Nonlinear Perturbation of Random Matrix Theory

Klaus M. Frahm and Dima L. Shepelyansky
Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

(Received 22 December 2022; revised 17 May 2023; accepted 25 July 2023; published 16 August 2023)

We consider a system of linear oscillators, or quantum states, described by random matrix theory and analyze how its time evolution is affected by a nonlinear perturbation. Our numerical results show that above a certain chaos border a weak or moderate nonlinearity leads to a dynamical thermalization of a finite number of degrees of freedom with energy equipartition over linear eigenmodes as expected from the laws of classical statistical mechanics. The system temperature is shown to change in a broad range from positive to negative values, and the dependence of system characteristics on the initial injected energy is determined. Below the chaos border the dynamics is described by the Kolmogorov-Arnold-Moser integrability. Owing to universal features of random matrix theory we argue that the obtained results describe the generic properties of its nonlinear perturbation.

DOI: 10.1103/PhysRevLett.131.077201

In far 1872, 150 years ago, Boltzmann developed the theory of statistical mechanics and thermalization originated from the dynamical laws of classical motion of many-body systems [1]. This result led to the famous Boltzmann-Loschmidt dispute on a possibility of thermalization and time irreversibility emerging from the reversible dynamical equations of particle motion [2,3] (see also Ref. [4]). The modern resolution of this dispute is based on the theory of dynamical chaos for generic nonlinear systems characterized by a positive maximal Lyapunov exponent and Kolmogorov-Sinai entropy leading to an exponential instability of motion (see, e.g., Refs. [5–8]). This instability leads to an exponential growth of errors which breaks time reversibility (see, e.g., an example in Ref. [9]).

The first numerical studies of how ergodicity, dynamical thermalization, and energy equipartition appear in a homogeneous 1D oscillator chain perturbed by a moderate nonlinearity were reported by Fermi, Pasta, Ulam in 1955 [10] (see system Hamiltonian in the Supplemental Material [11]). The conclusion was that "The results show very little, if any, tendency toward equipartition of energy between the degrees of freedom" [10]. It was argued in Ref. [18] that in the continuum limit the Fermi-Pasta-Ulam (FPU) problem is close to the Korteweg-de Vries equation with stable soliton solutions shown to be completely integrable [19], as well as the nonlinear Schrödinger equation [20]. In addition, at weak nonlinearity the FPU α model is close to the completely integrable Toda lattice [21,22]. Another explanation of equipartition absence in the FPU problem was given in Refs. [23–25] showing that below a certain strength of nonlinear interactions between oscillator modes the system is located in the regime of Kolmogorov-Arnold-Moser (KAM) integrability, and only above this border an overlap of nonlinear resonances takes place with the emergence of chaos and thermalization. Numerical simulations demonstrated a dynamical thermalization with energy equipartition reported in Refs. [24,25]. Thus, even 50 years after [10], various regimes of nonlinear dynamics of the FPU problem are actively discussed by the community of dynamical systems [26] (see, e.g., recent Ref. [27]). The variety of studies clearly demonstrates that this model played an important role in the investigations of nonlinear dynamics but also that it has multiple specific features indicating that it does not belong to a class of generic oscillator systems with nonlinear interactions.

To construct a generic model of many-body oscillator systems with nonlinear interactions between oscillators we take insight from quantum mechanics of many-body systems whose spectral properties are described by random matrix theory (RMT) invented by Wigner for a description of the spectra of complex nuclei, atoms, and molecules [28]. At present RMT finds applications in multiple areas of physics [29,30] including systems of quantum chaos whose dynamics is chaotic in the classical limit [31,32]. The properties of RMT eigenvalues and eigenstates were established in various studies and are well known. The RMT eigenstates are ergodic, i.e., uniformly distributed on the N-dimensional unit sphere, and the level spacing statistics is described by the universal RMT distribution [28–32]. Owing to the linearity of the Schrödinger equation the time evolution of a wave function ψ described by a RMT Hamiltonian also describes a time evolution of a system of N linear oscillators with random linear couplings. On its own, because of the universal properties of RMT, it is interesting to understand how a nonlinear perturbation affects RMT evolution.

With the aim of understanding the effects of nonlinear perturbation of RMT, we consider a simple model

described by the Schrödinger equation with a Hamiltonian given by a random matrix with an additional nonlinear interaction between linear modes:

$$i\hbar\frac{\partial\psi_n(t)}{\partial t}=\sum_{n'=1}^{N}H_{n,n'}\psi_{n'}(t)+\beta|\psi_n(t)|^2\psi_n(t). \quad (1)$$

Here $H_{n,n'}$ are elements of an RMT matrix $\hat{H}$ of size N taken from the Gaussian orthogonal ensemble (GOE) [29]; they have zero mean and variance $\langle H_{n,n'}^2\rangle=(1+\delta_{n,n'})/[4(N+1)]$. The averaged density of states is given by the semicircle law $dm/dE=(2N/\pi)\sqrt{1-E^2}$ with typical eigenvalues in the interval $E_m\in[-1,1]$ (we use dimensionless units with $\hbar=1$), and β is a dimensionless constant characterizing the nonlinear interaction strength in the original basis n.

The eigenmodes of $\hat{H}$ at energies E_m are $\phi_n^{(m)}$ which are ergodic with a uniform distribution on the N-dimensional unit sphere. The time evolution of the wave function can be expressed in the basis of eigenmodes as $\psi_n(t)=\sum_{m=1}^{N}C_m(t)\phi_n^{(m)}$ with coefficients $C_m(t)$ giving the occupation probability $\rho_m=\langle|C_m(t)|^2\rangle$ (with some long time or ensemble average; see below). The time evolution [Eq. (1)] has two integrals of motion being the probability norm $\sum_n|\psi_n(t)|^2=1$ and total energy $E=\sum_n[\langle\psi_n(t)|\hat{H}|\psi_n(t)\rangle+(\beta/2)|\psi_n(t)|^4]$. At $\beta=0$ the model [Eq. (1)] can be viewed as a quantum system or as a classical system of coupled linear oscillators whose Hamiltonian in the basis of oscillator eigenmodes is $\mathcal{H}=\sum E_mC_m^*(t)C_m(t)$ where C_m, C_m^* is a pair of conjugated variables and E_m plays the role of oscillator frequencies. Since RMT captures the universal features of quantum and linear oscillator systems we expect that the model [Eq. (1)] describes the universal properties of oscillator systems with chaotic dynamics induced by weak or moderate nonlinear couplings between oscillators. We call the model [Eq. (1)] the nonlinear random matrix model (NLIRM).

Above a certain chaos border with $\beta>\beta_c$ a moderate nonlinearity destroys KAM integrability leading to chaotic dynamics with a positive maximal Lyapunov exponent λ. The nonlinear frequency shift is $\delta\omega\sim\beta|\psi_n|^2\sim\beta/N$ and, as it was argued in Refs. [33–36], a developed chaos takes place when this shift $\delta\omega$ becomes comparable to a typical energy spacing between energies (or frequencies) of the linear system $\Delta\omega\sim 1/N$. Thus $\delta\omega>\Delta\omega$ implies chaos with the chaos border $\beta_c=\text{const}\sim 1$ being independent of system size N.

The issue of dynamical thermalization in finite size nonlinear lattices with disorder was studied in Refs. [36,37]. The time evolution in these systems is described by the discrete Anderson nonlinear Schrödinger equation (DANSE) with hopping between nearby sites (see DANSE in the Supplemental Material [11]). In the linear case the disorder leads to Anderson localization of modes [38] which is well visible when the localization length ℓ is smaller than the system size N. In this respect our RMT model [Eq. (1)] is rather different since the linear modes are delocalized and ergodic in a vector space of dimension N. We expect that our model [Eq. (1)] is generic and captures also certain features of the models of the Bose-Einstein condensate evolution in the chaotic Bunimovich stadium [39] or the Sinai oscillator [40] described by the nonlinear Gross-Pitaevskii equation (GPE) [41] (see GPE in the Supplemental Material [11]). Indeed, the linear eigenmodes of these systems have properties of quantum chaos similar to RMT [31,32]. There are however also certain differences discussed below.

For the GPE models [39,40] it is natural to assume that the dynamical thermalization induced by moderate nonlinearity leads to the Bose-Einstein (BE) distribution of probabilities ρ_m over quantum levels of the linear system. In the limit of high temperature T this distribution is reduced to a classical energy equipartition (EQ) distribution [4,42]. For the DANSE type models [36,37] the quantum Gibbs (QG) distribution was proposed to explain numerically obtained results. In fact QG and BE distributions give very close thermalization properties, and we mainly discuss the BE case here. Thus there are two options for the thermalized distributions of probabilities ρ_m:

$$\rho_m=\frac{1}{\exp[(E_m-\mu)/T]-1}\,(\text{BE}),\qquad \rho_m=\frac{T}{E_m-\mu}\,(\text{EQ}). \quad (2)$$

Here T is the system temperature and $\mu(T)$ is the chemical potential dependent on temperature. The parameters T and μ are determined by the norm and energy conservation $\sum_m\rho_m=1$ and $\sum_m E_m\rho_m=E$ (for E we assume the case of weak or moderate nonlinearity which gives only a weak contribution to the total energy). The entropy S of the system is determined by the usual relation [4,42]: $S=-\sum_m\rho_m\ln\rho_m$ with the implicit theoretical dependencies on temperature $E(T), S(T), \mu(T)$. The derivation of Eq. (2) is given in the Supplemental Material [11].

Based on classical statistical mechanics [4,42] the dynamical thermalization should lead to the EQ distribution [Eq. (2)] since the DANSE, GPE [36,37,39,40], and NLIRM [Eq. (1)] models describe classical nonlinear fields without second quantization. In contrast, in Refs. [36,37,39,40] it was argued that a moderate nonlinearity plays the role of an effective nonlinear thermostate that leads to quantum BE or QG distributions [Eq. (2)].

Of course, both BE and EQ approaches [Eq. (2)] give different thermal characteristics leading to a contradiction discussed in detail in Refs. [37,39,40]. The main argument in favor of the BE (or the QG) ansatz was based on a reasonably good agreement of numerical data for entropy vs energy with the theoretical thermal dependence $S(E)$

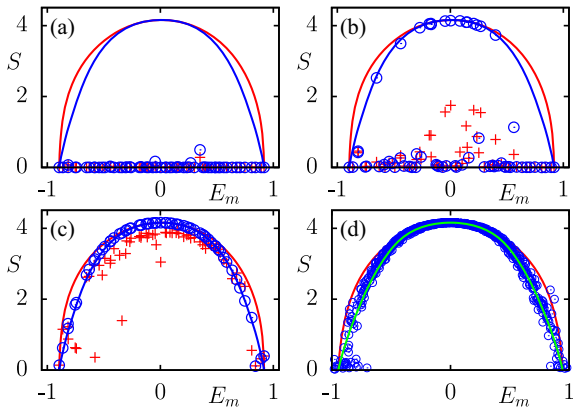

FIG. 1. Entropy S versus energy E_m of the initial state m at $t = 0$ for one RMT realization at $N = 64$ and $\beta = 0.02$ (a), $\beta = 0.1$ (b), and $\beta = 1$ (c) or 10 RMT realizations at $\beta = 1$ (d). The entropy S is computed from ρ_m averaged over the time range $2^{23} \leq t \leq 2^{24}$ (blue-black $\circ$) and $2^{16} \leq t \leq 2^{17}$ [red-gray + in (a), (b)] or $2^{11} \leq t \leq 2^{12}$ [red-gray + in (c)]. The theory curves $S(E)$ for BE (red-gray) and EQ [blue-black in (a),(b),(c) or green in (d)] are from ρ_m values of [Eq. (2)] with E_m values of the used RMT realization (a), (b),(c) or a fictitious spectrum according to the semicircle law in (d) [where E_m is the solution of $m - 1/2 = M(E_m)$, $m = 1, \ldots, N$ with $M(E)$ being the integrated density of states].

given by the BE (or QG) ansatz. The quantities S and E are extensive (self-averaging), and it was argued that their analysis is more preferable as compared to the direct study of the strongly fluctuating probabilities ρ_m [36,37,39,40]. Here we show that the ergodicity of RMT eigenstates of $\hat{H}$ allows one to reduce significantly the fluctuations and to obtain stable results for ρ_m that are clearly described by the EQ ansatz [Eq. (2)].

