Prethermalization in Fermi-Pasta-Ulam-Tsingou chains

Gabriel M. Lando ● and Sergej Flach ●

Center for Theoretical Physics of Complex Systems, Institute for Basic Science (IBS), Daejeon 34126, Korea

The observation of the Fermi-Pasta-Ulam-Tsingou (FPUT) paradox, namely the lack of equipartition in the evolution of a normal mode in a nonlinear chain on unexpectedly long times, is arguably the most famous numerical experiment in the history of physics. Seventy years after the original publication, most studies in FPUT chains still focus on long wavelength initial states similar to the original paper. It is shown here that all characteristic features of the FPUT paradox are rendered even more striking if modes with short(er) wavelengths are evolved instead. Since not every normal mode leads to equipartition, we also provide a simple technique to predict which modes, and in what perturbation order, are excited starting from an initial mode (root) in α-FPUT chains. The excitation sequences associated with a root are then numerically shown to spread energy at different speeds, leading to prethermalization regimes that become longer as a function of mode excitation number. This effect is visible in observables such as mode energies and spectral entropies and, surprisingly, also in the time evolution of invariant quantities such as Lyapunov times and Kolmogorov-Sinai entropies. Our findings generalize the original FPUT experiment, provide an original look at the paradox's source, and enrich the vast literature dedicated to studying equipartition in classical many-body systems.

I. INTRODUCTION

The numerical experiment of Fermi, Pasta, Ulam, and Tsingou (FPUT) [1] was simultaneously the first computational approach in the study of out-of-equilibrium many-body systems and a source of considerable headache shared by generations of physicists in the past 70 years [1–4]. A proper understanding of FPUT's findings—which contradicted the standard knowledge of the time and were considered paradoxical—relies heavily on modern mathematical and physical tools unavailable to the original authors, e.g., a mature formulation of Kolmogorov-Arnol'd-Moser (KAM) theory [5–7], resonance overlap criteria [8], and Birkhoff-Gustavson normalization [9–13], and also on the power of modern computers. Sitting at the intersection of several fields, it is no surprise that this topic has drawn sustained interest from researchers in nonlinear dynamical systems and statistical mechanics.

The original FPUT experiment consists of evolving the fundamental mode of a harmonic chain in a weakly perturbed, nonintegrable system in which this mode is no longer a stationary state. More specifically, the authors studied the evolution of the first normal mode of a discretized string governed by the Hamiltonian

$$H(\boldsymbol{p},\boldsymbol{x}) = \sum_{j=0}^{N}\left[\frac{p_j^2}{2} + \frac{(x_{j+1}-x_j)^2}{2} + \frac{\alpha(x_{j+1}-x_j)^3}{3}\right], \quad (1)$$

where p_j and x_j are the momentum and position of each discretized mass in the chain, $\alpha=1/4$ and fixed boundary conditions were used, namely $x_0=x_{N+1}=0$. A very small energy density was chosen for the initial state, i.e., $h=H(\boldsymbol{p},\boldsymbol{x})/N \ll 1$, which limited the contribution of the cubic term and rendered the system a weak perturbation of the integrable harmonic chain obtained by setting $\alpha=0$. Even with such a

weak perturbation, the dynamics under (1) is chaotic, which motivated the original expectation that the energy of the initially excited mode would rapidly spread among all other normal modes, leading to equipartition. However, what was observed was quite different: Not only did energy equipartition fail to occur over long timescales, but the system exhibited near recurrences, with energy periodically returning to the initial mode and many other modes remaining only weakly excited throughout the evolution [14].

Today, the convergence of numerous studies provides a coherent explanation for the originally observed lack of energy equipartition in finite, weakly anharmonic systems (for reviews, see Refs. [2,4,15–18])—although speculations on whether or not the original FPUT setup would eventually reach equilibrium have become a prominent topic in metaphysics. First, it is important to note that the energy density chosen by FPUT was extremely small; indeed, a subsequent investigation by Chirikov and Izrailev showed that larger energy densities do lead to energy equipartition [19]. Second, KAM theory was later shown to apply to many FPUT-like systems, rigorously proving the existence of an energy threshold below which equipartition does not take place [20–23]. Third, and most relevant to this manuscript, is the fact that the initial condition used by FPUT is highly *atypical*, corresponding to a stationary solution of the nearby harmonic chain [24]. Moreover, this atypical initial state lies close to exact, time-periodic stationary solutions of the anharmonic system, known as *q-breathers* [25,26]. This stands in sharp contrast to *typical* initial conditions which, when expressed in the normal mode basis, involve combinations of all of its elements. Since these modes form the canonical action variables of the harmonic chain, perturbing a typical initial condition leads to an immediate coupling of all actions, whereas in the FPUT case, only a single action is perturbed. It is therefore natural

that the resulting time evolution in the original FPUT study was highly atypical, reflecting the atypicality already present in the initial state itself.

A further important point is that atypical initial states are not all alike. Their time evolution depends sensitively on which initial mode, or *root*, is excited. The dynamics of roots is characterized by excitation sequences that spread in mode space and, depending on relatively simple algebraic rules, completely or only partially fill it. The case of complete filling corresponds to *thermal* roots, for which time-dependent observables generally reach equipartition. In contrast, if the root excites only a subset of modes, known as its *bush* [27], then equipartition fails to occur, and the corresponding root is termed *nonthermal*.

