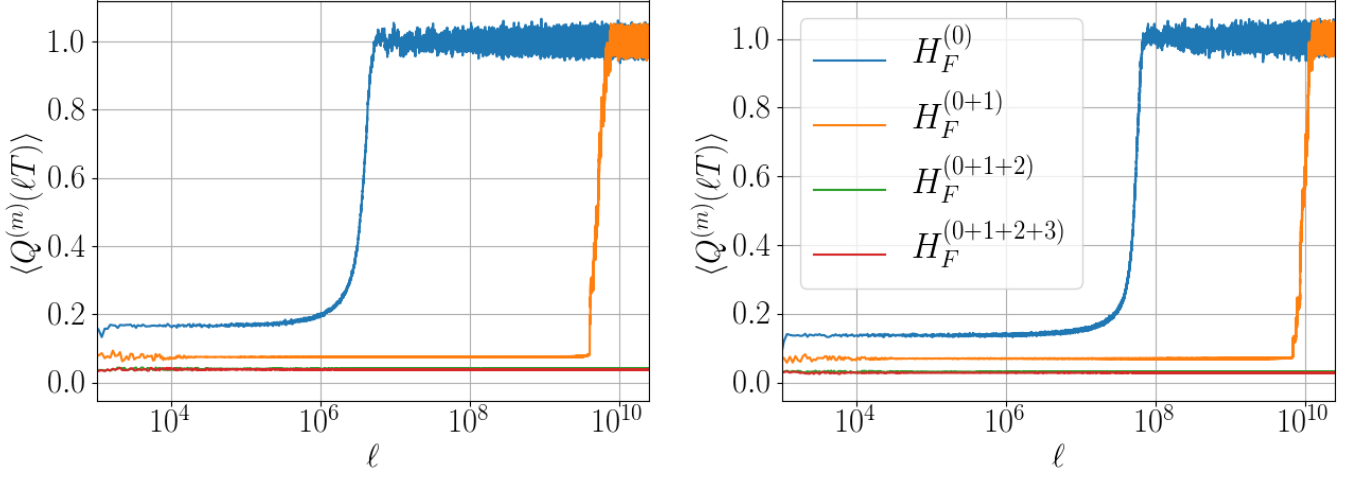


FIG. 9: Average energy absorption  $\langle Q(t) \rangle$  as a function of the number of driving cycles  $\ell$ , for different  $H_F^{(0+\dots+m)}$  for  $m \in [0, 1, 2, 3]$ , see Eq. (16). Higher order terms absorb less energy after the initial quench and exhibit less fluctuations in the prethermal plateau. The evolution is started in the ground state of  $H_F^{(0+\dots+m)}$ . The frequency values are  $\Omega/J = 3.8$  (left) and  $\Omega/J = 4.6$  (right). The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ .

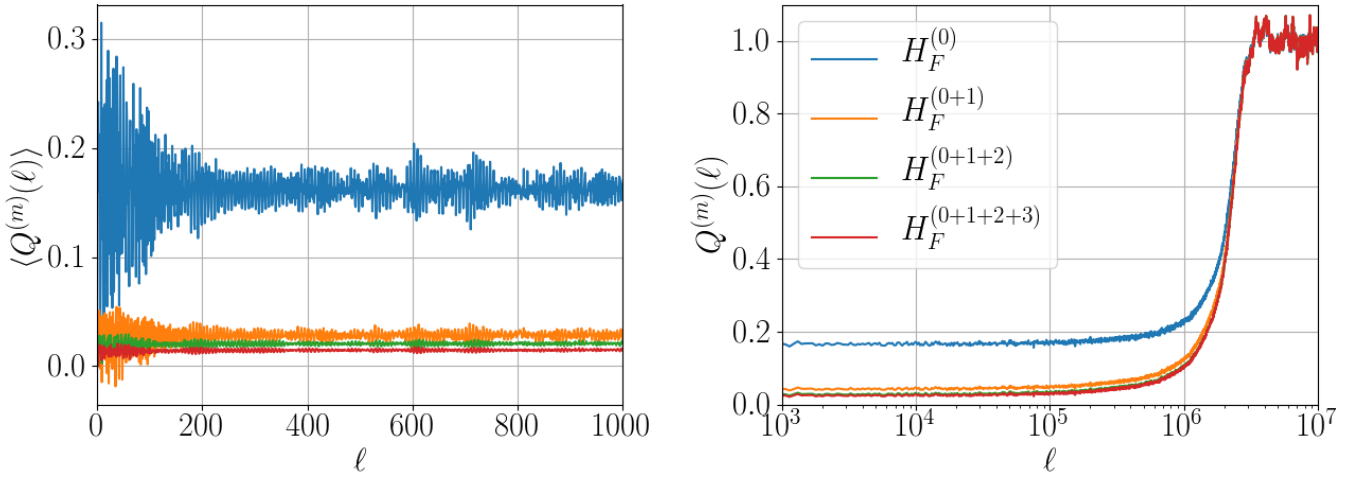


FIG. 10: Average energy absorption  $\langle Q(t) \rangle$  as a function of the number of driving cycles  $\ell$ , for different  $H_F^{(0+\dots+m)}$  for  $m \in [0, 1, 2, 3]$ . The evolution is started in the ground state of  $H_F^{(0)}$  for all curves. The frequency value is  $\Omega/J = 3.8$ . The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ . Note that when started in  $H_F^{(0)}$  all Floquet expansions thermalize after the same number of periods but the evolution of higher order Floquet terms exhibit less fluctuations after the initial quench and in the prethermal plateau.