The numerical integration of [Eq. (1)] is done with the symplectic scheme of order 4 [43–45] using a step size $\Delta t = 0.1$ up to maximal times $t = 4 \times 10^6 - 1.3 \times 10^8$ with exact norm conservation, energy conservation with accuracy $\sim 10^{-8}$, and for the GOE matrix size $N = 64$ (see the Supplemental Material [11] for more details and results for other values $N = 32$, 128, 256, 512). As initial condition, we choose an eigenmode $\phi_n^{(m)}$ of $\hat{H}$ at some index m (sometimes also noted m_0) such that the energy remains close to the initial energy $E \approx E_m$. Examples of the time dependence $S(t)$ are shown in the Supplemental Material [11], Fig. S1, demonstrating a steady-state regime reached at times $t > 10^4$ for $\beta = 1$. The obtained dependence $S(E)$ is shown in Fig. 1 at different β values for a specific RMT realization and two timescales and also for 10 RMT realizations at $\beta = 1$. At small values $\beta = 0.02$, 0.1 the system is close to an integrable KAM regime [7,8] while at $\beta = 1$ essentially all modes are thermalized (see Fig. 1, Fig. S2 in the Supplemental Material [11], and additional material in Ref. [46]). These results show that the critical border for thermalization is located at $\beta_c \sim 0.1$

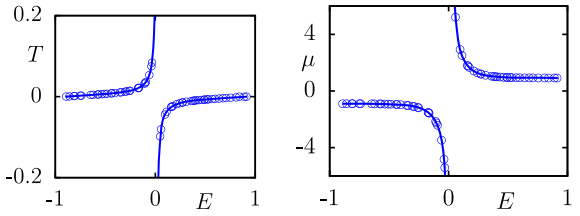

FIG. 2. Dependence of T and μ on energy E for EQ ansatz [Eq. (2)] (curves); data points are for $\beta = 1$, $N = 64$, and time range $2^{23} \leq t \leq 2^{24}$, with T and μ determined from norm and numerical entropy values [same RMT realization as in Fig. 1(c)].

independent of N. However, the exact determination of β_c is a rather complicated task due to the presence of many-body nonlinear effects like, e.g., the Arnold diffusion [7,8]. Also at the spectral borders $E \approx \pm 1$ the spacing between energies E_m increases according to the semicircle law [29], and therefore it is more difficult to reach thermalization there.

An important feature of Fig. 1 is that the theory curves $S(E)$ obtained with the BE and the EQ ansatz [Eq. (2)] are rather close to each other. Thus due to fluctuations of numerical data for $S(E)$ it is difficult to determine which theory BE or EQ better describes the numerical data. However, the data points are significantly closer to the BE curve, especially for moderate energies $|E| \approx 0.5$–0.8 where both curves are somewhat different [the difference between the QG and BE $S(E)$ curves, not visible on graphical precision, is ~ 0.003 at the spectral borders and much smaller at other E values, so that we discuss mainly the BE case].

For the EQ ansatz the dependencies $T(E)$, $\mu(E)$, obtained by the solution of the equations for energy and norm for a given RMT spectrum, are shown in Fig. 2 (the Supplemental Material [11], Fig. S3, for the BE ansatz) for the thermalized regime at $\beta = 1$. The numerical points obtained from E and norm values are by definition exactly located on the theory curves. If instead of E we use the numerical data of S then the points slightly deviate from the theory (Fig. 2 and Fig. S3 in the Supplemental Material [11]), but the T and μ values themselves are drastically different between BE and EQ cases.

The most direct way to distinguish between BE and EQ cases is to compare the probability dependence $\rho_m(E)$ with the theory [Eq. (2)]. Such a comparison is shown in Fig. 3 for four initial states at $m = m_0$, $\beta = 1$, and $N = 64$ (more data are in the Supplemental Material [11], Fig. S4, and Ref. [46]). The dynamical thermalization clearly follows the EQ ansatz and not at all the BE one, except for an initial state at $E_{m_0} \approx 0$ where both approaches are equivalent. This observation is in agreement with the classical statistical mechanics [4,42]. The probabilities ρ_m for all initial energies E_{m_0} are shown in Fig. 4 with a good agreement between the numerical data and the EQ ansatz (see Ref. [46] for figures such as Fig. 3 for all m_0 values).

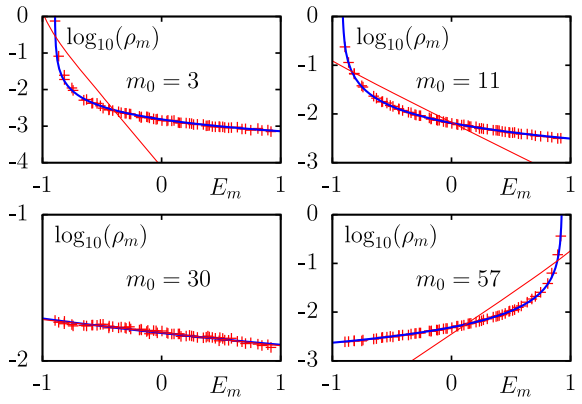

FIG. 3. Dependence of $\rho_m(E_m)$ on E_m for four initial states at $m_0 = 3$, 11, 30, and 57 with negative temperature $T < 0$; here $\beta = 1$, $N = 64$, and time average range $2^{23} \leq t \leq 2^{24}$. The blue curve shows the theory of EQ ansatz with $\rho_{\rm EQ}(E) = T/(E-\mu)$. The red line shows BE ansatz theory $\rho_{\rm BE}(E) = 1/(\exp[(E-\mu)/T]-1)$; T, μ theory [Eq. (2)] values are given in the Supplemental Material [11], Fig. S4, for BE and EQ cases.

The statistical distribution $p(x)$ of fluctuations of the rescaled quantity $x = (E_{m_0} - \mu)|C_m(t)|^2/T$ (with μ, T from the EQ ansatz for the energy E_{m_0}) also follows the Boltzmann law $p(x) = \exp(-x)$ (see the Supplemental Material [11], Fig. S5).

In Fig. 5 we show the energy dependence of the maximal positive Lyapunov exponent λ_m on energy E_m of initial state m for different β values (more data are in the Supplemental Material [11], Figs. S6–S10, and Ref. [46]). In the thermalized phase $\beta = 1$ we have a smooth variation of λ with E_m while below or close to the thermalization border at $\beta = 0.1$ high λ_m values appear only at specific E_m values. We attribute this to the existence of triplets of energies with very close E_m values. Indeed, in a hypothetic case of three equal E_m values the KAM theory is not valid and developed chaos exists at arbitrary small β values as is shown in Refs. [33,47]. Nonetheless, in RMT there is level repulsion, and double or triple degeneracies are forbidden leaving place only to quasidegeneracy of levels so that KAM becomes valid at $\beta \to 0$. Thus for $\beta = 0.02$ we have typically λ_m approaching to zero with increasing time. Our preliminary results show that in the thermal phase at larger $|T|$ (if $E_m \approx 0$) we have an approximate dependence $\lambda \sim \beta^\eta/N^\nu$ with $\eta \approx 1.52$, $\nu \approx 1.89$ (see the Supplemental Material [11], Figs. S6–S10). However, the Lyapunov exponent dependence on β and N requires further detailed studies.

Finally, we discuss the reasons why the nature of thermal equipartition, BE or EQ, was so difficult to establish in previous studies [36,37,39,40]. One of them is the proximity of $S(E)$ curves for both approaches. At the same time the direct determination of the $\rho_m(E)$ dependence is rather difficult due to significant fluctuations, as it was pointed out previously. These fluctuations are especially large for the DANSE case at a large disorder ($W = 4$ in Ref. [36]) when the localization length ℓ is significantly smaller than system size N ($\ell/N \approx 0.1$ at $N = 64$). We illustrate this in the Supplemental Material [11], Figs. S11–S12, showing that at smaller disorder $W = 2$ with larger localization length ℓ the fluctuations of ρ_m are reduced and at long times we have an agreement of $\rho_m(E)$ with the EQ ansatz and strong deviations from the BE ansatz. For the NLIRM model [Eq. (1)] the linear eigenmodes are ergodic, i.e., no localization, and the fluctuations of $\rho_m(E)$ are significantly reduced, which allows one to distinguish clearly between EQ and BE cases.

The cases of GPE in the Bunimivich stadium [39] and the Sinai-oscillator trap [40] are somewhat different. Indeed, in these models the spectrum of the linear system is unbounded so that, even if linear eigenstates are in the quantum chaos regime, the probability spreading to high energies is rather slow due to small coupling transitions induced by nonlinearity between states with significantly different energies. Thus in these systems there is a

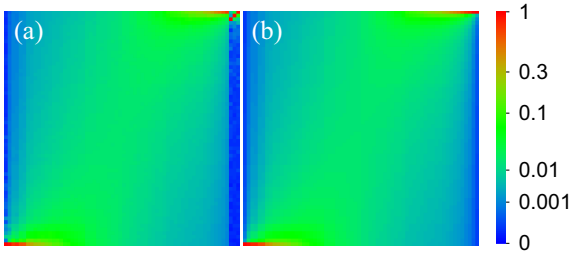

FIG. 4. Density plot of ρ_m for parameters of Fig. 3 with initial state index $1 \leq m_0 \leq 64$ in the x axis and $1 \leq m \leq 64$ in the y axis. The colorbar shows ρ_m values in a nonlinear scale to increase the visibility of small ρ_m values. (a) Numerical data for $\beta = 1$, $N = 64$. (b) The EQ ansatz $\rho_{\rm EQ}(E_m)$ (see also Fig. 3).

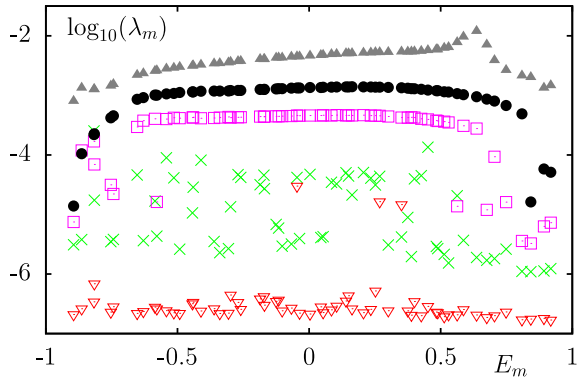

FIG. 5. Lyapunov exponent λ_m dependence on E_m with m being the index of the initial state for $N = 64$; λ_m is determined from the fit $\ln\|\Delta\psi(t)\| = a + b\ln(t) + \lambda_m t$ for $\beta = 2$ (gray triangle, top); 1 (black circle); 0.5 (pink square); 0.1 (green cross) at $t \leq 2^{22}$; $\beta = 0.02$ for $t \leq 2^{27}$ (red $\bigtriangledown$, bottom).

formation of a relatively compact probability packet at low energies which spreads to high energies very slowly in time. Such an energy packet of ρ_m gives $S(E)$ values compatible with the curve of the BE ansatz; however the fluctuations of $\rho_m(E)$ are very strong with a significant difference from the BE distribution at high energies (see, e.g., Fig. 5 in Ref. [39] and Figs. 8 and 11 in Ref. [40]). To analyze these features in more detail, we add to the diagonal RMT matrix element $H_{n,n}$ an additional diagonal energy fn with a constant $f > 0$. Then the variation of linear energies fN becomes rather large and exceeds significantly those of the RMT case. The results for this model at $\beta = 1$, $f = 0.25$ show that at times $t = 2^{15}$ for $N = 32$ (or $t = 2^{20}$ for $N = 64$) the probabilities $\rho_m(E)$ form a compact packet of approximate BE shape and the EQ thermal distribution is reached (with fluctuations) only at very large times $t = 2^{27}$ (see the Supplemental Material [11], Figs. S13 and S14). Such large timescales were out of reach in Refs. [39,40] due to the complexity of the numerical integration of GPE.

In conclusion, we showed that a nonlinear perturbation of RMT leads to dynamical thermalization with energy equipartition corresponding to the laws of classical statistical mechanics [4,42]. Such a thermalization appears due to dynamical chaos in finite systems with a moderate or large number of degrees of freedom at weak or moderate perturbation of a linear RMT system. At very weak perturbations the system dynamics is characterized by a quasi-integrable KAM regime. We argue that the proposed NLIRM model captures the generic features of dynamical thermalization in systems weakly perturbed by classical nonlinear fields and does not depend on the specific form of the nonlinear term (see detailed discussion in the Supplemental Material [11] and Figs. S15 and S16 there). Of course, for finite many-body quantum systems with second quantization the interactions lead to quantum dynamical thermalization and distributions of Bose-Einstein for bosons or Fermi-Dirac for fermions, as has been demonstrated in numerical studies [48] and Refs. [49–51] respectively.

This work has been partially supported through the Grant NANOX No ANR-17-EURE-0009 in the framework of the Programme Investissements d'Avenir (project MTDINA). This work was granted access to the HPC resources of CALMIP (Toulouse) under the allocation 2022-P0110.

Note added.—After submission of this work a dynamical thermalization at negative temperature in EQ [Eq. (2)] has been observed in optical fibers [52]. See also the discussion in the Supplemental Material [11].

[1] L. Boltzmann, Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen, Wiener Ber. **66**, 275 (1872).

[2] J. Loschmidt, Über den Zustand des Wärmegleichgewichts eines Systems von Körpern mit Rücksicht auf die Schwerkraft, Sitzungsber. Akad. Wiss. Wien Math. Naturwiss. Kl. Abt. 1 **II 73**, 128 (1876).

[3] L. Boltzmann, Über die Beziehung eines allgemeine mechanischen Satzes zum zweiten Haupsatze der Wärmetheorie, Sitzungsber. Akad. Wiss. Wien Math. Naturwiss. Kl. Abt. 1 **II 75**, 67 (1877).

[4] J. E. Mayer and M. Goeppert-Mayer, *Statistical Mechanics* (John Wiley & Sons, New York, 1977).

[5] V. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).

[6] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory* (Springer-Verlag, New York, 1982).

[7] B. V. Chirikov, A universal instability of many-dimensional oscillator systems, Phys. Rep. **52**, 263 (1979).

[8] A. Lichtenberg and M. Lieberman, *Regular and Chaotic Dynamics* (Springer, New York, 1992).

[9] D. L. Shepelyansky, Some statistical properties of simple classically stochastic quantum systems, Physica (Amsterdam) **8D**, 208 (1983).

[10] E. Fermi, J. Pasta, and S. Ulam, Studies of Non Linear Problems, Los Alamos Report LA-1940, 1955; published later in E. Fermi *Collected Papers*, edited by E. Serge (University of Chicago Press, Chicago IL, 1965), 2, 491; see also historical overview in T. Dauxois Fermi, Pasta, Ulam and a mysterious lady, Phys. Today **61**, No. 1, 55 (2008).