A particularly striking example of nonthermal roots is provided by *q*-breathers: families of exact, time-periodic solutions that remain localized in mode space and reduce to a single-mode excitation as the anharmonicity tends to zero [25,26]. These states show that certain normal modes and their perturbations can remain stationary even in the presence of nonlinearity, offering remarkable counterexamples to the expectation that observables in nonlinear classical many-body systems necessarily undergo some form of equilibration. They also reveal that ergodicity-breaking behavior can arise not only as a function of decreasing energy density, as in the original FPUT experiment, but also as a consequence of the choice of initial state, as we shall explore in detail here.

This manuscript is devoted to investigating the properties of atypical initial conditions (normal modes) in FPUT chains, and also comparing them to the ones of typical (random) ones. We provide explicit algebraic rules to determine which modes are excited given an initial root, together with their short-time perturbation order in α. We then numerically investigate the time evolution of thermal roots of several different excitation numbers, showing that the paradox associated to the fundamental mode is substantially maximized for higher excited roots, e.g., the prethermal regime and its metastable plateaus become increasingly longer as a function of root excitation number for a fixed energy density. This is shown both by a direct tracking of mode energies as a function of time and by computing equipartition times from spectral entropies.

We also show that the root excitation number impacts quantities that do not depend on the initial state. Indeed, albeit Lyapunov times and Kolmogorov-Sinai (KS) entropies being invariant with respect to a large number of transformations and being constant in an ergodic system regardless of the initial state used for computing them [28], we show here that their *transient* properties are strikingly different from those of typical initial states and clearly reflect the fact that the system is trapped in phase space [24]. More specifically, both the maximal Lyapunov exponent (the inverse of the Lyapunov time) and the Kolmogorov-Sinai entropy show a dip in their time evolution when computed from roots, which is absent if the initial state chosen is typical. Moreover, the duration and depth of this dip increases with excitation number, showing that excited modes remain trapped for even longer in near-integrable portions of phase space, likely due to their proximity to *q*-breathers.

The paper is structured as follows. Section II discusses the model, which is entirely focused in Eq. (1), introducing normal modes, spectral entropies and Lyapunov data which will be thoroughly employed in the rest of the paper. Section III presents the rules for computing excitation sequences for roots, which are then evolved in an FPUT chain with $N = 63$ masses in Sec. IV. A discussion of our findings can be found in Sec. V, together with a conclusion in Sec. VI. An Appendix is also included providing examples of pathological numerical behavior due to bush instability.

II. MODEL AND OBSERVABLES

We restrict our analysis to cubic FPUT chains, also known as the α-FPUT model, with Hamiltonian (1). In the following we recall how to rewrite the FPUT Hamiltonian in normal mode (or phonon, or eigenmode) basis, and introduce quantities such as mode energies, spectral entropies and equipartition times. We then describe how to obtain important invariant quantities that will be used in the following sections, such as Lyapunov times and Kolmogorov-Sinai entropies.

A. Normal mode basis

Since Eq. (1) has N degrees of freedom its harmonic limit has a set of N normal modes, denoted $\{P_j, Q_j\}_j$, with which we can rewrite the position and momentum in Eq. (1) at any time t as

$$\begin{pmatrix} p_j(t) \\ x_j(t) \end{pmatrix} = \sqrt{\frac{2}{N+1}} \sum_{n=1}^{N} \begin{pmatrix} P_j(t) \\ Q_j(t) \end{pmatrix} \sin\left(\frac{\pi jn}{N+1}\right), \quad (2)$$

each mode having frequency and energy

$$\omega_j = 2\sin\left[\frac{\pi j}{2(N+1)}\right], \quad E_j = \frac{P_j^2 + \omega_j^2 Q_j^2}{2}. \quad (3)$$

In these canonical coordinates, the Hamiltonian (1) takes the form [20]

$$H(\boldsymbol{P}, \boldsymbol{Q}) = \frac{1}{2}\sum_{j=1}^{N}\left(P_j^2 + \omega_j^2 Q_j^2\right) + \frac{\alpha}{3\sqrt{2(N+1)}}$$

$$\times \sum_{i,j,k=1}^{N} B_{i,j,k}\, \omega_i\, \omega_j\, \omega_k\, Q_i Q_j Q_k, \quad (4)$$

where [29]

$$B_{i,j,k} = \delta_{i+j,k} + \delta_{k+i,j} + \delta_{j+k,i} - \delta_{i+j+k,2(N+1)}. \quad (5)$$

Note that $B_{i,j,k}$ is invariant with respect to permutations of its indices.

B. Spectral entropy

Equation (1) is much easier to numerically deal with than Eq. (4). However, we are interested in several properties that need to be computed in the normal mode basis, such as the mode energies themselves and the [normalized] spectral entropy

$$\eta(t) = \frac{S(t) - \log N}{S(0) - \log N}, \quad (6)$$

where the Shannon entropy and normalized energies are given, respectively, by

$$S(t) = -\sum_{j=1}^{N} \rho_j(t) \log \rho_j(t), \quad \rho_j(t) = \frac{E_j}{\sum_{j=1}^{N} E_j}. \quad (7)$$