For this setup, applying the Magnus expansion is equivalent to the Baker-Campbell-Hausdorff formula, which to third order in the inverse frequency gives:

$$\begin{aligned}
 \hat{H}_{\text{ave}} &= \hat{H}_F^{(0)} = \frac{1}{2}(\hat{A} + \hat{B}), \\
 \hat{H}_F^{(1)} &= \frac{T}{8}[\hat{B}, \hat{A}], \\
 \hat{H}_F^{(2)} &= \frac{T^2}{96}([\hat{B}, [\hat{B}, \hat{A}]] + [\hat{A}, [\hat{A}, \hat{B}]]) , \\
 \hat{H}_F^{(3)} &= -\frac{T^3}{288}[\hat{A}, [\hat{B}, [\hat{B}, \hat{A}]]].
 \end{aligned} \tag{14}$$

In Ref. [3] it was noted that if we formally replace the commutator  $[\cdot, \cdot]$  by the Poisson bracket  $\{\cdot, \cdot\}$ , the expansion translates over to classical systems. In our model, this leads to

$$\begin{aligned}
H_F^{(0)} &= \frac{1}{2} \sum_{j=1}^N JS_j^z S_{j+1}^z + hS_j^z + gS_j^x, \\
H_F^{(1)} &= \frac{-gT}{8} \sum_{j=1}^N (J(S_{j+1}^y + S_{j-1}^y) + h)S_j^y, \\
H_F^{(2)} &= \frac{T^2}{96} \sum_{j=1}^N Jg^2 (S_j^y (S_{j+1}^y + S_{j-1}^y) - S_j^z (S_{j+1}^z + S_{j-1}^z)) - g^2 h S_j^z \\
&\quad - 2Jgh (S_{j+1}^z + S_{j-1}^z) S_j^x - gh^2 S_j^x - J^2 g (2S_{j+1}^z S_{j-1}^z + (S_{j+1}^z)^2 + (S_{j-1}^z)^2), \\
H_F^{(3)} &= \frac{T^3}{284} \sum_{j=1}^N Jg^2 h S_j^x (S_{j+1}^z + S_{j-1}^z) + J^2 g^2 S_{j+1}^z S_j^x (S_{j+1}^z + S_{j-1}^z) + J^2 g^2 S_j^z S_{j+1}^x (S_{j+2}^z + S_j^z). \tag{15}
\end{aligned}$$

The above approach is equivalent to starting from a classical periodically-driven system, quantizing it, applying the Magnus expansion, and then taking the classical limit. While this procedure is straightforward to apply, the absence of a Floquet theorem for classical chaotic systems casts doubt on how controlled this expansion is.

To study this, for every order  $m$  in the Magnus expansion, we start from the GS of the corresponding approximate classical Floquet-Hamilton function  $H_F^{(0+\dots+m)}$ . The exact GS of  $H_F^{(0+\dots+m)}$  can be computed by first making use of the two-site translational invariance of the system to effectively reduce the number of degrees of freedom. Except for  $m = 0$ , we have a nonzero  $S_j^y$  component, induced by the  $y$ -field in the first-order Floquet-Hamilton function. To find the correct ground state we, therefore, parametrize the spin vector  $\vec{S}_j$  by spherical coordinates  $(\theta_j, \phi_j)$ ,  $j = 1, 2$ , and variationally determine the angles through minimizing the energy  $E_{\text{GS}}^{(0+\dots+m)}(\theta_j, \phi_j)$ . The true ground state must thus be solved for numerically. Note that the ground state for  $m > 0$  is also dependent on the driving frequency  $\Omega$ .

We then evolve the initial state with the exact time-dependent Hamilton function  $H(t)$ , and measure the energy  $H_F^{(0+\dots+m)}$ . Figure 9 shows a comparison between the different order Magnus expansion for energy absorption [see main text for a comparison on the staggered magnetization]. The normalized energy-absorption quantity  $Q$  is more generally defined as:

$$\langle Q^{(m)}(\ell T) \rangle = \frac{\langle H_F^{(0+\dots+m)}[\{\vec{S}_j(\ell T)\}] \rangle - E_{\text{GS}}^{(0+\dots+m)}}{\langle H_F^{(0+\dots+m)} \rangle_{\beta=0} - E_{\text{GS}}^{(0+\dots+m)}}. \tag{16}$$

Starting from the GS of the Magnus-corrected Hamiltonian, we find that the higher-order terms have the following salient effects, see Fig. 9. Firstly, the initial energy fluctuations arising from the ferromagnetic structure of the ground state during the constrained thermalization [stage (i) in Fig. 1, main text] are reduced. In addition, the unconstrained thermalization transient [stage (iii) in Fig. 1, main text] shows a sharper transition with increasing order of the expansion.