[11] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.131.077201 for additional description, data, figures, and Refs. [12–17].

[12] P. Aschieri, J. Garnier, C. Michel, V. Doya, and A. Picozzi, Condensation and thermalization of classsical optical waves in a waveguide, Phys. Rev. A **83**, 033838 (2011).

[13] K. Baudin, A. Fusaro, K. Krupa, J. Garnier, S. Rica, G. Millot, and A. Picozzi, Classical Rayleigh-Jeans Condensation of Light Waves: Observation and Thermodynamic Characterization, Phys. Rev. Lett. **125**, 244101 (2020).

[14] E. V. Podivilov, F. Mangini, O. S. Sidelnikov, M. Ferraro, M. Gervaziev, D. S. Kharenko, M. Zitelli, M. P. Fedoruk, S. A. Babin, and S. Wabnitz, Thermalization of Orbital Angular Momentum Beams in Multimode Optical Fibers, Phys. Rev. Lett. **128**, 243901 (2022).

[15] F. Mangini, M. Gervaziev, M. Ferraro, D. S. Kharenko, M. Zitelli, Y. Sun, V. Couderc, E. V. Podivilov, S. A. Babin, and S. Wabnitz, Statistical mechanics of beam self-cleaning in GRIN multimode optical fibers, Opt. Express **30**, 10850 (2022).

[16] V. E. Zakharov, V. S. L'vov, and G. Falkovich, *Kolmogorov Spectra of Turbulence I* (Springer, Berlin, 1992).

[17] A. Altland and B. Simons, *Condensed Matter Field Theory* (Cambridge University Press, Cambridge UK, 2006), p. 58.

[18] N. J. Zabusky and M. D. Kruskal, Interaction of "Solitons" in a Collisionless Plasma and the Recurrence of Initial States, Phys. Rev. Lett. **15**, 240 (1965).

[19] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Method for Solving the Korteweg—de Vries Equation, Phys. Rev. Lett. **19**, 1095 (1967).

[20] V. E. Zakharov and A. B. Shabat, Interaction between solitons in a stable medium, Sov. Phys. JETP **37**, 823 (1973).

[21] M. Toda, Studies of a non-linear lattice, Phys. Rep. **18**, 1 (1975).
[22] G. Benettin, H. Christodoulidi, and A. Ponno, The Fermi-Pasta-Ulam problem and its underlying integrable dynamics, J. Stat. Phys. **152**, 195 (2013).
[23] B. V. Chirikov and F. M. Izrailev, Statistical properties of a non-linear string, Sov. Phys. Dokl. **11**, 30 (1966).
[24] B. V. Chirikov, F. M. Izrailev, and V. A. Tayursky, Numerical experiments on statistical behavior of dynamical systems with a few degrees of freedoms, Comput. Commun. Phys. **5**, 11 (1973).
[25] R. Livi, M. Pettini, S. Ruffo, and A. Vulpiani, Chaotic behavior in nonlinear Hamiltonian systems and equilibrium statistical mechanics, J. Stat. Phys. **48**, 539 (1987).
[26] *The Fermi-Pasta-Ulam Problem: A Status Report*, Lecture Notes in Physics Vol. 728, edited by G. Gallavotti (Springer, Berlin, 2008).
[27] M. Gallone, M. Marian, A. Ponno, and S. Ruffo, Burgers Turbulence in the Fermi-Pasta-Ulam-Tsingou Chain, Phys. Rev. Lett. **129**, 114101 (2022).
[28] E. P. Wigner, Random matrices in physics, SIAM Rev. **9**, 1 (1967).
[29] M. L. Mehta, *Random Matrices* (Elsvier, Amsterdam, 2004).
[30] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, Random matrix theories in quantum physics: Common concepts, Phys. Rep. **299**, 189 (1998).
[31] O. Bohigas, M.-J. Giannoni, and C. Schmit, Characterization of Chaotic Quantum Spectra and Universality of Level Fluctuation Laws, Phys. Rev. Lett. **52**, 1 (1984).
[32] F. Haake, *Quantum Signatures of Chaos* (Springer, Berlin, 2010).
[33] B. V. Chirikov and D. L. Shepelyanskii, Dynamics of some homogeneous models of classical Yang-Mills fields, Sov. J. Nucl. Phys. **36**, 908 (1982).
[34] D. L. Shepelyansky, Delocalization of Quantum Chaos by Weak Nonlinearity, Phys. Rev. Lett. **70**, 1787 (1993).
[35] I. Garcia-Mata and D. L. Shepelyansky, Delocalization induced by nonlinearity in systems with disorder, Phys. Rev. E **79**, 026205 (2009).
[36] M. Mulansky, K. Ahnert, A. Pikovsky, and D. L. Shepelyansky, Dynamical thermalization of disordered nonlinear lattices, Phys. Rev. E **80**, 056212 (2009).
[37] L. Ermann and D. L. Shepelyansky, Quantum Gibbs distribution from dynamical thermalization in classical nonlinear lattices, New J. Phys. **15**, 123004 (2013).
[38] P. W. Anderson, Absence of diffusion in certain random lattices, Phys. Rev. **109**, 1492 (1958).
[39] L. Ermann, E. Vergini, and D. L. Shepelyansky, Dynamical thermalization of Bose-Einstein condensate in Bunimovich stadium, Europhys. Lett. **111**, 50009 (2015).
[40] L. Ermann, E. Vergini, and D. L. Shepelyansky, Dynamics and thermalization of a Bose-Einstein condensate in a Sinai-oscillator trap, Phys. Rev. A **94**, 013618 (2016).
[41] L. Pitaevskii and S. Stringari, *Bose-Einstein Condensation* (Oxford University Press, Oxford, 2003).
[42] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Wiley, New York, 1976).
[43] E. Forest and R. D. Ruth, Fourth-order symplectic integration, Physica (Amsterdam) **43D**, 105 (1990).
[44] R. I. McLachlan and G. R. W. Quispel, Splitting methods, Acta Numer. **11**, 341 (2002).
[45] S. MacNamara and G. Strang, *Operator Splitting/ Splitting methods in Communication, Imaging, Science, and Engineering, Scientific Computation*, edited by R. Glowinski, S. Osher, and W. Yin (Springer, Cham, 2016), pp. 95-114.
[46] https://www.quantware.ups-tlse.fr/QWLIB/nonlinrmt/ (Accessed Dec 22, 2022).
[47] M. Mulansky, K. Ahnert, A. Pikovsky, and D. L. Shepelyansky, Strong and weak chaos in weakly nonintegrable many-body Hamiltonian systems, J. Stat. Phys. **145**, 1256 (2011).
[48] P. Schlageck and D. L. Shepelyansky, Dynamical thermalization in Bose-Hubbard systems, Phys. Rev. E **93**, 012126 (2016).
[49] A. R. Kolovsky and D. L. Shepelyansky, Dynamical thermalization in isolated quantum dots and black holes, Europhys. Lett. **117**, 10003 (2017).
[50] K. M. Frahm and D. L. Shepelyansky, Dynamical decoherence of a qubit coupled to a quantum dot or the SYK black hole, Eur. Phys. J. B **91**, 257 (2018).
[51] K. M. Frahm, L. Ermann, and D. L. Shepelyansky, Dynamical thermalization of interacting fermionic atoms in a Sinai-oscillator trap, MDPI Condens. Matter **4**, 76 (2019).
[52] K. Baudi, J. Garnier, A. Fusaro, N. Berti, C. Michel, K. Krupa, G. Millot, and A. Picozzi, Observation of Light Thermalization to Negative-Temperature Rayleigh-Jeans Equilibrium States in Multimode Optical Fibers, Phys. Rev. Lett. **130**, 063801 (2023).

Supplementary Material for

Nonlinear perturbation of Random Matrix Theory

by K. M. Frahm and D. L. Shepelyansky
Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

See also [46], corresponding to `https://www.quantware.ups-tlse.fr/QWLIB/nonlinrmt/`, (Accessed Dec 22, 2022), for additional and more detailed figures.

I. HAMILTONIANS AND EQUATIONS FOR FPU, GPE, DANSE MODELS

For convenience we give here the Hamiltonian of the Fermi-Pasta-Ulam (FPU) model and the time evolution equations for the Discrete Anderson Nonlinear Schrödinger Equation (DANSE) and the nonlinear Gross-Pitaevskii equation (GPE).

The Hamiltonian of the FPU model is:

$$H = \frac{1}{2}\sum_{n=0}^{N}[p_n^2 + (x_{n+1}-x_n)^2] + \frac{\alpha}{3}\sum_{n=0}^{N}(x_{n+1}-x_n)^3 + \frac{\beta}{4}\sum_{n=0}^{N}(x_{n+1}-x_n)^4$$

where the first term gives the Hamiltonian H_0 of the linear oscillator waves and the second and third terms represent the interaction $H_{\rm int}$. Here p_n, x_n are conjugated variables of momenta and coordinates. For $\beta = 0$ we have the α-FPU model and for $\alpha = 0$ we have the β-FPU model. The FPU model is described in detail in Refs.[10,16-20].

The time evolution of the DANSE model is described by the equation:

$$i\hbar\frac{\partial\psi_n}{\partial t} = \varepsilon_n\psi_n + \beta|\ \psi_n\ |^2\psi_n + V(\psi_{n+1}+\psi_{n-1})\ ,$$

where β determines the strength of the nonlinearity, V is the hopping matrix element, on-site energies ε_n are randomly distributed in the range $-W/2 < \varepsilon_n < W/2$, and the total probability is normalized to unity $\sum_n |\ \psi_n\ |^2 = 1$. Here we use units with $\hbar = V = 1$. For $\beta = 0$ and weak disorder all eigenstates are exponentially localized with the localization length $l \approx 96(V/W)^2$ at the center of the energy band. The DANSE model belongs to the same class of models described by Eq. (1) but with $H_{n,n'} = \varepsilon_n\delta_{n,n'} + V(\delta_{n,n'+1} + \delta_{n+1,n'})$ and the theoretical considerations of the next section also apply to the DANSE model assuming that the hypothesis of thermalization is verified. Detailed descriptions of the DANSE model are given in Refs. [34-37].

The GPE time evolution is described by

$$i\hbar\frac{\partial}{\partial t}\psi = \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \beta|\psi|^2\right)\psi\ .$$

We do not study this equation in this work. More detailes about its properties can be found in Ref. [41].

II. STATISTICAL CLASSICAL THEORY

The nonlinear Schrödinger equation (1) has two integrals of motion. By neglecting the energy of the weak nonlinear term $\sim \beta/N$ and assuming global chaos and ergodicity, we expect that the system becomes ergodic or thermalizes on the manifold fixed by the two constraints :

$$\sum_m E_m|C_m|^2 = E \quad , \quad \sum_m |C_m|^2 = 1$$

where C_m are the coefficients of the state in the expansion of the eigenbasis of the matrix $\hat{H}$. This situation corresponds in principle to a micro canonical ensemble with an additional constraint which is technically quite complicated. One can use $|C_1|^2 = 1 - \sum_{m=2}^{N}|C_m|^2$ to remove the first coordinate C_1 from the phase space to obtain a pure micro canonical ensemble for $C_2, \ldots, C_N$ with:

$$E - E_1 = \sum_{m=2}^{N}(E_m - E_1)|C_m|^2$$

but there is still the condition $\sum_{m=2}^{N}|C_m|^2 = 1 - |C_1|^2 \leq 1$ which creates technical complications. For small temperature or energy (with E being close to E_1, assuming an ordered eigenvalue spectrum $E_1 < E_2 < \ldots < E_N$) one can neglect this condition and in this case it is not difficult to show by standard text book techniques of statistical physics that in the limit $N \gg 1$ the marginal distribution of a field C_m (integrating out the other fields of the micro-canonical ensemble) is a (complex) Gaussian

$$p(C_m) \sim \exp\left(-\frac{(E_m - E_1)|C_m|^2}{T_{mc}}\right)$$

with the micro-canonical temperature $T_{mc} = (E-E_1)/N$ and providing the equipartition average : $\rho_{m,mc} = \langle |C_m|^2 \rangle = T_{mc}/(E_m - E_1)$.