Spectral entropy has been used several times to quantify equipartition in nonlinear systems [4,24,30,31]. The logic for employing it is simple: If equipartition is reached, then $\rho_j(t)$ will be stationary and the same for every mode j, freezing $\eta(t)$ at some value $\langle \eta \rangle$. If one assumes a Gibbs distribution at equilibrium, then $\langle \eta \rangle$ can be calculated analytically as a function of the Shannon entropy of the initial state and is given by

$$\langle \eta \rangle = \frac{1 - \gamma}{\log N - S(0)}, \quad (8)$$

where $\gamma \approx 0.5772$ is the Euler constant [32]. Since $\eta(0) = 1$ for any initial state, one expects $\eta(t)$ to decrease and follow a complicated and possibly highly oscillatory evolution until eventually hitting its stationary value. Once this happens the dynamics cannot, at least on average, depart significantly from thermal oscillations around $\langle \eta \rangle$. It is then natural to consider that equipartition is reached once $\eta(t)$ hits its equilibrium value for the first time [31]. Such first-passage time furnishes our definition of *equipartition time*, τ_{eq}, and is appealing because it does not require any *ad hoc* assumptions.

C. Lyapunov invariants and Kolmogorov-Sinai entropy

Equipartition time as defined in Sec. II B is, evidently, attached to the observables being monitored, namely the mode energies. Different observables will generally reach equipartition at different times (if at all), such that observable equipartition does not actually measure properties intrinsic to the system. Nevertheless, since mode energies are conserved in the unperturbed harmonic chain, one knows well how these observables should behave upon approaching an integrable limit, i.e., equipartition should slow down and eventually stop for sufficiently small energy densities.

There is, however, an observable-independent timescale that is completely intrinsic to the system and should not depend on the initial state, given by the Lyapunov time (the inverse of the maximal Lyapunov exponent). Since, like the maximal Lyapunov exponent (MLE), all exponents in the Lyapunov spectrum (LS) are invariant with respect to homeomorphisms [28], the LS provides a valuable set of timescales related to the average "pull" along each 1-dimensional submanifold that forms the hyperbolic skeleton of the system's dynamics in phase space [33]. The information in the LS can also be condensed in the form of a scalar, the Kolmogorov-Sinai entropy, which by Pesin's theorem can be obtained as the sum of all positive exponents in the LS [34].

To compute the LS we employ the well-known prescription of [35,36], which amounts to computing the spectrum of the symmetric matrix

$$\Gamma(\boldsymbol{p}_0, \boldsymbol{x}_0) = \lim_{t \to \infty} [\mathcal{M}^T(\boldsymbol{p}_0, \boldsymbol{x}_0; t) \mathcal{M}(\boldsymbol{p}_0, \boldsymbol{x}_0; t)]^{1/2t}. \quad (9)$$

In the above, $\mathcal{M}$ represents the system's stability matrix, obtained by solving Hamilton's equations in tangent space:

$$\frac{\mathrm{d}\mathcal{M}(\boldsymbol{p}, \boldsymbol{x}; t)}{\mathrm{d}t} = \mathcal{J}\text{Hess}[H(\boldsymbol{p}, \boldsymbol{x})]\mathcal{M}(\boldsymbol{p}, \boldsymbol{x}; t), \quad (10)$$

where Hess is the Hessian with respect to $(\boldsymbol{p}, \boldsymbol{x})$. Since the stability matrix has to be computed along a trajectory, the above equation is solved in parallel with Hamilton's equations for an initial phase-space point $(\boldsymbol{p}_0, \boldsymbol{x}_0)$ until a final time long enough to result in a converged LS. During this evolution, one must also perform QR-diagonalizations of $\mathcal{M}$ to avoid numerical denegenacies, as is well-known and described in, e.g., Eq. [36].

As can be seen in Eq. (9), the stability matrix will generally depend on the trajectory along which it is calculated, i.e., on the initial phase-space point $(\boldsymbol{p}_0, \boldsymbol{x}_0)$. However, for a sufficiently large number of degrees of freedom and energy density, phase space will consist of a single metrically intransitive set that is equally accessible to all trajectories in the system [37,38]. Equivalently, one can formulate this previous property as stating that, for sufficiently large N and $h = H(\boldsymbol{p}, \boldsymbol{x})/N$, the system lies outside the KAM regime in which phase space can be split into a chaotic web and pockets of near-integrability. Thus, all trajectories are "free to roam" around the entire accessible phase space, and the LS is expected to be independent of the trajectory along which it is computed. This independence is a stronger evidence of ergodicity than any conclusion drawn from time-dependent observables, and can be verified more succinctly by comparing KS entropies instead of full LS.

III. EXCITATION SEQUENCES

We start this section by writing an abridged version of Newton's second law for a given mode i, obtained from Eq. (4):

$$\ddot{Q}_i + \omega_i^2 Q_i \sim \alpha \sum_{j,k=1}^{N} B_{i,j,k} Q_j Q_k. \quad (11)$$

Thus, the mode Q_i is excited by a single mode Q_j or a mixture of modes Q_j and Q_k. Excitations from single modes have $j=k$ and the coupling tensor assumes the form

$$B_{i,j,j} = \delta_{i+j,j} + \delta_{j+i,j} + \delta_{2j,i} - \delta_{i+2j,2(N+1)}$$
$$\implies \quad B_{i,j,j} = \delta_{i,2j} - \delta_{i,2(N+1)-2j}, \quad (12)$$

where the first two terms vanished because j and i are never zero. Now, for mixed excitations we have $j \neq k$, so

$$B_{i,j,k} = \delta_{i+j,k} + \delta_{k+i,j} + \delta_{j+k,i} - \delta_{i+j+k,2(N+1)}$$
$$\implies \quad B_{i,j,k} = \delta_{i,k+j} + \delta_{i,|k-j|} - \delta_{i,2(N+1)-(j+k)}. \quad (13)$$

The above expression is composed of an ascending term, namely $\delta_{i,k+j}$, which dictates which higher-order mode to excite starting from j and k; a right reflection term, $\delta_{i,2(N+1)-(j+k)}$, which dictates which term to excite if $j+k > (N+1)$, i.e., if the next mode is too large and leaves mode space; and a descending term, $\delta_{i,|k-j|}$, which propagates the excitation backward in mode space and is also responsible

for controlling left reflections, which take place when the next excited mode is smaller than zero.