To study the effect of the initial state on the dynamics, we can fix the initial GS to be the GS of  $H_F^{(0)}$ , and repeat the procedure, measuring the energy functions  $H_F^{(0+\dots+m)}$ , see Fig. 10. Clearly, while the pre-thermal plateau is reduced with increasing the order of the expansion, the heating time now remains unchanged. However, starting the evolution in the GS of a higher order Floquet Hamiltonian increases the duration of the prethermal plateau. If the evolution was completely described by quench with respect to some high order Floquet Hamiltonian, then the corresponding energy function  $H_F^{(0+\dots+m)}$  of that order should be approximately conserved except for some energy absorption when the drive is turned on. We observe that this is true in the prethermal plateau but the heating time is not changed. The fact that the heating time does not change means that the heating mechanism is not completely explained by the Magnus expansion and the heating time is a function of the initial state.

One might wonder how well the approximate Floquet-Magnus Hamiltonians describe the dynamics of observables within the prethermal plateau. In order study this, we prepare the system in the GS of  $H_F^{(0)}$ , and evolve it with the time-independent approximate  $H_F^{(0+\dots+m)}$ . Our aim is to compare the approximate dynamics, generated by  $H_F^{(0+\dots+m)}$ , to the exact periodic-step evolution. To compare the two kinds of evolution, we look into the staggered  $z$ -magnetization,  $\langle M^z(\ell) \rangle$ , which serves as an order parameter for the anti-ferromagnetic correlations in the GS, as mentioned in the main text. Since it is difficult to compare rapidly oscillating observables over long times, we also

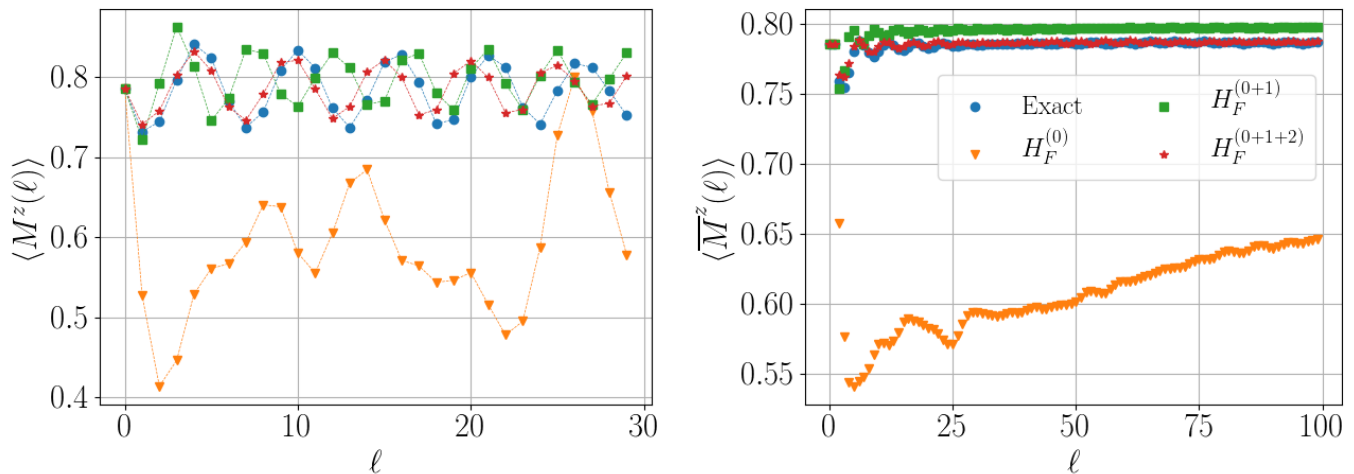


FIG. 11: Comparison of stroboscopic staggered magnetization  $\langle M^z(\ell) \rangle$  and average stroboscopic staggered magnetization  $\langle \bar{M}^z(\ell) \rangle$ , see Eq. (17), as a function of the number of driving periods. The thin dashed lines are only a guide to the eye and were not used to do time averaging. The stroboscopic dynamics is generated by the exact step-drive evolution and different orders of the time-independent Floquet Hamiltonian, respectively, started from the same initial state. Note that going to higher order in the Magnus expansion increases the accuracy of the dynamics both stroboscopically (left panel) and in long-time average (right panel). The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ ,  $N = 100$  and  $\Omega/J = 4.0$ .

look at the stroboscopic time-average of  $M^z$ :

$$\langle \bar{M}^z(\ell) \rangle = \frac{1}{\ell} \sum_{\kappa=0}^{\ell} \langle M^z(\kappa) \rangle. \quad (17)$$

We find that quenching with a higher-order Floquet Hamiltonian reproduces the exact dynamics with greater fidelity, cf. Fig. 11. This shows that the dynamics following the exact evolution is better captured as the order of the expansion is increased, although this behavior might reverse at some higher order, as is common for asymptotic expansions [57, 58]. That is an intriguing result because there is no straightforward generalization of Floquet's theorem to nonlinear systems.