However, for larger energies the additional inequality for the coefficients C_m cannot be neglected. Therefore, we treat the system as a grand-canonical ensemble, which is equivalent for $N \gg 1$. In this approach the fields C_m can freely fluctuate and the constraints are only verified in average. The classical grand canonical partition function is given by

$$\begin{aligned} Z &= \int \prod_m d^2 C_m \, \exp\left(-\frac{1}{T} \sum_m (E_m - \mu) |C_m|^2 \right) \\ &\sim T^N \prod_m \frac{1}{E_m - \mu} \quad \Rightarrow \\ \ln(Z) &= N \ln(T) - \sum_m \ln(E_m - \mu) + \text{const.} \end{aligned}$$

with two parameters being the (grand canonical) temperature T and the chemical potential μ which are determined by the implicit equations

$$1 = \sum_m \rho_m \quad , \quad E = \sum_m E_m \rho_m \tag{S.1}$$

with ρ_m being the statistical average :

$$\rho_m = \langle |C_m|^2 \rangle = \frac{T}{E_m - \mu} \equiv \rho_{EQ}(E_m) \ .$$

Here we have either $T > 0$ and $\mu < E_1$ or $T < 0$ and $\mu > E_N$ in order to have well defined Gaussian integrals in the partition function and only solutions for T and μ satisfying this condition are valid. From

$$\begin{aligned} E - \mu &= \left\langle \sum_m (E_m - \mu) |C_m|^2 \right\rangle = T^2 \frac{\partial \ln(Z)}{\partial T} \\ &= T^2 \frac{N}{T} \quad \Rightarrow \quad T = \frac{E - \mu}{N} \end{aligned} \tag{S.2}$$

we find that μ is a solution of the implicit equation:

$$1 = T \sum_m \frac{1}{E_m - \mu} = \frac{1}{N} \sum_m \frac{E - \mu}{E_m - \mu} \ . \tag{S.3}$$

For a given value of E and a given spectrum E_m this equation can be solved numerically by standard techniques and using (S.2) we also obtain T once μ is known. Depending on the sign of $E - \sum_m E_m < 0$ (or > 0) we have either $\mu < E_1$ and $T > 0$ (or $\mu > E_N$ and $T < 0$) as unique and physically valid solution (mathematically there are typically many other but invalid solutions of (S.3) in the interval $E_1 < \mu < E_N$). Once $\mu(E)$ and $T(E)$ are known one can use ρ_m to compute the entropy

$$S_{EQ}(E) = -\sum_m \rho_{EQ}(E_m) \ln(\rho_{EQ}(E_m)) \ .$$

This expression was used to compute the theoretical $S(E)$ curves in the equi-partition approach based on the grand-canonical classical theory shown in Figs. 1, S2, S11, S13 for various examples.

We mention that the grand canonical temperature (S.2) is similar to the micro-canonical temperature if we replace $E_1 \to \mu$ and it is not difficult to verify that in the limit $E \searrow E_1$ we have $\mu \nearrow E_1$ with $T \searrow 0$ (or if $E \nearrow E_N \quad \Rightarrow \quad \mu \searrow E_N$ with $T \nearrow 0$; see also Figs. 2,S3). Also the micro-canonical expression for ρ_m provides numerically correct $S(E)$ curves (identical to the grand canonical curve) for the lower 20%-30% of the energy spectrum where $\mu \approx E_1$ with a rather good accuracy.

The Bose-Einstein ansatz with

$$\rho_m = \rho_{BE}(E) \equiv \frac{1}{e^{(E_m - \mu)/T} - 1} \tag{S.4}$$

cannot be directly justified by the classical field approach. From a purely formal point of view it can be obtained by replacing in the partition function $|C_m|^2 \to c_m$ with integer c_m and replacing the Gaussian integrations by sums over $c_m = 0, 1, 2, \ldots$ thus resulting in (S.4). In the framework of this approach T and μ are computed by solving numerically the implicit equations (S.1) with $\rho_m = \rho_{BE}(E_m)$ which is technically a bit more complicated as for the EQ case. In the limit of large $|T|$ we can expand in (S.4) the exponential and both approaches become equivalent.

The difference between both approaches in the $S(E)$ curves is not very strong but the numerical data of long time averages of $\rho_m = \langle |C_m(t)|^2 \rangle$ clearly show the validity of the EQ model provided the state is sufficiently thermalized as can be seen in Figs. 3, S4.

Furthermore, according to both the micro-canonical and grand-canonical approaches the statistical distribution of C_m is a complex Gaussian which corresponds to an exponential distribution of $|C_m|^2$, i.e. the distribution of the rescaled variable $x = (E_m - \mu)|C_m|^2/T$ is theoretically $p(x) = \exp(-x)$ which is clearly confirmed by the numerical data for quite large values of x as can be seen in Fig. S5 providing an additional confirmation of the classical model.

Both approaches require the use of a given fixed energy spectrum E_m which is typically obtained by diagonalizing a certain realisation of an RMT matrix (or another

matrix for the variants as DANSE or the model with additional diagonal elements). However, in Fig. 1(d), we show the data for 10 different RMT realisations which would provide individually slightly different $S(E)$ curves. For this figure we used, for both theoretical $S(E)$ curves, a fictitious spectrum with E_m being the solution of $m-1/2 = M(E_m)$ for $m = 1, \ldots, N$ where

$$M(E) = \frac{2N}{\pi}\int_0^E \sqrt{1-E'^2}\, dE' = \frac{N}{2} + \frac{N}{\pi}\left(\arcsin(E) + E\sqrt{1-E^2}\right)$$

is the integrated density of states of the RMT semi-circle law such that $M(-1) = 0$ and $M(1) = N$. This fictitious spectrum corresponds to a constant uniform level spacing in the unfolded spectrum.

The link between the radius (here being unity) of the semi-circle law of a GOE matrix and the variance of its matrix elements $\langle H^2_{n,n'}\rangle = (1+\delta_{n,n'})/(4(N+1))$ is rather standard [29]. However, it can be easily verified by computing the average

$$\left\langle \mathrm{Tr}(\hat{H}^2)\right\rangle = \frac{1}{4(N+1)}\big(2N + N(N-1)\big) = \frac{N}{4}$$

which should coincide with

$$\sum_m \langle E_m^2\rangle = \frac{2N}{\pi}\int_{-1}^{1} dE\, E^2\sqrt{1-E^2} = \frac{N}{4}\ .$$

The above derivations of the thermal distributions ρ_{BE} and ρ_{EQ} are done for finite size systems with a finite number of degrees of freedom. However, they mainly follow the textbook approach of statistical physics for systems in the thermodynamical limit with an infinite number of degrees of freedom.

III. SYMPLECTIC INTEGRATOR

Here we remind some basic facts about symplectic integrators and the particular implementation for our case. For further details, its derivation, we refer for example to [43,44,45], especially for the 4th order variant [43]. These kind of methods are also known as splitting methods [44,45].

A. General method

Let A and B be two non-commuting operators of a general Lie algebra for which it is possible to compute exactly and efficiently (by some exact numerical/analytical method) $\exp(tA)$ and $\exp(tB)$ individually and for arbitrary values of t (or more precisely these operators applied to some given vector or function) while the numerical problem to compute $\exp[t(A+B)]$ is very difficult (very inefficient) or even impossible (as far as an exact method is concerned).

To solve this problem it is sufficient to compute $\exp[\Delta t(A+B)]$ for small Δt (with some given precision) and then to apply: $\exp[t(A+B)] = \exp[\Delta t(A+B)]^n$ with $n = t/\Delta t$ (assuming that t is an integer multiple of Δt). To compute $\exp[\Delta t(A+B)]$ approximately one can write:

$$\exp[\Delta t(A+B)] \approx \prod_{j=1}^{p}\Big[\exp(d_j \Delta t A)\, \exp(c_j \Delta t B)\Big]$$

where the product is ordered with increasing j-values from right to left. The coefficients c_j, d_j, $j = 1, \ldots, p$ are determined such that the error (for one step) is $\sim (\Delta t)^{p+1}$ for a given order p and implying a global error $\sim (\Delta t)^p$ (for many steps and fixed t). The simplest case is $p = 1$ with $c_1 = d_1 = 1$ corresponding to the usual Trotter formula. For $p = 2$, we have the symmetrized Trotter formula with $c_1 = 0$, $c_2 = 1$, $d_1 = d_2 = \frac{1}{2}$. For $p = 3$ there is a non-symmetric solution which can also be found in [43] (see references therein for the proper credit) but which is not really simpler (with all 6 coefficients being different from zero) than the fourth order solution. For $p = 4$ there is a symmetric solution which according to [43] is:

$$c_1 = 0,\ c_2 = c_4 = 2x+1,\ c_3 = -4x-1,$$
$$d_1 = d_4 = x + 0.5,\ d_2 = d_3 = -x$$

where $x = (2^{1/3} + 2^{-1/3} - 1)/6$ is the real solution of $48x^3 + 24x^2 - 1 = 0$. Note that these coefficients verify the sum rule $\sum_j c_j = \sum_j d_j = 1$ due to the first order terms in both exponential expressions. The fourth order formula requires as the third order formula the multiplication of 6 exponential factors for one step if one uses an optimization to merge the d_4-factor with the d_1-factor of the next step (a similar optimization is possible for the symmetrized Trotter formula).

In typical applications one applies this method to solve numerically the time evolution of a classical Hamiltonian or a quantum system where the Hamiltonian is a sum of two terms $H_1 + H_2$ for which the individual exponentials (of either the Liouville operator associated to H_j or $-iH_j$, $j = 1, 2$) can be computed analytically or by an efficient exact numerical method. The splitting method can also be applied to a certain type of partial differential equations [44,45] with potential complications due

to time steps of different signs (i.e. coefficients c_j or d_j having different signs). However, in our situation where we have a system of *ordinary differential equations* for a finite number of degrees of freedom, with an additional imaginary factor i applied to the time variable, there is no numerical nor stability problem with respect to time steps of different signs.

The advantage of the method is that it respects the symplectic/unitary symmetry of the problem. Furthermore, even if one chooses a low order variant with a not so small time step Δt, one can argue that the approximate time evolution (with respect to "$A+B$") represents in reality the *exact* time evolution of a slightly different operator $S \approx A + B$ such that $\exp(\Delta t S)$ coincides exactly with the above product of exponential terms and that many physical features of the modified time evolution are still very relevant since they apply to the same "class" of systems.

B. Numerical implementation

In our case, we chose $A = -i\hat{H}$ (in the quantum point of view or the Liouville operator associated to $\hat{H}$ in the classical point of view) and $B = -i\hat{V}(\psi)$ where $\hat{V}(\psi)$ is an effective potential depending on ψ and with matrix elements $V_{n,n'}(\psi) = \beta|\psi_n|^2\,\delta_{n,n'}$. In this case $e^{-it\hat{V}(\psi)}$ provides the *exact* time evolution of the pure nonlinear equation (assuming $\hat{H} = 0$):

$$\frac{\partial \psi_n(t)}{\partial t} = -i\beta|\psi_n(t)|^2\,\psi_n(t) \quad \Rightarrow$$
$$\psi_n(t) = e^{-it\beta|\psi_n(0)|^2}\,\psi_n(0)$$

which can be easily verified by writing $\psi_n = r_n\,e^{-i\theta_n}$ such that $\dot{r}_n = 0 \quad \Rightarrow \quad r_n(t) = r_n(0) = \text{const.}$ and $\dot{\theta} = \beta r_n^2 \quad \Rightarrow \quad \theta(t) = \theta(0) + t\beta r_n^2(0)$. The conservation of $|\psi_n(t)| = \text{const.}$ (for the pure nonlinear equation) is a feature of the particular form of the nonlinear term and due to this $V(\psi)$ does not depend on ψ nor on t (during the purely nonlinear time evolution) and the time evolution due to the quantum exponential of $-it\hat{V}(\psi)$ coincides exactly with the time evolution of the exponential of the classical Liouville operator associated to the nonlinear term.

In the numerical implementation, we choose a certain initial condition of the state in the representation of the eigenbasis of $\hat{H}$, e.g. $C_m(0) = \delta_{m,m_0}$ with m_0 being the index of the initial state. Then, we apply the first exponential factor with coefficient d_1 (and given value of Δt) which corresponds to $e^{-iE_m d_1 \Delta t}\,C_m \to C_m$. Then, using the unitary matrix that diagonalizes $\hat{H}$, we transform $C_m \to \psi_n$ and we apply the exponential factor with c_2 (since $c_1 = 0$ if $p = 2$ or $p = 4$) which corresponds to $e^{-ic_2\Delta t\beta|\psi_n|^2}\,\psi_n \to \psi_n$ which represents *exactly* the purely nonlinear time evolution. Then we transform $\psi_n \to C_m$ and apply the next exponential factor with coefficient d_2 etc. (If one uses a non-symmetric variant, with $c_1 \neq 0$, for $p = 1$ or $p = 3$ one has first to transform the initial condition to ψ_n, apply the first c_1-factor and transform back to C_m.)

We have implemented and tested all four variants of the method. In particular, we have verified that the classical energy is conserved, i.e. its residual numerical fluctuations ($\sim 10^{-8}$ for the fourth order variant at $\Delta t = 0.1$) scale with $(\Delta t)^p$ and also that the errors of other quantities scale with $(\Delta t)^p$. For the case of a RMT with an extra diagonal where the values E_m become larger, we have also tested the precision by comparing some data with $\Delta t = 0.0125$ (for reduced iteration times) which does not change the values of S etc. (apart from statistical fluctuations).

IV. ADDITIONAL FIGURES

In this section, we present additional SupMat Figures for the main part of this article. The figure captions and figure notes contain physical discussions or additional information for figures in the main part; in particular the values of T and μ for both approaches and the four states shown in Fig. 3 are given in the caption of Fig. S4 below.

FIGURE NOTES

Notes Fig. S1: The initial states are linear eigenstates $\phi_n^{(m_0)}$ of $\hat{H}$ (i.e. $\sum_{n'} H_{n,n'}\phi_{n'}^{(m_0)} = E_{m_0}\phi_n^{(m_0)}$) with specific values of m_0 given in the figure. The entropy is computed from $S(t) = -\sum_m \rho_m \ln(\rho_m)$ where ρ_m is obtained as the time average $\rho_m = \langle |C_m(t)|^2 \rangle$ for successive time intervals with increasing lengths by a factor of two corresponding to the plateau intervals of constant $S(t)$ visible in the figure. The thick horizontal lines represent the theoretical entropy S_{EQ} for EQ (blue) and S_{BE} for BE (red) for the energy of the state at $\beta = 1$ and $m_0 = 3$ (pink open squares). At intermediate times $t \approx 2 \times 10^4$ the entropy of this state is close to S_{BE} while at longer times $t \geq 10^6$ it decreases to S_{EQ} showing that the EQ ansatz describes the correct long time thermalization but also that at intermediate times the entropy is larger and closer to the BE ansatz. The other states $m_0 = 11, 30, 57$ at $\beta = 1$ thermalize rather quickly at $t \geq 10^4$-10^5 to their final value S_{EQ} (with $S_{BE} \approx S_{EQ}$ for $m_0 = 30$). For $\beta = 0.1$ the state $m_0 = 30$ (cyan full squares) thermalizes to the same entropy value as with $\beta = 1$ (green crosses) but only for very long time scales $t \geq 10^6$.