A. First excited mode and *q*-breathers

Initializing the system in a root Q_r means setting $Q_j = 0$ unless $j = r$. The first excitation cycle, therefore, starts with a lone mode Q_r, which by (12) excites the modes $i = 2r$ and $i = 2(N-r+1)$. If the first condition is met, then the second δ will be one iff $r = (N+1)/2$, with $N+1$ even. Thus, if the chain has an odd number of sites and the chosen root is $Q_{(N+1)/2}$, then all coefficients of the coupling tensor vanish and the energy remains forever localized in the root: this mode is a *q*-breather.

Now, assuming that only the second Kronecker δ in Eq. (12) vanishes, then the only mode that is excited starting from Q_r is Q_{2r} [26,39,40]. If, however, only the first δ vanishes, then the first excited mode is obtained after a reflection and is given by $Q_{2(N-r+1)}$. These scenarios are fundamentally identical but, for simplifying writing excitation sequences, we will always assume that the root mode fulfills $2(r-1) < N$, so that the first excited mode will always be Q_{2r}.

Moving foward, a naïve perturbative argument allows us to estimate how the amplitude of Q_{2r} scales with the anharmonicity parameter, α, for short times. It goes as follows: Since Q_r carries all the initial energy in the system and α is small, the solution $Q_r(t)$ starts close to a sine or a cosine, as can be seen by setting $\alpha = 0$ in Eq. (11). This approximation scales as $\mathcal{O}(\alpha^0)$ [26]. Now, Q_{2r} emerges from the product $Q_r Q_r$ in Eq. (11), which is also of 0th order in α, but the α multiplying the right-hand side of Eq. (11) gives $Q_{2r}(t)$ a $\mathcal{O}(\alpha^1)$ dependence for short times [41].

B. First ascending sequence

The second excited mode is obtained from Q_r and Q_{2r}. Evidently, Q_{2r} excites Q_{4r} due to Eq. (12) and, since Q_{2r} is of $\mathcal{O}(\alpha^1)$, the amplitude of $Q_{4r}(t)$ goes as $\alpha\, Q_{2r}Q_{2r} = \mathcal{O}(\alpha^4)$ for short times. Mode $Q_{3r=r+2r}$ is also excited according to Eq. (13), but now the short-time amplitude will scale as $\alpha\, Q_r Q_{2r} = \mathcal{O}(\alpha^2)$. By induction, the first ascending sequence for a root Q_r is given by

$$\{Q_r, Q_{2r}, Q_{3r}, Q_{4r}, Q_{5r}, \ldots, Q_{ar}\}. \qquad (14)$$

It includes all multiples of the root mode ordered by how their short-time dynamics scales with α, namely $\mathcal{O}(\alpha^{a-1})$, until a reflection takes place in Eq. (13), i.e., until $j+k = 2(N+1)$ for some Q_j and Q_k. The reflection, however, will only excite new modes if $2(N+1)-(j+k)$ is not a multiple of r, since otherwise the terms would have already been excited during the ascending sequence.

C. First descending sequence

At the end of the first ascending sequence every term is a multiple of the root, r. Thus, at the reflection, substituting $j+k = a\,r$ in the last δ of Eq. (13) singles out $r = 2(N+1)/a$, allowing us to state two important facts:

(1) If r divides $2(N+1)$, then the first ascending sequence is reflected into itself. This shows, in particular, that Q_2 is nonthermal for all N, since $2(N+1)$ is always divisible by 2

[39,40]. It also shows that Q_1 is a *trivial* thermal root, since it excites all modes in the first ascending sequence for any N;

(2) The reflection takes place when a multiple of r is larger than N for the first time, i.e., when $Q_j = Q_{(a+1)r}$. Evidently this mode lies outside mode space and the excited mode is actually $Q_{2(N+1)-(a+1)r}$ after the reflection. A consequence is that even roots can never excite odd modes, since for even modes $2(N+1)-(a+1)r$ is always even [40]. This shows that Q_{2r} are nonthermal roots for all r and all N.

Once a reflection occurs, the selection rule for $|j-k|$ generates a descending sequence starting from $Q_{2(N+1)-(a+1)r}$, namely,

$$(Q_{2(N+1)-(a+2)r}, Q_{2(N+1)-(a+3)r}, \ldots). \qquad (15)$$

The descent continues in multiples of r until the next mode would exit the chain, i.e., until $2(N+1) - (a+b)r < 0$. Despite the reflection, it is important to note that each term in the descending sequence comes from a product like $Q_{ar}Q_{br}$, and will therefore have an amplitude that scales perturbatively as $\mathcal{O}(\alpha^{a+b-1})$ for short times, just as in the ascending sequence. This shows both ascending and descending sequences can be tied together and monotonically ordered based on how their short-time solutions scale with α.