## V. PRETHERMAL PHASE

The behavior of the system in the prethermal phase looks like a thermal state for most local observables like magnetization. However, we argue that the prethermal phase is not completely described by a thermal distribution. This is shown in Figure 3 of the main text but we expand on this argument here. Consider the following numerical experiment: We take an ensemble of states in the prethermal regime and evolve with exact evolution for 1000 periods and then evolve with  $H_F^{(0)}$  or  $H_F^{(0+1)}$  for 1000 periods. Next, we take the same initial state and evolve with  $H_F^{(0+1+2)}$  for 1000 periods and quench with respect to  $H_F^{(0)}$  or  $H_F^{(0+1)}$  for 1000 periods.

Figure 12 shows that when evolving with  $H_F^{(0+1+2)}$  the heating rate of quenching with respect to  $H_F^{(0+1)}$  is much less than that of  $H_F^{(0)}$ . When we evolve with the exact evolution we find that the rates of heating of  $H_F^{(0+1)}$  and  $H_F^{(0)}$  are almost identical. This implies that the description of the prethermal phase is not completely captured by a thermal ensemble. The heating we see as a result of restarting is inconsistent with the picture that we simply quench to the wrong Hamiltonian because then we would observe a very strong difference of the heating rate for restarting with  $H_F^{(0)}$  and  $H_F^{(0+1)}$ . This observation leads us to believe that behavior of system in the prethermal phase is not completely described by a thermal state and may contain additional slow coherent dynamics that are destroyed by quenching or restarting processes.

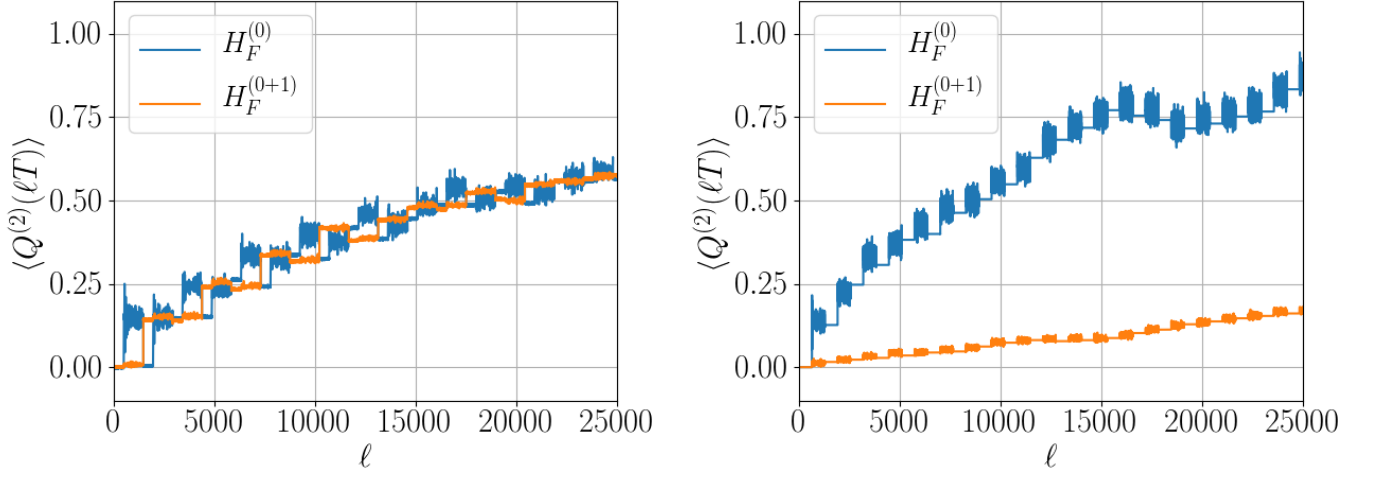


FIG. 12: The left panel shows average energy absorption  $\langle Q^{(2)}(t) \rangle$  as a function of the number of driving cycles  $\ell$  for dynamics with alternating quenches between  $H_F^{(0)}$  and exact evolution (left panel blue curve) and  $H_F^{(0+1)}$  and exact evolution (left panel orange curve). The right panel shows average energy absorption  $\langle Q^{(2)}(t) \rangle$  as a function of the number of driving cycles  $\ell$  for dynamics with alternating quenches between  $H_F^{(0)}$  and  $H_F^{(0+1+2)}$  evolution (right panel blue curve) and  $H_F^{(0+1)}$  and  $H_F^{(0+1+2)}$  evolution (right panel orange curve). The length of each quench is 1000 periods. The graphs shown have  $\Omega/J = 4.0$ . The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ .