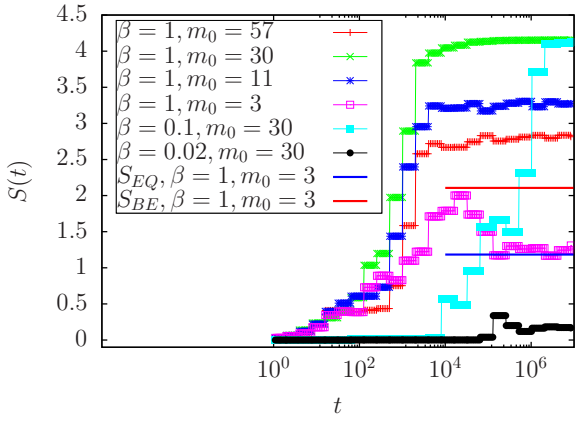

FIG. S1: Time dependence of the entropy $S(t)$ for the four states shown in Fig. 3 and two other states at $\beta = 0.02, 0.1$ with initial index $m_0 = 30$ (for $\beta = 0.02, 0.1$ the other three index values $m_0 = 3, 11, 57$ correspond to very small entropy $S(t)$ values clearly below the case $m_0 = 30$). See details in FIGURE NOTES of Fig. S1.

Notes Fig. S2: For $\beta = 1$, $N = 256$ the time average corresponds to $2^{21} \leq t \leq 2^{22}$ (red +, all 256 initial conditions) and $2^{25} \leq t \leq 2^{26}$ (blue $\circ$, 35 initial conditions with $1 \leq m \leq 35$). The curves represent the theoretical $S(E)$-curves from the EQ (blue) and BE (red) approaches using the exact spectrum of the used RMT realisation. The data point with $S > 1$ for $N = 32$, $\beta = 0.02$ is not saturated and still increasing at the given maximal time $t = 2^{27}$. The data for $\beta = 1$, $N = 128$ and $\beta = 1, 2$, $N = 256$ coincide very well with the EQ ansatz. Also for $N = 32$ the EQ ansatz is more appropriate. Here the small differences to the theoretical EQ-curve are due to the fact that on the x-axis the initial energy E_m is used and not the averaged linear energy $\langle E \rangle = \sum_{m'} E_{m'}\rho_{m'}$ using the long time average $\rho_{m'}$ and which is slightly different from E_m due to the nonlinear term. Using $\langle E \rangle$ the data points (for the cases with good thermalization) fall nearly exactly on the theoretical curve. For $\beta = 1$, $N = 256$ the data points with $m \geq 10$ thermalize well and rather *early* to the EQ curve already for the time average interval $2^{21} \leq t \leq 2^{22}$. The data points for $1 \leq m \leq 5$ and $m = 7$ do not thermalize at all even for $2^{25} \leq t \leq 2^{26}$ with entropy values clearly below the EQ and BE curves and being rather constant between $2^{22} \leq t \leq 2^{26}$. The two data points for $m = 6$ and $m = 9$ thermalize *late* to the EQ curve for $2^{25} \leq t \leq 2^{26}$ while for $2^{21} \leq t \leq 2^{22}$ their entropy values are clearly below the EQ and BE curves. The data point at $m = 8$ also approaches *late* the EQ curve ($2^{25} \leq t \leq 2^{26}$) but from above, i.e. with *early* entropy values ($2^{21} \leq t \leq 2^{22}$) slightly above the

EQ curve but still clearly below the BE curve. Additional and more detailed figures (higher resolution and more data points at different times) for theses points are available at [46].

We note that for $N = 128, 256$ detains states close to the spectral border $E = \pm 1$ are not thermalized even at very large times. We attribute this to the fact that at such energies the energy level spacing is significantly increased as compared to the band center and thus a stronger nonlinearity β is required for thermalization. Indeed, for the larger nonlinearity parameter $\beta = 2$ more states of those border states are thermalized as compared to $\beta = 1$. More data for $N = 256$ are available in [46].

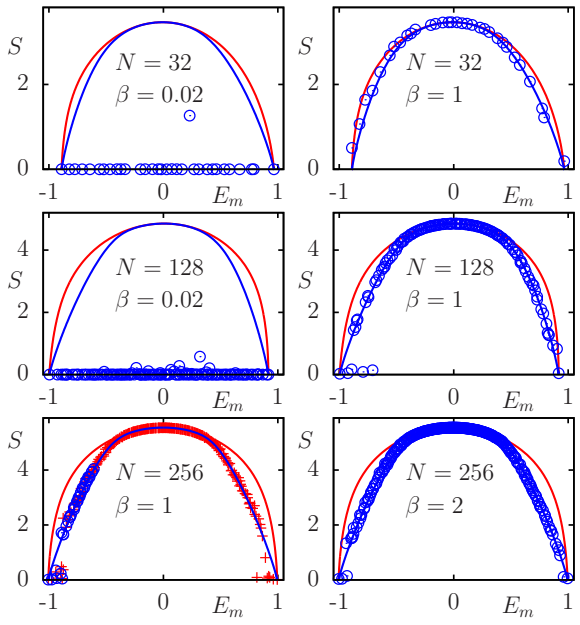

FIG. S2: Dependence of entropy on energy $S(E)$. As in Fig. 1 data are for one RMT realisation, $\beta = 0.02, 1$ for $N = 32, 128$ and $\beta = 1, 2$ for $N = 256$, ρ_m obtained by the time average for $2^{21} \leq t \leq 2^{22}$ for $N = 128, 256$ and $2^{26} \leq t \leq 2^{27}$ for $N = 32$ (blue $\circ$, all panels except $\beta = 1$, $N = 256$). See details in FIGURE NOTES of Fig. S2.

Notes Fig. S3: The data points in Figs. 2 and S3 were obtained by computing μ and T from the implicit set of the two equations : $S_{\rm num.} = -\sum_m \rho_m \ln \rho_m$ and $1 = \sum_m \rho_m$ using the expressions Eq. (2) for both approaches and the numerical values of $S_{\rm num.}$ (data points in Figs. 1 and S2). Therefore, if $S_{\rm num.}$ and $S(E)$ are not identical (due to statistical fluctuations or lack of thermalization), these data points do not need to fall exactly on the theory curves which were obtained by solving another set of two equations : $E = \sum_m E_m \rho_m$ and $1 = \sum_m \rho_m$ (using Eq. (2)). The deviations of the data points with respect to the theory curves are significantly weaker for the EQ case than for the BE case but the latter are still quite weak, even though better visible in Fig. S3 as compared to Fig. 2 (with no visible difference on graphical precision). This observation confirms somehow that the EQ ansatz fits better the numerical data but the analysis shown in Figs. 2 and S3 does not allow to distinguish very clearly between the validity of either the EQ or the BE ansatz.

For this the direct comparison the numerical values of ρ_m with the expressions (2) (see Figs. 3 and S4), provide a much stronger argument in support of the EQ ansatz.

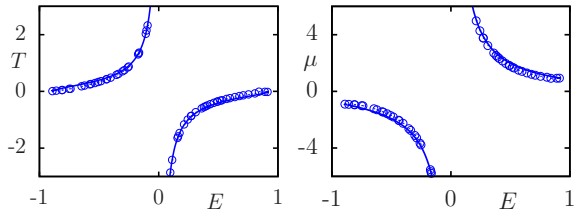

FIG. S3: As Fig. 2 but for the BE case. Here the data points for T and μ, computed from the numerical data of S, show some small deviations from the theoretical curves which are visible in the figure and significantly larger than in Fig. 2 where no deviations for the EQ case (on graphical precision) are visible. Furthermore, in comparison to the EQ case of Fig. 2 the typical values of T and μ for the BE case are considerably larger. See details in FIGURE NOTES of Fig. S3.

Notes Fig. S4: Here, the blue curve shows the theoretical values based on EQ with $\rho_{\rm EQ}(E) = T/(E - \mu)$ and $T = 0.002007, 0.004986$ ($T = 0.0006853, 0.001414$), $\mu = -0.9995, -1.041$ ($\mu = -1.006, -1.009$) for $m_0 = 11, 31$ and $N = 128$ ($N = 256$). T and μ were determined from the solution of the equations $1 = \sum_m \rho_{\rm EQ}(E_m)$ and $\langle E \rangle = \sum_m E_m \rho_{\rm EQ}(E_m)$ with $\langle E \rangle = \sum_m E_m \rho_m \approx E_{m_0}$. The red line shows the theoretical values based on BE with $\rho_{\rm BE}(E) = 1/(\exp[(E - \mu)/T] - 1)$, $T = 0.1983, 0.5562$ ($T = 0.1301, 0.2686$), $\mu = -1.434, -2.914$

($\mu = -1.300, -1.875$) for $m_0 = 11, 31$ and $N = 128$ ($N = 256$). Here T, μ were determined from the solution of the equations $1 = \sum_m \rho_{\rm BE}(E_m)$ and $\langle E\rangle = \sum_m E_m \rho_{\rm BE}(E_m)$. Furthermore, the energy values for $m_0 = 11, 31$ and $N = 128$ ($N = 256$) are $E_{m_0} \approx \langle E\rangle = -0.7426, -0.4031$ ($E_{m_0} \approx \langle E\rangle = -0.8307, -0.6469$). The thermalization of all four states according to the EQ theory is very good (essentially perfect) despite the shorter averaging time for the case $N = 128$ as compared to Fig. 3. For $N = 256$ the averaging time is rather long but here the selected states are closer to the band edge at $E_{\rm min} = -1$ and with lower temperature values (than for $N = 128$) such that thermalization is more difficult. Additional similar figures for other values of m_0 at different values of N are available at [46].

These results clearly show that the dynamical thermalization of ρ_m is very well described by the EQ ansatz (2). High quality figures of ρ_m for all initial states values m_0 at $N = 256$ are available in [46], including all thermalized states with negative temperatures $T < 0$ which appear at energies $E_{m_0} > 0$.

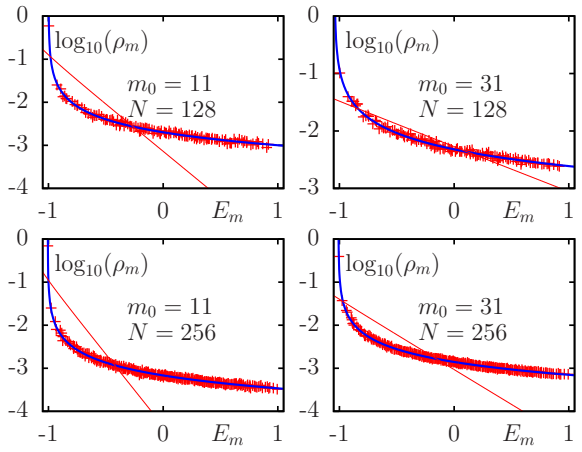

FIG. S4: Dependence of ρ_m on E_m for two states with initial state $m_0 = 11, 31$ for $\beta = 1$, $N = 128$ ($N = 256$) obtained by an time average in the interval $2^{21} \le t \le 2^{22}$ ($2^{25} \le t \le 2^{26}$). *Complementary information for* Fig. 3: The T and μ values for the EQ ansatz and the four states $m_0 = 3, 11, 30, 56$ with $N = 64$ and $\beta = 1$ of Fig. 3 are $T = 0.001372, 0.005984, 0.07585, -0.004538$, $\mu = -0.8964, -0.9178, -4.892, 0.9293$ and the corresponding values for the same states and the BE ansatz are $T = 0.112, 0.3581, 4.913, -0.2649$, $\mu = -1.062, -1.794, -20.52, 1.496$. See details in FIGURE NOTES of Fig. S4.

Notes Fig. S5: This figure clearly shows that the statistical distribution of $C_m(t)$ (or of the quantity $u \equiv |C_m(t)|^2$) is very well described by the thermal Boltzmann Gaussian distribution $\exp(-(E_m - \mu)|C_m(t)|^2/T)$ (or exponential distribution $\exp(-(E_m - \mu)\, u/T)$ in u) for values up to $u \approx$ (8-10) $\times \langle u\rangle$.

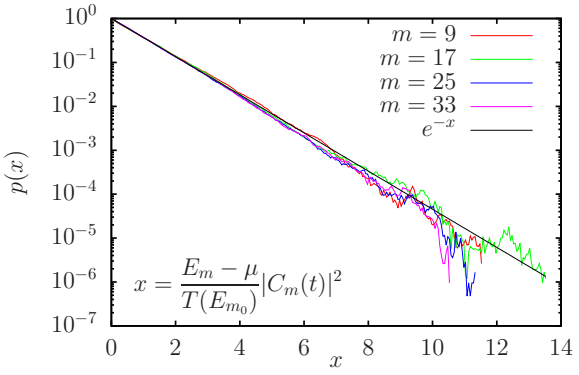

FIG. S5: Statistical distribution of the rescaled variable $x = (E_m - \mu)|C_m(t)|^2/T(E_{m_0})$ for $\beta = 1$, $N = 64$, $m_0 = 9$, $m = 9, 17, 25, 33$ using the time values in the interval $2^{23} \le t \le 2^{24}$ and a histogram of bin width 0.05. The thin black line shows the theoretical distribution $p(x) = e^{-x}$ according to the EQ approach. The numerical distributions follow the theoretical distribution for values up to $x \approx$ 8-10 providing an additional confirmation for the validity of the EQ ansatz. See details in FIGURE NOTES of Fig. S5.