D. Full excitation sequences

Once the first descending sequence is over, another reflection takes place, now going from left to right, mediated by $\delta_{i,|j-k|}$ in Eq. (13). Substituting the last mode number from the descending sequence in this δ, we see that the excited mode after this reflection is $Q_{2(N+1)-(a+b)r+r}$. The second ascending sequence will start from this mode number, once again ascending in steps of r until $2(N+1) - (a+b-c)r > N$. Then, we have another reflection from left to right that excites mode $2(N+1) - [2(N+1) - (a+b-c)r - r] = (a+b-c+1)r$, and so on. Instead of writing full abstract sequences, we now provide a few examples showing how easy it is to write them in practice.

Let us start with a chain with $N = 11$ masses. Since $2(N+1) = 24$, the only three nontrivial thermal roots in this chain are Q_5, Q_7 and Q_{11}. The root Q_3 is evidently nonthermal since it divides 24 (cf. Sec. III C), but let's take a look at its excitation sequence. Denoting a sequence starting from root r in a system with N normal modes by S_r^N, the first ascending sequence is $S_3^{11} = \{Q_3, Q_6, Q_9\}$, and we stop here because the next term falls outside mode space. The reflection term comes from $i = 2(N+1)-(6+9) = 9$, which is already excited. This will be now propagated backwards to Q_6 and Q_3 according to Eq. (13), retracing the ascending sequence and never exciting any new mode. Now, for the leftmost reflection, the excited mode is $Q_{|3-3|}$, which does not exist. Thus, the bush of Q_3 is given by $\{Q_3, Q_6, Q_9\}$. This contradicts a statement made in Ref. [40], which states that all odd modes should be excited if r and N are relatively prime, but is verified numerically in Fig. 1(a).

Let us now consider what happens to Q_3 for $N = 12$ instead. The first ascending sequence is now updated to $\{Q_3, Q_6, Q_9, Q_{12}\}$. The reflection gives $i = 2(N+1)-(12+3) = 11$, which was not previously excited and is not a multiple of r. The next excited modes
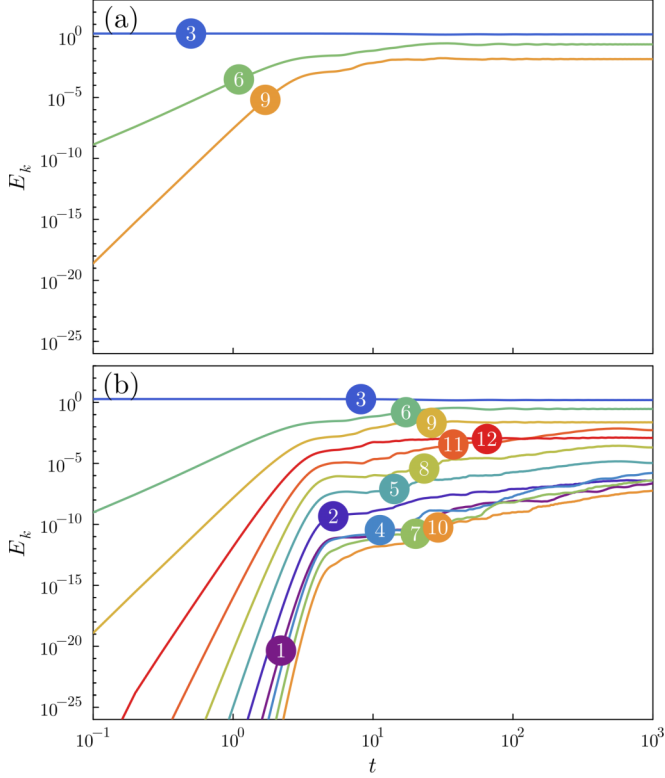
FIG. 1. Evolution of mode energies for the root Q_3 in an α-FPUT chain with (a) $N = 11$ and (b) $N = 12$ masses, both for $h \approx 0.16$ and $\alpha = 1/4$. The root Q_3 is only thermal in the latter case, while in the former its bush is composed of all multiples of 3 that are smaller than 11. The mode excitation sequence in panel (b) is the same as the one predicted using the coupling coefficients, Eq. (17).

will then be $11-3 = 8$, $11-6 = 5$, etc, and the full excitation sequence is

$$S_3^{12} = \{Q_3, Q_6, Q_9, Q_{12}, Q_{11=26-(12+3)},$$
$$Q_8, Q_5, Q_2, Q_{1=|2-3|}, Q_4, Q_7, Q_{10}\}, \qquad (16)$$

where modes are ordered by how their short-time solutions scale with α. This excitation sequence is seen to match numerics in Fig. 1(b).

As another example, for S_3^{13} we have

$$S_3^{13} = \{Q_3, Q_6, Q_9, Q_{12}, Q_{13=28-(12+3)},$$
$$Q_{10}, Q_7, Q_4, Q_1, Q_{2=|1-3|}, Q_5, Q_8, Q_{11}\}, \qquad (17)$$

and, for S_5^{13},

$$S_3^{13} = \{Q_5, Q_{10}, Q_{13=28-(10+5)}, Q_8, Q_3, Q_{2=|3-5|},$$
$$Q_7, Q_{12}, Q_{11=28-(12+5)},$$
$$Q_6, Q_1, Q_{4=|1-5|}, Q_9\}. \qquad (18)$$

E. Mode connections and amplitudes

Excitation sequences such as Eqs. (17) and (18), and also the trivial sequence generated by Q_1, completely fill mode space for $N = 13$. However, these sequences do not provide insight into how energy is exchanged between excited modes—an aspect we will briefly explore in the following.