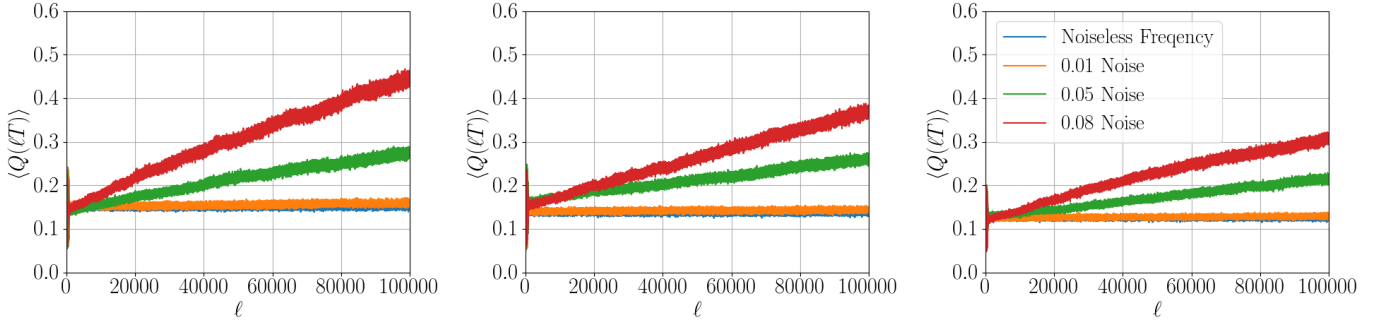


FIG. 13: Average energy absorption  $\langle Q(t) \rangle$  as a function of the number of driving cycles  $\ell$  for different noisy driving durations [see text]. The evolution goes up to  $10^5$  cycles. The graphs shown have  $\Omega/J = 4.0$  (left),  $\Omega/J = 4.2$  (middle) and  $\Omega/J = 4.4$  (right). The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ . The legend applies to all three graphs.

## VI. NOISY DRIVES AND SYNCHRONIZATION

In this paper we have considered only driving periods of constant duration. If small amounts of noise are added to the duration of each driving frequency  $T$  then we find a much faster rate of heating. This suggests that the long lifetime of the prethermal regime as seen in Figure 1 in the main text may be due to the emergence of quasi-conserved quantities caused by synchronization effects.

To test this, before each new driving cycle, we slightly change the cycle duration  $T'$  nondeterministically. In particular, using the relation  $\Omega' = 2\pi/T'$ , we draw a new drive frequency  $\Omega' = \Omega + \mu$ ;  $\mu$  is a random number drawn from the uniform probability distribution supported on the interval  $[0, 2\sigma\sqrt{3}]$ , and  $\sigma$  is the variance of the distribution which sets the strength of the noise. Figure 13 shows that the rate of heating is very susceptible to such noise. This suggests that the existence of the long-lived prethermal regime in the noiseless case is due to some form of coherence effects.

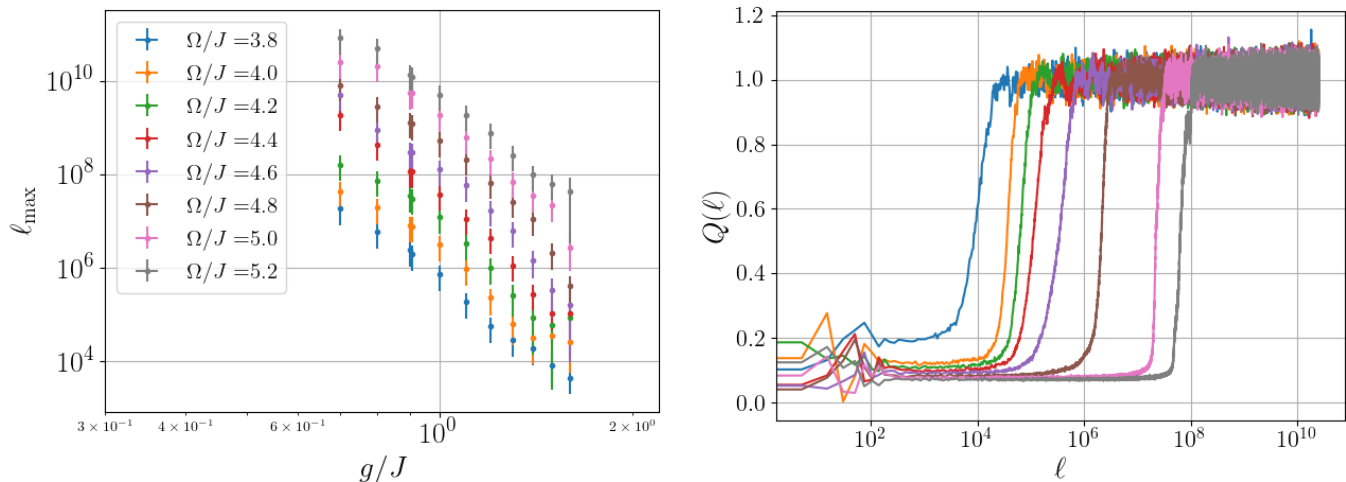


FIG. 14: The left panel shows the heating time as a function of  $g/J$  for different  $\Omega/J$  values for  $h/J = 0$ . The error bars are determined using a bootstrap method. There are 15 realizations per data point. Note that the distance between two points of different frequency for the same  $g$  value is approximately constant on logarithmic scale. The right panel shows the energy absorption as a function of time for  $g/J = 1.6$ .