Notes Fig. S6: This figure is similar to Fig. 5 but with additional β values : Lyapunov exponent λ_m dependence on E_m with m being the index of the initial state $\phi^{(m)}$ for $N = 64$. Here λ_m is determined from the fit $\ln\|\Delta\psi(t)\| = a + b\ln(t) + \lambda_m t$ for $t \le 2^{22}$ and $\beta = 2$ (grey ▲; top), $\beta = 1.5$ (orange △), $\beta = 1$ (black •), $\beta = 0.75$ (cyan ■), $\beta = 0.5$ (pink □), $\beta = 0.25$ (blue ∗), $\beta = 0.1$ (green ×), $\beta = 0.02$ (red +), $\beta = 0.02$ for $t \le 2^{27}$ (red ▽; bottom). The numerical data suggests that most λ_m for $\beta = 0.02$ decay as $\lambda_m \sim 1/\sqrt{t}$ for $t \ge 10^7$ (see Fig. S10 below). However, three λ_m values for $\beta = 0.02$ do not decay with time (data points with same red + and ▽; e.g. $m = 45$ and $E_m \approx 0.35$) and have significantly larger values $\lambda_m > 10^{-5}$ indicating a trajectory in a chaotic region while other initial conditions correspond to trajectories in bounded KAM regions. These cases are also visible in Fig. 1 (a) with entropy values slightly above 0.

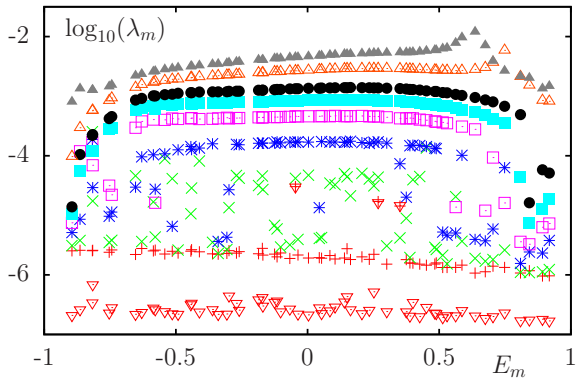

FIG. S6: Dependence of Lyapunov exponent λ_m on the energy E_m of the initial state $\phi^{(m)}$, same as in Fig. 5. See details in FIGURE NOTES of Fig. S6.

Notes Fig. S7: In the *bottom* panel the fits are done using all data points such that the values for $\beta \leq 0.05$ have a smaller weight due to the reduced number of different initial conditions. Very long time computations for 35 random initial conditions at $\beta = 0.02$ for $t \leq 3 \times 10^9$ ($t \leq 10^9$) for $N = 32$ ($N = 64$) indicate a chaotic behavior with no further global decrease of λ with time for $t > 2^{27} \approx 1.3 \times 10^8$. There are however considerable fluctuations between different initial conditions and in the time dependence (but with very long correlation times) of the order of 10-15%. Additional figures for this point are available at [46].

We point out that for a localized initial condition with only one mode m the Lyapunov exponent at $\beta = 0.02$ (*top* panel) decreases with time as $\lambda \sim t^{-1/2}$ (see Fig. S10) indicating a non-chaotic behavior in the limit of very large times. In contrast for states $\psi(t) = \sum_m C_m(t)\phi^{(m)}$, with uniform random initial C_m amplitudes at $t = 0$ (*bottom* panel), the Lyapunov exponents are well stabilized at large times $t < 1.3 \times 10^8$ even for $\beta = 0.02$ (see Figs. in [46]). These state have automatically an average energy $E \approx 0$ close to the band-center. It is important to stress that all states with initial random configurations have approximately the same values of $\lambda > 0$. This means that even at small value $\beta = 0.02$ the measure of the chaotic component (at $E \approx 0$) is close to unity. This result is very different from many-body nonlinear systems studied in [47] where the measure of the chaotic component is $\sim \beta$.

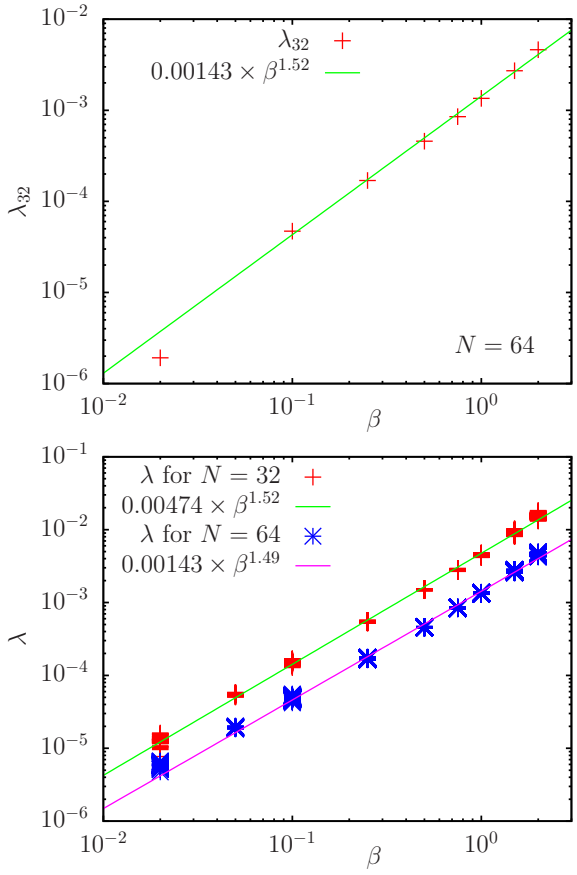

FIG. S7: *Top:* Dependence of Lyapunov exponent λ_{32} on β in the energy band center for the localized initial condition at $m = 32$, $N = 64$ and computation time $t = 2^{22}$. The straight green line shows the power law fit $\lambda_{32} = a\beta^{\eta}$ with $a = 0.00143 \pm 0.00005$ and $\eta = 1.52 \pm 0.03$. For this fit the smallest data point at $\beta = 0.02$ was not used since for this value the Lyapunov exponent continues to decrease with increasing computation time t and it is most likely below the chaos border. *Bottom:* Dependence of Lyapunov exponent λ on β for many different random uniform C_m configurations (most energy values $\langle E \rangle \approx 0$ and some cases with $\langle E \rangle \approx \pm 0.1$) for $N = 32, 64$ and computation time $t = 2^{22}$ ($t = 2^{27}$) for $\beta \geq 0.1$ ($\beta \leq 0.05$). The number of initial conditions is 64 for $\beta \leq 0.05$, and (at least) $10N = 320,640$ for $\beta \geq 0.1$; symbols at fixed β mark λ for different initial configurations. The straight green line shows the power law fit $\lambda = a\beta^{\eta}$ with $a = 0.00474 \pm 0.000009$ and $\eta = 1.524 \pm 0.0015$ for the case $N = 32$. The straight pink line shows the power law fit $\lambda = a\beta^{\eta}$ with $a = 0.001426 \pm 0.000002$ and $\eta = 1.491 \pm 0.0013$ for the case $N = 64$. See details in FIGURE NOTES of Fig. S7.

Notes Fig. S8: In the *bottom* panel symbols at fixed N show λ_m values for different initial condition (one mode $\phi^{(m)}$ with energy close to the band center); all obtained λ_m values are rather close to each other. This indicates that at $\beta = 1$ the measure of the chaotic component is close to unity for $N \leq 512$ (see also NOTES of Fig. S7).

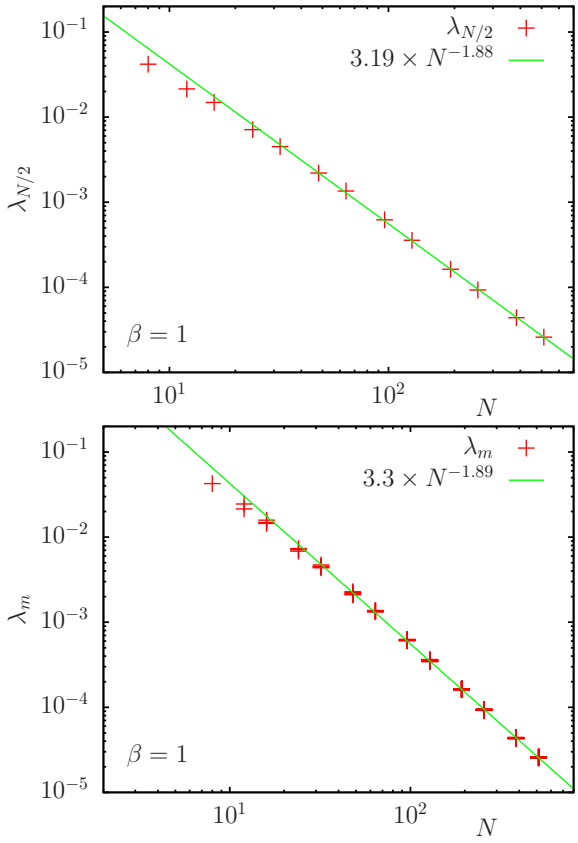

FIG. S8: Dependence of Lyapunov exponent $\lambda_{N/2}$ on N at $\beta = 1$. *Top:* data for the band center and computation time $t = 2^{22}$ ($t = 2^{24}$) for $N \leq 384$ ($N = 512$). The straight green line shows the power law fit $\lambda_{N/2} = aN^{-\nu}$ with $a = 3.19 \pm 0.19$ and $\nu = 1.88 \pm 0.01$ using the data for $N \geq 32$. *Bottom:* Same as top but showing all λ_m values corresponding to $|E_m| \leq 0.1$ (for $N \leq 128$) or 35 values in the band center with $|m - N/2| \leq 17$ (for $N \geq 192$). The straight green line shows the power law fit $\lambda_{N/2} = aN^{-\nu}$ with $a = 3.30 \pm 0.04$ and $\nu = 1.886 \pm 0.002$ using the data for $N \geq 32$. The data of both panels correspond to the case of localized initial conditions at some value m (with E_m being close to the band center). See details in FIGURE NOTES of Fig. S8.

Notes Fig. S9: At $\beta = 1$ the scaling $\lambda \propto 1/N^{\nu}$ with $\nu = 1.89$ works well in the energy band center. Certain

deviations from this scaling are seen in the vicinity of the energy edges $E \approx \pm 1$. We attribute this to a significant increase of level spacing at band edges that may modify chaos properties at different N at band edges.

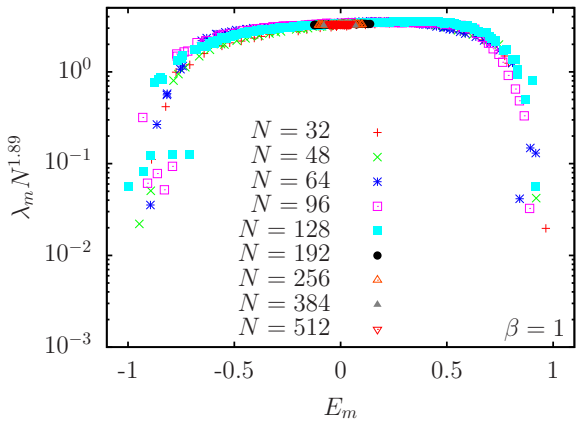

FIG. S9: Dependence of rescaled Lyapunov exponent $\lambda_m N^\nu$ on the initial energy E_m using the exponent $\nu \approx 1.89$ found in the fit of (the bottom panel of) Fig. S8 for different values of N with $32 \le N \le 512$ and $\beta = 1$. For $N \ge 192$ only 35 values of λ_m (per N value) in the band center with $|m - N/2| \le 17$ are available and shown. See also FIGURE NOTES of Fig. S9.

Notes Fig. S10: Here $\lambda_m(t)$ has been obtained by the fit $\ln \|\Delta\psi(\tau)\| = a + b \ln(\tau) + \lambda_m \tau$ for $0 \le \tau \le t$ and for values $t \le 2^{27}$ where $\Delta\psi(\tau) = \tilde{\psi}(\tau) - \psi(\tau)$ is the difference vector between two close initial conditions with $\psi(0) = \phi^{(m)}$, $\tilde{\psi}(0) = \psi(0) + \Delta\psi(0)$ and $\Delta\psi(0)$ being a random vector with initial norm $\|\Delta\psi(0)\| = 10^{-12}$. During the computation the difference vector $\Delta\psi(\tau)$ is regularly renormalized to the norm 10^{-12} when its norm has become larger than 10^{-10} such that both trajectories stay close and the logarithm of the renormalization factor is added to a special variable measuring the quantity $\ln \|\Delta\psi(\tau)\|$ which is used for the computation of the Lyapunov exponent. The rescaled logarithmic growth $(\ln \|\Delta\psi(t)\|)/t$ shows roughly the same behavior as $\lambda_m(t)$, with a final slope somewhat closer to the exponent $-1/2$ than for $\lambda_m(t)$ (in logarithmic representation and for $t \ge 10^7$).