Starting from Eq. (17), we see that the pair that most strongly connects to Q_{13} in S_5^{13} is (Q_5, Q_{10}), since the right-hand side of Eq. (11) gives $\alpha Q_5 Q_{10} = \mathcal{O}(\alpha^2)$. There are also connections formed by other pairs, such as Q_5 and Q_8, but these connections are of higher order in α, e.g., $\alpha Q_8 Q_5 = \mathcal{O}(\alpha^4)$. Thus, the short-time regime will see Q_{13} exchanging energy primarily with Q_5 and Q_{10}. The situation is similar in S_3^{13} in Eq. (18), where (Q_{12}, Q_3) and (Q_9, Q_6) both connect to Q_{13} and have weighs of $\mathcal{O}(\alpha^4)$. The case of S_1^{13} is rather extreme, since the sum or difference of any two mode numbers that equals 13 or 15 all connect to Q_{13}, and all connections scale as $\mathcal{O}(\alpha^{12})$.

The argument above sheds light on why energy remains localized in a subset of modes for short times, but the simplified form of Eq. (11) glosses over the multiplier $\omega_i \omega_j \omega_k$ in Eq. (4). These frequencies ω, as seen in Eq. (3), increase monotonically for higher mode numbers. A consequence is that, despite Q_3 in S_1^{13} and Q_{13} in S_5^{13} both scaling perturbatively as $\mathcal{O}(\alpha^2)$, the amplitude of Q_{13} in S_5^{13} will lie above the one of Q_3 in S_1^{13} for short times if the roots are excited with the same energy density. The same conclusion could be drawn from noticing that, for a fixed energy density, energy has to be shared among a smaller subset of excited modes in S_5^{13} when compared to S_1^{13} in the short-time regime, rendering local energies larger. These larger energies take longer to spread to other modes, since Q_{13} in S_5^{13} has exactly the same *number* of connections as Q_3 in S_1^{13} (and they are also of the same order). Note also that the next thermal root in S_r^{13} would be $r = 7$, but this mode is a q-breather for $N = 13$. Thus, increasing mode number is tantamount to approaching a state for which energy does not spread in mode space, which offers a more qualitative way of interpreting the slowing down of energy equipartition for increasing r.

IV. NUMERICAL EXPERIMENTS

In this section we perform computations in the α-FPUT chain (1) with $N = 63$ sites. This value of N is chosen because $2(N+1) = 128$ is not divisible by any odd integer other than 1, such that all odd roots are thermal for this chain. Although initial states, final propagation times and energy densities will vary, evolution is always computed using an optimized second order symplectic integrator with step size $\Delta t = 0.2$ [42]. This results in a maximum relative energy error of around 10^{-3} for the largest energy densities chosen, and much smaller for most of them. We will focus on the evolution of random initial conditions and the first five odd roots Q_1, Q_3, Q_5, Q_7, and Q_9. The sampling of all initial conditions starts by setting $x(0) = 0$. Then, for random states, each momentum component is drawn from a Maxwell-Boltzmann distribution, while for mode Q_k we have momentum components fixed as $p_j(0) = \sin[\pi k j/(N+1)]$. Once sampled, momenta are then renormalized to achieve the desired energy density for both types of initial conditions.