## VII. DEPENDENCE OF $g/J$ COUPLING ON THERMALISATION TIMESCALE

Our model is ergodic as long as the  $g/J$  dimensionless parameter is non-zero, when  $g/J = 0$  the model becomes the standard Ising model and is exactly solvable. The parameter  $h/J$  is irrelevant for the physics of pre-thermalization since  $H_F^{(0)}$  is non-integrable even for  $h/J = 0$  whenever the on-site spin degree of freedom  $s > 1/2$  [for  $s = 1/2$  one has the transverse-field Ising model which is integrable]. The right panel of Figure 14 shows the heating time as a function of  $g/J$  parameter when  $h/J = 0$ . Interestingly, the spacing between different  $\Omega/J$  values for fixed  $g/J$  is approximately constant on a log-log scale meaning that for  $h/J = 0$  the heating time can be parametrized as  $\ell_{\max}(\Omega) = r(g/J) \exp[-\gamma(g/J) \times \Omega/J]$ , with  $\gamma(g/J)$  a slowly-varying function of  $g/J$ , and  $r(g/J) \propto (g/J)^\alpha$  with numerically determined exponent  $\alpha \approx -2.12$ . This would suggest that for fixed  $g/J$  the exponential scaling (with the possibility of logarithmic corrections) is a *universal* property. The left panel of Figure 14 shows that for  $g/J$  greater than 1.7 the prethermal plateau is no longer present for low frequencies and the system begins to absorb energy immediately. This possibly explains why data for large  $g/J$  the data does not have the same clean linear relationship on log-log scale as seen for low  $g/J$  values. Simulating  $g/J$  values less than 0.6 is not feasible due to the divergence in heating time. This constrains the available window of meaningful data to  $0.7 \leq g/J \leq 1.6$

## VIII. MONO- VS. POLYCHROMATIC DRIVES. THE EFFECT OF MICROMOTION.

Throughout this study we used a time-periodic step drive function. This is a key property which ultimately allows to reach astronomically long times required to study pre-thermalization. However, typically in experiments it is easier to apply periodic modulations with few harmonics, while step drives require a more precise apparatus. One may, therefore, wonder how the heating properties of the system might change, if we used few-harmonic drives instead.

Here, we give evidence that the form of the drive is relatively unimportant when considering the qualitative behavior of thermalizing dynamics. Let us cast the Hamilton function in the generalized form

$$H(t) = \frac{1}{2}(A + B) + \frac{f(t)}{2}(A - B), \quad (18)$$

see Eq. (12). For a step drive, we can Fourier-decompose the protocol as follows

$$f(t) = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{1}{2n-1} \sin((2n-1)\Omega t) = \begin{cases} +1 & \text{for } t \in [0, T/2] \text{ mod } T \\ -1 & \text{for } t \in [T/2, T] \text{ mod } T \end{cases} \quad (19)$$

Note that each term of the expansion of  $f(t)$  has zero average over a period. Thus, no matter the number of terms at which the series is truncated,  $H_F^{(0)} = \frac{1}{2}(A + B)$ .

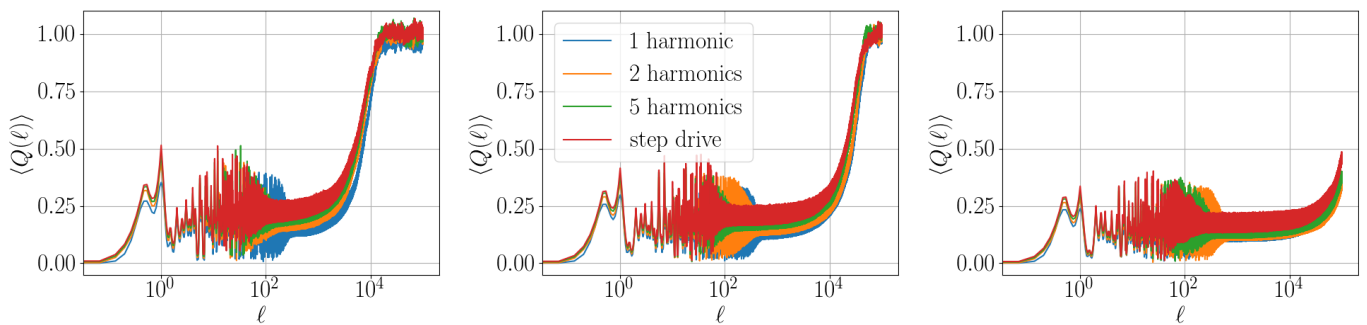


FIG. 15: Average energy absorption  $\langle Q(t) \rangle$  as a function of the number of driving cycles  $\ell$  for different truncations of  $f(t)$ . The graphs shown have  $\Omega/J = 3.0$  (left),  $\Omega/J = 3.2$  (middle) and  $\Omega/J = 3.4$  (right). The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ . This plot shows 15 points per period going beyond stroboscopic dynamics.