The two cases at $m = 31$, $r = 0, 1$ indicate a vanishing Lyapunov exponent in the limit $t \to \infty$ and a trajectory in a bounded KAM region. The Lyapunov exponent for the other two cases at $m = 45$ (with $E_{45}(r = 0) \approx 0.351$ and $E_{45}(r = 1) \approx 0.310$) saturate to the values $\lambda_{45}(r = 0) \approx 1.47 \times 10^{-5}$ and $\lambda_{45}(r = 1) \approx 8.89 \times 10^{-6}$ in the limit $t \to \infty$ indicating a trajectory in a chaotic region probably due to the effect of a near triple quasi-resonance for the given RMT realisation. For the first realisation $r = 0$ there are three cases like this as can be seen in Figs. 5 and S6 (see also caption therein). The observation that for both realisations there are saturated Lyapunov values at the same index $m = 45$ is a coincidence and for example for $m = 32$ (not shown in the figure) there is a stabilized Lyapunov exponent for $r = 1$ but not for $r = 0$.

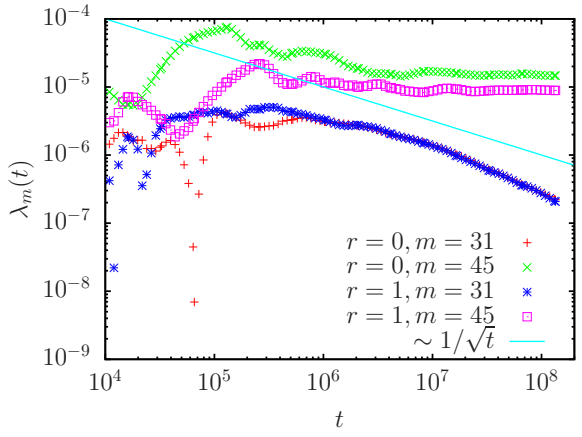

FIG. S10: Dependence of Lyapunov exponent $\lambda_m(t)$ on time t for $\beta = 0.02$, $N = 64$ and two initial states in the band center, $m = 31, 45$. Data are for two RMT realisations $r = 0$ (same realisation as for most main and SupMat figures concerning the RMT case) and $r = 1$. The cyan full line shows $10^{-2}/\sqrt{t}$ to indicate an empirical $t^{-1/2}$ power law $\lambda \propto t^{-1/2}$ at large times. See details in FIGURE NOTES of Fig. S10.

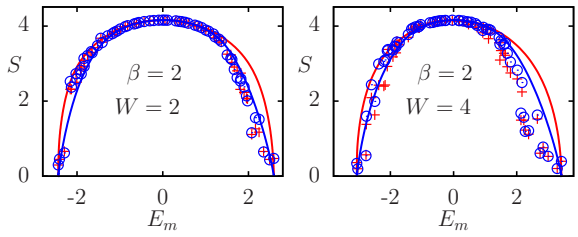

FIG. S11: As Fig. 1 but for one realisation of the DANSE model of [36] at disorder strength $W = 2$ and $W = 4$ for $\beta = 2$ and $N = 64$. The data points correspond to the averaging time $2^{23} \leq t \leq 2^{24}$ (blue $\circ$) and $2^{20} \leq t \leq 2^{21}$ (red $+$; similar t values as in [36]). See also FIGURE NOTES of Fig. S11 and Fig. S12.

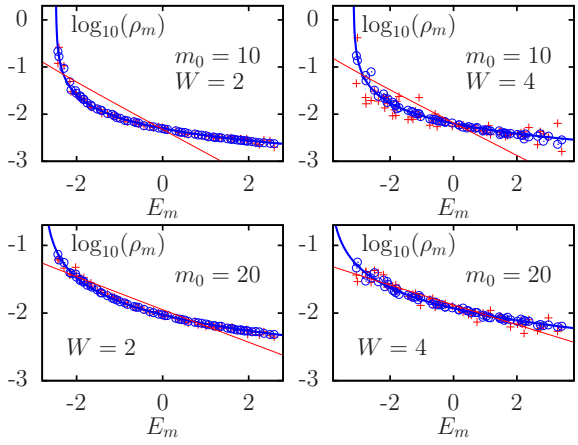

FIG. S12: As Figs. 3 and S4 but for one realisation of the DANSE model of [30] at disorder strengths $W = 2$ and $W = 4$ for $\beta = 2$, $N = 64$ and two initial states with $m_0 = 10, 20$. The data points correspond to ρ_m obtained by the averaging time $2^{23} \leq t \leq 2^{24}$ (blue $\circ$) and $2^{20} \leq t \leq 2^{21}$ (red $+$; similar t values as in [30]). The values of T and μ for the EQ approach are $T = 0.0124, 0.02626, 0.01997, 0.04636$ and $\mu = -2.484, -2.784, -3.157, -3.952$ for $W = 2$ with $m_0 = 10, 20$ and $W = 4$ with $m_0 = 10, 20$. The values of T and μ for the BE approach are $T = 0.8794, 1.815, 1.26, 3.011$ and $\mu = -4.709, -8.171, -6.346, -13.07$ for the same states. For $W = 2$ both states are well thermalized according to the EQ case. For $W = 4$ the thermalization also corresponds to the EQ case but there are still stronger fluctuations, especially for the data with shorter averaging time (and corresponding to the data of [30]). See also FIGURE NOTES of Fig. S11 and Fig. S12.

Notes Fig. S11 and Fig. S12: These two figure correspond to the case of the DANSE model studied in [36]. In the limit $N \to \infty$ and $\beta = 0$ the model is reduced to the Anderson model in one-dimensions with exponentially localized eigenstates and the localization length $\ell \approx 96/W^2$ in the band center. Here W is the strength of the diagonal disorder. For $N = 32, 64$ and $W = 2, 4$ the value of ℓ is comparable to the system size and chaos induced by the nonlinearity β leads to dynamical thermalization EQ (2) as it is shown in Fig. S11 and Fig. S12. More thermalization figures for all initial eigenmodes, including those leading to negative temperature $T < 0$ are available at [46].

Notes Fig. S13 and Fig. S14: These figures correspond to the case with an additional linearly growing

term fn added to the diagonal matrix elements $H_{n,n}$. At $\beta = 1$, $N = 32$ and $f = 0.25$ the dynamical thermalization is reached at large times but it is not completely the case for $f = 0.5$ (see Fig. S13). As shows Fig. S14 at initial times we have an approximate exponential drop of probabilities ρ_m with E_m (red crosses) which is similar to a BE or quantum Gibbs distribution. However, at larger times the distribution ρ_m approaches the theoretical EQ curve (2). We assume that there is a relatively rapid process of chaotic mixing of modes being close to the initial $m = m_0$ value and those with lower energies at $m < m_0$. Somehow it is easy to go to low energies while the propagation of excitations to higher energies, with m being significantly higher than m_0, goes as a slow diffusion requiring significantly longer times. Indeed, at $f = 0.5$, $N = 64$ the whole energy range is close to $E_N - E_1 \approx 32$ being much larger than the range $E_N - E_1 \approx 2$ at $f = 0$. We argue that such a slow diffusion in energy is at the origin of the approximate BE distribution found in numerical simulations with the Bunimovich stadium [33] and the Sinai oscillator [34] which have a very broad energy range and the time of numerical simulations was not very high due to the complexity of the integration of GPE.

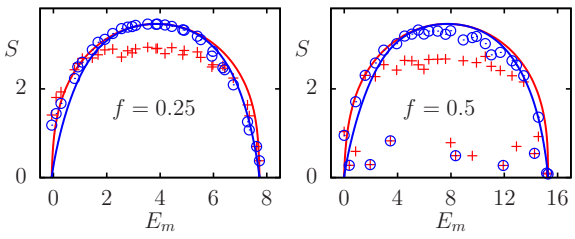

FIG. S13: As Fig. 1 for the case of an RMT plus extra diagonal matrix elements fn with parameter $f = 0.25$ or $f = 0.5$ for $\beta = 1$ and $N = 32$. The data points correspond to the averaging time $2^{26} \le t \le 2^{27}$ (blue $\circ$ for $f = 0.25, 0.5$), $2^{19} \le t \le 2^{20}$ (red $+$ for $f = 0.5$) and $2^{14} \le t \le 2^{15}$ (red $+$ for $f = 0.25$). See also FIGURE NOTES of Fig. S13 and Fig. S14.

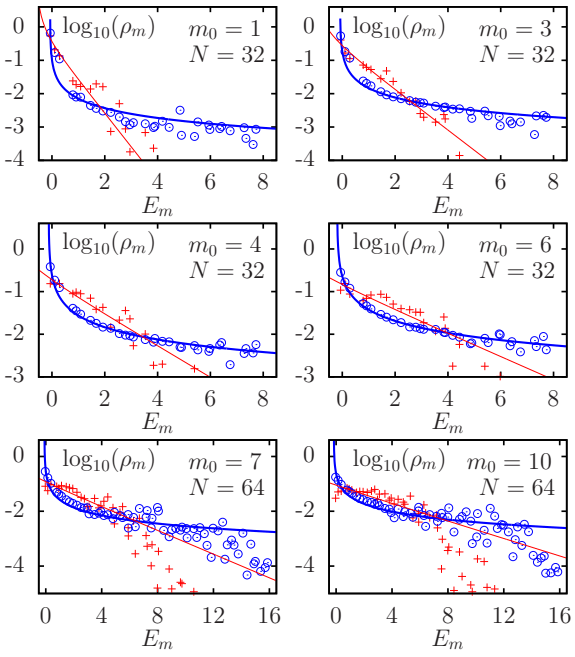

FIG. S14: As Fig. 3 for the case of an RMT plus extra diagonal matrix elements fn with parameter $f = 0.25$ for $\beta = 1$, $m_0 = 1, 3, 4, 6$ ($N = 32$) or $m_0 = 7, 10$ ($N = 64$). The data points correspond to ρ_m obtained by the averaging time $2^{26} \le t \le 2^{27}$ (blue $\circ$), $2^{19} \le t \le 2^{20}$ (red $+$; for $N = 64$) and $2^{14} \le t \le 2^{15}$ (red $+$; for $N = 32$). At longer times $t = 2^{27}$ the states are (quite) well thermalized according to the EQ case (with somewhat stronger fluctuations for $N = 64$). However, at the intermediate time scale the values of ρ_m are closer to the BE line thus explaining that the corresponding entropy values are also closer to the BE curve. See also FIGURE NOTES of Fig. S13 and Fig. S14.

Scaling of Lyapunov exponent and chaos border

The numerical results presented in Figs. 5, S6-S10 are reasonably well described by the scaling relation:

$$\lambda \sim \beta^{\eta}/N^{\nu}, \ \eta = 3/2, \ \nu = 2. \qquad \text{(S.5)}$$

Indeed, the fits of data give values $\eta = 1.52$ and $\nu = 1.89$ being close to (S.5) and we assume that in the limit of large N and small β we will have the exponents of (S.5).

We find that most states with localized initial conditions ($C_m(t = 0) = \delta_{m,m_0}$) have zero Lyapunov exponents at our smallest value $\beta = 0.02$ (with a few exceptions due to strong quasi degenerate levels as discussed above). But at the same time the Lyapunov exponent is positive for random initial configurations with random

and uniform initial values of C_m (which gives automatically an initial energy close to the energy band center). All such states have approximately the same values of λ indicating that the measure of the chaotic component is close to unity. At present, we cannot say what is the precise chaos border β_c for such states. For the moment, we do not have theoretical arguments for the found dependence (S.5).

We only note that equation (1) for the time evolution can be rewritten in the basis of linear eigenmodes (see eqs. Eq.(3) in [28] or Eq.(2) in [29]). In this representation the transitions between modes are induced only by β-terms with 4-mode interaction (or 4-wave interaction) $C_{m_1}C_{m_2}C^*_{m_3}C^*_m \exp[-i(E_{m_1}+E_{m_2}-E_{m_3}-E_m)t]$. In the RMT case the amplitudes of this interaction have a typical value $Q \sim 1/N^{3/2}$ (see also [28,29]). Thus the lowest energy difference between these 4 energies is of the order of $\delta E \sim 1/N^2$ that can be at the origin of $\nu \approx 2$ and rather low chaos border with $\beta_c < 0.02$. We note that the same estimate for δE remains valid even in presence of the diagonal term fn that stress the importance of 4-mode interactions.

However, the above estimates remain insufficient and the understanding of the relation (S.5) requires further studies.

V. DYNAMICAL THERMALIZATION IN MULTIMODE OPTICAL FIBERS

Very recent remarkable experiments (published at 8 February 2023; after the submission date 22 December 2022 of our work) with multimode optical fibers (MMF) [46] demonstrated dynamical thermalization in MMF with negative temperature. It is stressed there that this is a dynamical thermalization resulting from pure Hamiltonian dynamics without an external thermal bath [46]. The equilibrium state is a thermal state with energy equipartition over fiber modes described by the EQ ansatz (2) also known in optics as Rayleigh-Jeans distribution. As we pointed out in Eq,(2) the EQ ansatz is a limiting case of the BE ansatz when the temperature is large compared to $E_m - \mu$ in the BE exponent. This can be considered as the case when the field has many photons of linear modes.