A. Metastable plateaus

Section III described how energy can remain localized within a small subset of modes before spreading throughout mode space: The higher excited the root, the smaller the subset
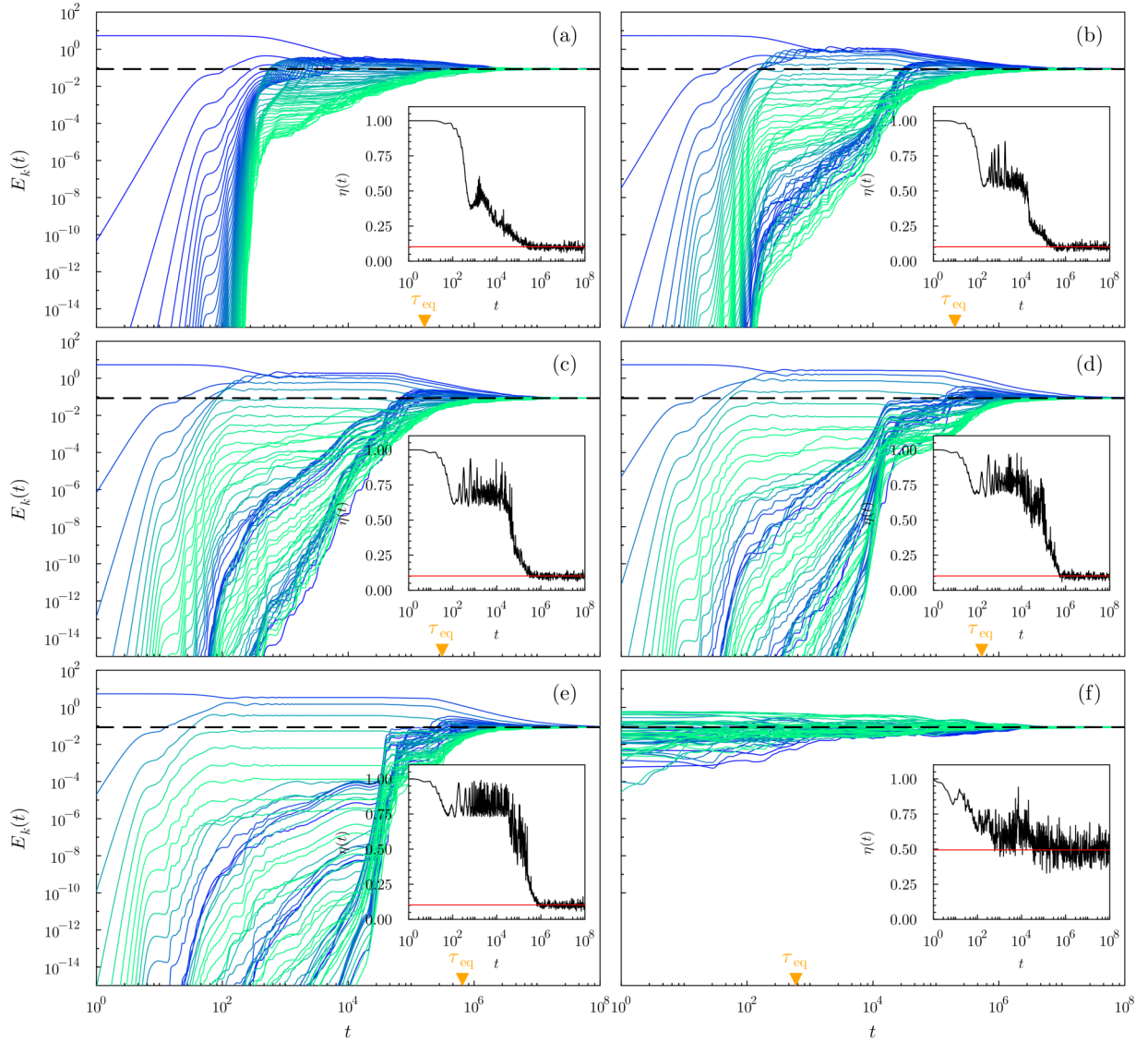
FIG. 2. Evolution of mode energies for the first five odd modes, namely: (a) Q_1; (b) Q_3; (c) Q_5; (d) Q_7; (e) Q_9; and (f) a random initial condition, all with energy density $h = H(\boldsymbol{p}, \boldsymbol{x})/N \approx 0.085$ in an α-FPUT chain with $N = 63$ and $\alpha = 1/4$. Line colors correspond to mode frequencies, going from lower (blue) to higher (green), with their equilibrium value shown as a black dashed line. Equipartition times (orange triangles) are defined as the instant the time-evolving spectral entropy, shown in black in the inset, first attains its equilibrium value $\langle \eta \rangle$, given by the horizontal red line.

of excited modes and the larger their amplitudes, such that more time is needed to distribute their energy among other modes. During this transfer time the dynamics is mostly localized within the subset of excited modes, which essentially "wait" and form the metastable plateaus prominent in FPUT literature [14,18,43].

The well-known fact that the metastable plateaus associated to roots become more visible at lower energy densities is unsurprising, since in the limit of zero energy density such modes are normal modes of the associated harmonic lattice. However, the behavior of plateaus for excited roots was not previously addressed. To this end, Fig. 2 shows a comparison of evolving mode energies and spectral entropies for the α-FPUT chain and initial states described in the introduction to this section. Figure 2(a) displays the standard mode energy dynamics found in literature: energy initially placed in

Q_1 monotonically spreads to all other modes and eventually results in energy equipartition, with all E_k's converging to the mean energy $h = H(\boldsymbol{p}, \boldsymbol{x})/N$ (black dashed line). Metastable plateaus are clearly visible in the prethermal regime, in which a measurement would indicate that the system is not ergodic and cannot be described by equilibrium statistical mechanics. This statement is evidently incorrect, as waiting longer would result in a system that does display statistical behavior.