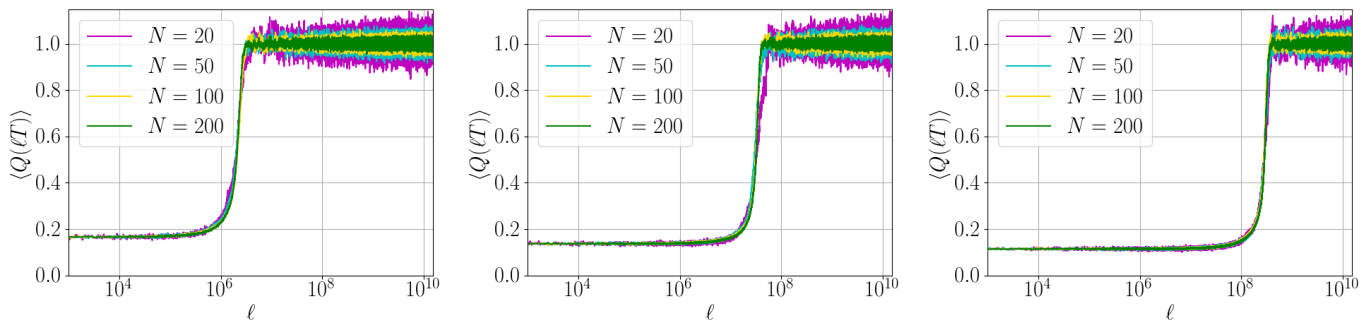


FIG. 16: Average energy absorption  $\langle Q(t) \rangle$  as a function of the number of driving cycles  $\ell$  for system sizes  $N = 20, 50, 100, 200$ . The graphs shown have driving frequencies  $\Omega/J = 3.8$  (left),  $\Omega/J = 4.2$  (middle) and  $\Omega/J = 4.6$  (right). The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ . Note how the fluctuations decrease with system size while the steepness increases with system size.

Figure 15 shows a comparison of the energy absorption with  $f(t)$  truncated at different orders in the Fourier expansion. The graphs show that the presence of higher order Fourier modes only marginally increases the rate of heating throughout the frequency range of interest. Hence, the picture developed in the main text about the properties of the thermalizing dynamics remains robust to the number of harmonics the driving protocol is composed of. Furthermore, this graphs the behavior for all times not just stroboscopically. This suggests that irrespective of the phase of the drive (i.e. the Floquet gauge), the qualitative behavior of dynamics remains the same. We used QuSpin [81, 82] to simulate the nonlinear EOM.

## IX. SYSTEM SIZE DEPENDENCE

Figure 16 shows that the timescale at which the prethermal plateau ends and thermalisation occurs is insensitive to the system size. However, the steepness of this heating process does increase with increasing the system size. Additionally, in the infinite-temperature state [stage (iv), cf. Fig. 1, main text] reaching larger system sizes reduces the size of the temporal fluctuations, as anticipated from the laws of Statistical Mechanics.

## X. STABILITY OF THE NUMERICAL INTEGRATION OF THE HAMILTON EQUATIONS

The Hamilton EOM for our model can be solved using an analytic recursion relation during in each half period. This means that the stroboscopic evolution of the system is described by an iterative application of a discrete nonlinear map [see main text]. Iterative recursive relations are easier to simulate numerically than a system of time-dependent non-linear differential equations.

Nevertheless, when we deal with times as large as  $10^{10}$  driving cycles, it is imperative to ensure that the numerical

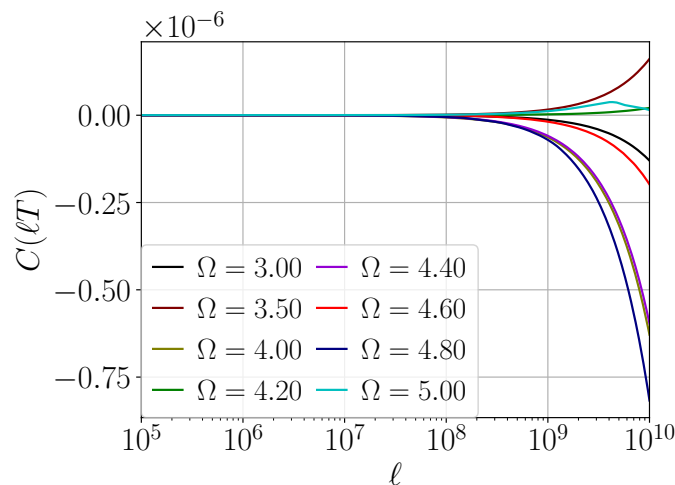


FIG. 17: The maximum error of the Casimir operator  $C(\ell T)$  as a function of the number of driving cycles  $\ell$  for different  $\Omega/J$  values. The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ .

error growth is under control at all times. To test it, recall that the spin algebra requires that the sum of the squared on-site magnetizations

$$C_j(\ell T) = |\vec{S}_j(\ell T)|^2 \quad (20)$$

is an integral of motion. In particular, it is fixed to unity for every lattice site and at all times. The origin of this integral of motion can be traced back to the Casimir operator of the rotor algebra.