There is a significant literature with discussions, numerical simulations and experiments on dynamical thermalization in MMF (see e.g. Refs. S1,S2,S3,S4,[46]). The emergence of Rayleigh-Jeans distribution is explained in the frame work of the weak turbulence approach (see Refs. S5,S6,S1,S2). However, it should be pointed out that the weak turbulence theory (see Refs. S5,S6) assumes an existence of a certain weak randomizing force that disappears in the final equilibrium state. This is in direct contradiction with the dynamical Hamiltonian equations leading to the equilibrium thermal state. In fact it is clear that the origin of dynamical thermalization in MMF is dynamical chaos and its exponential instability of motion is related to a positive maximal Lyapunov exponent. However, strangely enough no notion of dynamical chaos and Lyapunov exponent appeared in theoretical arguments of Refs. S1,S2.S3.S4. Also from the theory of chaos it is clear that no thermalization appears if the nonlinear perturbation is sufficiently weak and below the chaos border (KAM integrability, see Refs.[5-8]). In fact, we should note that in contrast to our RMT case the spectrum of MMF discussed in Refs. S1-S4,[46] has a form $E \propto (m_x + m_y + const)$ thus with exact degenerate energy levels for the lowest 45 modes considered practically in all MMF cases (and also in [46]). As was shown in [27,41] for such a case with degeneracy of modes the KAM theory is not valid and dynamical chaos appears at an arbitrarily small nonlinear perturbation. However, such chaos is localized only on degenerate modes and does not lead to dynamical thermalization over all modes.

Another interesting note about dynamical thermalization in MMF experiments is about the validity of BE or EQ ansatz (2). It is possible to assume that the light waves are classical and then one should observe the EQ or Rayleigh-Jeans distribution over modes. However, the real life is of course described by quantum mechanics with second quantization of photons and their interactions that should lead to the Bose-Einstein distribution (BE ansatz (2)). It is possible that in MMF experiments the number of photons was very large, dynamical temperature was high and the BE distribution was transferred to its classical limit with the EQ ansatz (or Rayleigh-Jeans). However, it is interesting to know if MMF can operate in a quantum regime with the BE thermal distribution.

We also point out that all discussed MMF systems have very simple integrable spectrum with $E \propto (m_x + m_y + const)$ being rather far from the RMT spectrum which corresponds to a generic case. Of course, it is difficult to realize such an RMT case with MMF. However, it is possible to have cases when a fiber cross-section have a form of a chaotic billiard. It may be the Bunimovich stadium (two semi-circles connected by two parallel straight lines), or a circle with a line cut. In such systems the classical dynamics is chaotic and the level spacing statistics is the same as for RMT [25,26]. So we assume that such sections can be realized technologically thus allowing to study nonlinear effects for MMF in a regime of quantum chaos.

VI. GENERIC FEATURES OF DYNAMICAL THERMALIZATION IN THE NLIRM MODEL

The emergence of dynamical thermalization and its properties appeared as far as 150 years ago in 1872 in the work of Boltzmann who established the foundations of statistical mechanics and thermalization from dynamical equations [1] (see also the related Boltzmann-Loschmidt dispute [2,3,4]). The first attempt to obtain dynamical thermalization in a nonlinear oscillator system, known as the FPU problem [10], was not successful due to certain specific features of the FPU model.

In this work, we considered the NLIRM model (1), which describes the classical dynamics of nonlinear oscillators, coupled by a Gaussian random matrix, and in which a moderate nonlinearity leads to the emergence of dynamical chaos followed by the classical dynamical thermal distribution Eq. (2) corresponding to the energy equipartition between oscillator modes of the unperturbed linear system.

We argue that, in contrast to the FPU problem [10], our NLIRM model captures the generic features of linear oscillator systems with moderate nonlinear interactions between linear eigenmodes.

First, the statistical classical theory given above in section I, is very generic and applies to generic linear couplings and generic interactions as long as we have the two integral of motions and as long as the system is sufficiently chaotic to ensure thermalization.

Furthermore, without the nonlinearity the oscillators are described by Random Matrix Theory (RMT) which captures the generic features of such diverse quantum systems as complex atoms, molecules and nuclei, mesoscopic electronic systems and systems of quantum chaos [22,23,24,25,26].

In this work, we mostly used a nonlinear onsite interaction which is broadly used in condensed matter systems and is known as the Hubbard interaction (see e.g. Ref. S7). We showed that this interaction leads indeed to the EQ dynamical thermal distribution (2), perfectly confirming the theory of section I, not only for the linear oscillator system described by the RMT model but also by the DANSE model studied previously in [28,29,30] provided the iteration time is sufficiently long (see Figs. S11, S12). This confirms the generic properties of the NLIRM model concerning *the linear oscillator couplings*.

However, one can question if this model is also generic concerning the specific form of the onsite interaction and if the latter captures the generic features of dynamical thermalization in the NLIRM model. To study this question, we have also considered two modified interaction models which are not restricted to onsite interactions only and which correspond (i) to nearest neighbors interactions (NNI) and (ii) to long range "Coulomb type" interactions (COULI). In these models the wavefunction evolution is described by the equation

$$i\hbar\frac{\partial\psi_n(t)}{\partial t} = \sum_{n'=1}^{N} H_{n,n'}\psi_{n'}(t) + \beta\left(\sum_j V_j|\psi_{n+j}(t)|^2\right)\psi_n(t)\ . \tag{S.6}$$

with interaction couplings $V_j = 1$ for $j = -2,-1,0,1,2$ and $V_j = 0$ for other $|j| > 2$ in the NNI case and $V_j = 1/(1+|j|)$ for $-N/2+1 < j < N/2+1$ in the COULI case; the linear term with $H_{n,n'}$ (taken as a Gaussian random matrix) remains unchanged. In (S.6), if $n+j < 0$ or $n+j \geq N$ we apply periodic boundary conditions, i.e. $n+j \to n+j+N$ if $n+j<0$ and $n+j \to n+j-N$ if $n+j \geq N$. One can easily verify that for these types of interactions, we also have two integrals for motion being the conserved norm $1 = \sum_n |\psi_n(t)|^2$ and the conserved classical energy which now reads :

$$E = \sum_n \Big(< \psi_n(t)|\hat{H}|\psi_n(t) > + \frac{\beta}{2}|\psi_n(t)|^2\sum_j V_j|\psi_{n+j}(t)|^2\Big)\ . \tag{S.7}$$

Therefore, the statistical classical theory given above in section I, equally applies to these kind of interactions.

Furthermore, the considerations of section IIB can be generalized for these interactions. In particular in absence of the linear coupling (if $H_{nn'} = 0$) the pure nonlinear dynamics conserves the individual values of $r_n = |\psi_n(t)| = \text{const.}$ and only the phases $\theta_n(t)$ evolve such that (for the pure nonlinear dynamics) we have:

$$\psi_n(t) = e^{-it\beta\sum_j|\psi_{n+j}(0)|^2}\,\psi_n(0)\ .$$

This point is important to justify the use of the symplectic integrator which requires to compute the exact exponential $\exp(tB)$ of the operator B corresponding to the nonlinear term (see section IIB). We have verified that (i) the 4th order symplectic integrator, applied to both modified interaction models, still produces results such that the (global) error of the method scales with $(\Delta t)^4$ and (ii) that the classical energy (S.7) is indeed conserved with small numerical variations $\sim 10^{-9}$-10^{-8} for $\Delta t = 0.1$.

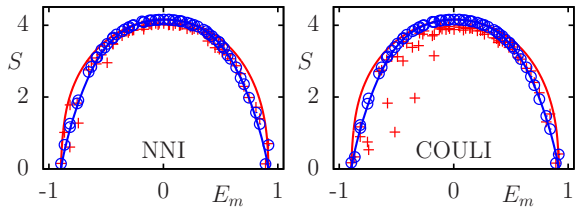

FIG. S15: Dependence of entropy on energy $S(E)$ for both modified interaction models NNI, COULI, parameters $\beta = 1$, $N = 64$ and localized initial conditions (as in Fig. 1). The entropy S is computed from ρ_m obtained by the time average $2^{23} \leq t \leq 2^{24}$ (blue $\circ$) or $2^{11} \leq t \leq 2^{12}$ (red $+$). Both panels have to be compared with Fig. 1(c) which corresponds to the same values of β, N and average interval for t but for the onsite interaction.

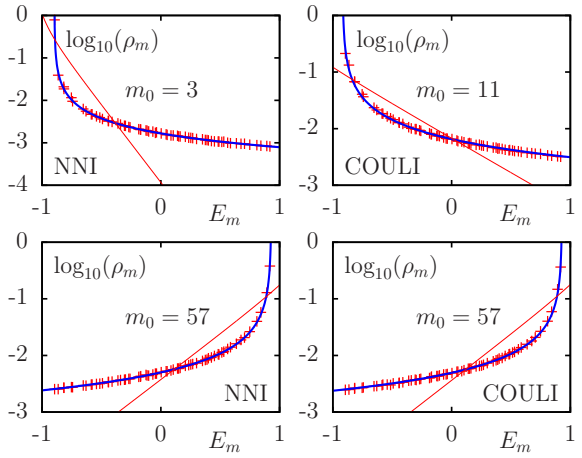

FIG. S16: Dependence of ρ_m on E_m for both modified interaction models NNI and COULI and two states for each case with initial state $m_0 = 3, 57$ for NNI and $m_0 = 11, 57$ for COULI obtained by an time average in the interval $2^{23} \leq t \leq 2^{24}$ (similar states and same parameters, $\beta = 1, N = 64$, as in Fig. 3 except for the modified interaction model). As in Fig. 3 the blue curve shows theory of the EQ ansatz with with $\rho_{\rm EQ}(E) = T/(E - \mu)$ and the red line shows the BE ansatz $\rho_{\rm BE}(E) = 1/(\exp[(E - \mu)/T] - 1)$ with μ and T determined from the norm and energy conservation as explained in the main text below Eq. (2).

The entropy dependence on energy $S(E)$ shown in Fig. S15 clearly confirms for both modified interaction models a thermalization to the classical EQ ansatz. The data points have to be compared with Fig. 1(c) which corresponds to the same values of β, N and same time average intervals for t but for the onsite interaction. The secondary set of data points for the reduced time interval, $2^{11} \leq t \leq 2^{12}$, is actually closer to the theoretical EQ-curve as compared to Fig. 1(c) showing that the thermalization time scale is even reduced, in particular for the NNI case. For the longer time scale, $2^{23} \leq t \leq 2^{24}$, the data points lie nearly perfectly on the theoretical EQ-curve. Actually, our numerical data show that, for $\beta = 1$, the thermalization is already very good for $t \geq 2^{15}$.

Fig. S16 shows two examples for each modified interaction model NNI and COULI of the dependence of ρ_m on E_m for similar initial values m_0 as in Fig. 3. Also here the data matches perfectly the theoretical EQ-curves. More detailed figures for the full set of initial conditions, both modified interaction models, and $\beta = 0.5, 1$, $N = 64$ are available at [40].

Therefore, the results presented in both figures clearly show that for both modified interaction models governed by Eq. (S.6), the steady-state of the system is still very well described by the dynamical thermal distribution corresponding of Eq. (2) for the EQ case (with T and μ determined by two implicit equations as explained below Eq. (2)).

The physical reasons why a modification of the interaction range does not affect the steady-state thermal distribution are (i) the theory of section I does not depend on the particular choice of the interaction, as long as it mixes the linear modes and (ii) the generic features of the linear RMT term corresponding to "ergodic linear oscillator eigenmodes" (i.e. "ergodic" in one particle quantum/oscillator space) such that all types of moderate interactions lead to a nonlinear coupling of these modes with randomly fluctuating amplitudes (the same holds for the DANSE model if the "linear quantum" localization length is comparable to the system size N, see also [28,30]).

We also point out that the dynamical thermal distribution EQ (2) has been observed in experiments with multimode optical fibers (see [45], Ref. S3, Ref. S4).

SupMat References

These Refs.S1-S6 appear in the main part of the article numbered as [12-17] respectively.

Ref.S1. P. Aschieri, J. Garnier, C. Michel, V. Doya, and A. Picozzi, *Condensation and thermalization of classsical optical waves in a waveguide*, Phys. Rev. A **83**, 033838 (2011).

Ref.S2. K. Baudin , A. Fusaro, K. Krupa, J. Garnier, S. Rica, G. Millot, and A. Picozzi, *Classical Rayleigh-Jeans condensation of light waves: pbservation and thermodynamic characterization*, Phys. Rev. Lett. **125**, 244101 (2020).

Ref.S3. E.V. Podivilov, F. Mangini, O.S. Sidelnikov, M. Ferraro, M. Gervaziev, D.S. Kharenko, M. Zitelli, M.P. Fedoruk, S.A. Babin, and S. Wabnitz, *Thermalization of orbital angular momentum beams in multimode optical fibers*, Phys. Rev. Lett. **128**, 243901 (2022).

Ref.S4. F. Mangini, M.Gervaziev, M. Ferraro, D.S. Kharenko, M. Zitelli, Y. Sun, V. Couderc, E.V. Podivilov, S.A. Babin, and S.Wabnitz, *Statistical mechanics of beam self-cleaning in GRIN multimode optical fibers*, Optics Express **30(7)**, 10850 (2022).

Ref.S5. V.E. Zakharov, V.S. L'vov, and G. Falkovich, *Kolmogorov spectra of turbulence I*, Springer, Berlin, (1992).

Ref.S6. S. Nazarenko, *Wave turbulence*, Lectures Notes in Physics, Springer, New York (2011).

Ref.S7. A. Altland, and B. Simons, *Condensed Matter Field Theory*, p.58 Cambridge Univ. Press, Cambridge UK, (2006).