The relative deviation of  $C_j(\ell T)$  from unity represents a reliable measure of how large the numerical integration error is. Since the squared magnetization is defined for each lattice site, an absolute value for the error is given by the maximum deviation over the entire chain:

$$C(\ell T) = \max_j |C_j(\ell T) - 1|. \quad (21)$$

Figure 17 shows the time-dependence of  $C(\ell T)$ . One can see that the maximum error is no bigger than  $10^{-5}$  at the end of the time evolution. This is a good indicator that the numerical simulation is sufficiently accurate to render our analysis reliable. Indeed, note that (exponential) blow up of the curves due to numerical error happens at times order of magnitudes larger than stage (iii) [see Fig. 1, main text], during which unconstrained thermalization to an infinite-temperature state occurs for a given frequency. Nevertheless, the pre-factor is so small that this blow-up is under control for all times of interest.

For comparison, the average error  $N^{-1} \sum_{j=1}^N |C_j(\ell T) - 1| \sim 10^{-10}$  is a few orders of magnitude smaller at  $\ell = 10^{10}$ , due to self-averaging effects.

## XI. DEPENDENCE ON THE SIZE OF THE NOISE DISTRIBUTION FOR THE INITIAL STATE

As we explained in the main text, in order to make our results robust to the chaotic character of many-body dynamics, we add small noise in the azimuthal angle of the initial antiferromagnetic state. By studying the long time behavior of energy absorption, here we show that the size of the support of the noise distribution is irrelevant to the qualitative picture of the thermalizing dynamics, provided the energy of the initial state remains unaffected. Figure 18 shows the time-dependence of the energy absorption for three different values of the support of the uniform noise distribution:  $[-\pi/W, \pi/W]$ , with  $W \in \{50, 150, 500\}$ .

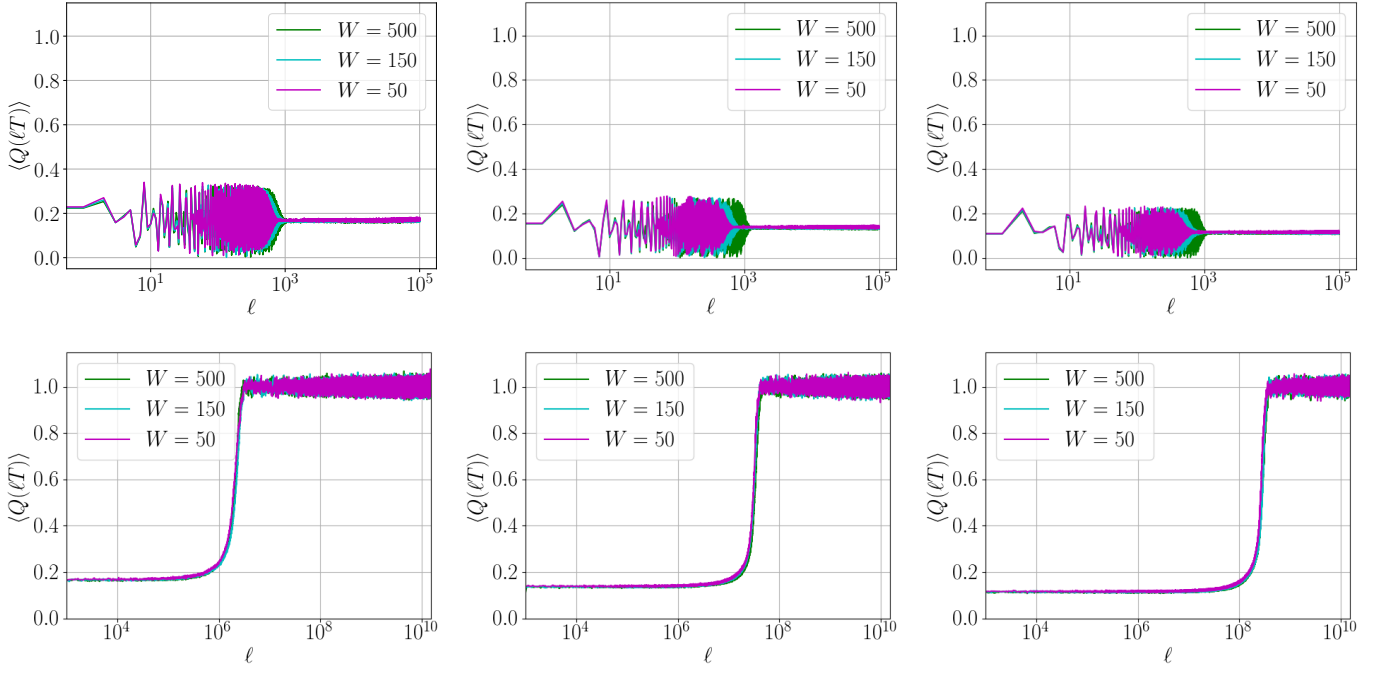


FIG. 18: Average energy absorption  $\langle Q(t) \rangle$  as a function of the number of driving cycles  $\ell$ , for  $\Omega/J = 3.8$  (left),  $\Omega/J = 4.2$  (middle) and  $\Omega/J = 4.6$  (right). The above panels show the initial fluctuations due to two particle behavior and the lower show behavior for long times. The  $W$  value corresponds to the strength of the initial noise. The simulation parameters are  $h/J = 0.809$ ,  $g/J = 0.9045$ , and  $N = 100$ .