The approach from an out-of-equilibrium initial state toward a thermal one is most clearly seen in tracking the spectral entropy as a function of time, shown in the insets: the moment it touches the red line, which represents its equilibrium value (8), is where we consider equipartition to have been achieved. The associated equipartition times are displayed in the main plots as an orange triangle and slowly increase as we move from Q_1 up to Q_9, with the shortest time corresponding to
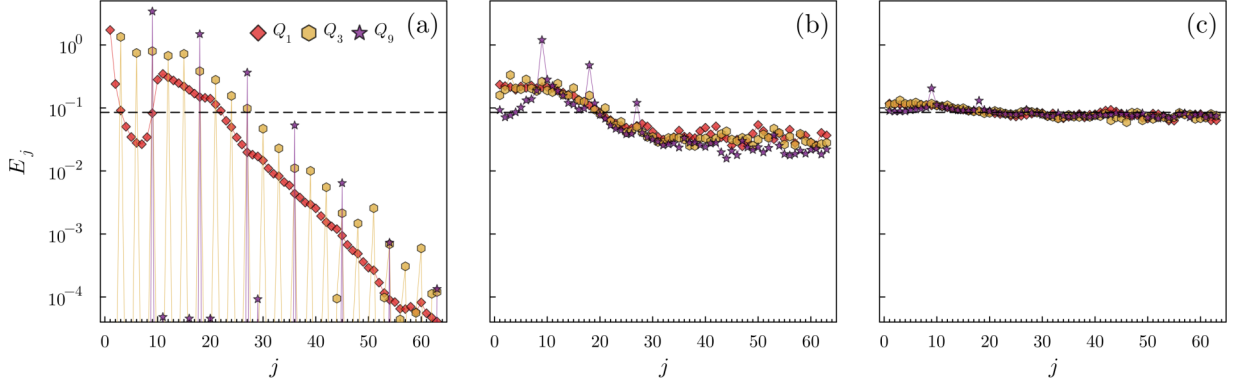
FIG. 3. Mode energies, E_j, as a function of mode number, j, for roots Q_1, Q_3 and Q_9 for fixed times, with their equilibrium (average) value, h, shown as a black dashed line. Each panel is a vertical slice of a panel in Fig. 2 at (a) $t = \tau_{eq}/100$; (b) $t = \tau_{eq}$; and (c) $t = 10\tau_{eq}$ (note that τ_{eq} is different for each root). The short time of panel (a) allows for a clear visualization of the energy cascade flowing between modes that are multiples of 1, 3, and 9, as described in Sec. III. At equipartition most of the site energies have the same order of magnitude, as seen in panel (b), with stronger mixing still taking place at high mode numbers while the lower half of the excitation sequence remains slightly isolated from the other modes. Even after ten equipartition times the nearly homogeneous energy distribution hints at which root was initially excited, as can be seen in the above-average energies of modes Q_9 and Q_{18} for the root Q_9 in panel (c).

the random initial condition. Evidently, this latter case is less interesting since the equipartition time depends strongly on the initial sampling, but for roots the structures formed by mode energies as they evolve in time show that, indeed, higher-excited roots isolate an increasingly smaller subset of modes in mode space. Figure 2(a) also shows a dense subset of modes homogeneously concentrated around the mean, while Fig. 2(e) clearly shows that two modes, namely Q_9 and Q_{18}, remain relatively isolated from the others for much longer times. To see this more clearly, in Fig. 3 we take vertical slices of Fig. 2 at three different fixed times for Q_1, Q_3 and Q_9. At $t = \tau_{eq}/100$ the contents of Sec. III E can be clearly visualized, namely the fact that increasing root number traps energy in a smaller subset of modes with higher amplitudes in the short-time regime. At $t = \tau_{eq}$ low frequency modes that are multiples of 3 and 9 can still be seen to have above-average energies for roots Q_3 and Q_9. At $t = 10\tau_{eq}$ this effect has already dissipated for Q_3, but the energies of modes Q_9 and Q_{18} remain slightly above the mean energy when evolving Q_9, as visible in Fig. 3(c).

Evidently, the longer it takes for a subset of modes to reach equipartition, the longer the prethermal regime. This is clearly seen in Fig. 2: The relatively obfuscated metastability associated to the Q_1-root seen in Fig. 2(a) is already maximized in Fig. 2(b), where Q_3 is evolved, and becomes more and more blatant as higher-excited modes are used as roots. Evidently, the random initial condition in Fig. 2(f) does not display metastability, since here all modes are excited at once and the slow, selection-rule-mediated transferring of energy from excited modes to previously unexcited ones does not take place. Since the mean energy is a constant of motion, the prethermal regime's duration can also be visualized by plotting the time-dependent standard deviation of mode energies in Eq. (2),

which will be approximately constant while the energy is still strongly localized in a subset of normal modes. This is indeed to be true in Fig. 4, where it is also clear that the metastability's lifetime increases monotonically as a function of root number for this energy density.

B. Equipartition and Lyapunov times

Spectral entropy (6) is a function of the initial state, so it is no surprise that the equipartition times computed from it in Fig. 2 depend on it. These times have no direct relation to the duration of prethermalization, except for the fact that they are necessarily longer than the length of metastable plateaus. Since random initial conditions do not, at least generally, undergo prethermal regimes, it is expected that the equipartition times of such typical initial conditions will be shorter than the ones of roots for any energy density chosen. This expectation is confirmed in Fig. 5, where we compare equipartition times for 280 typical conditions with the ones obtained from odd roots as a function of the energy density. Clearly, the former are bounded from above by the latter, which slowly increase as a function of excitation number.
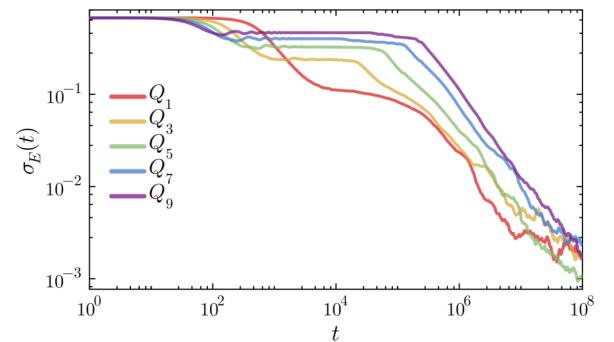

FIG. 4. Standard deviation of mode energies in Fig. 2 as a function of time for $h \approx 0.085$. Despite keeping the energy density fixed, the prethermal regime becomes longer as a function of root number.

$$\sigma_E(t) = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(E_j(t)-h)^2}, \qquad (19)$$
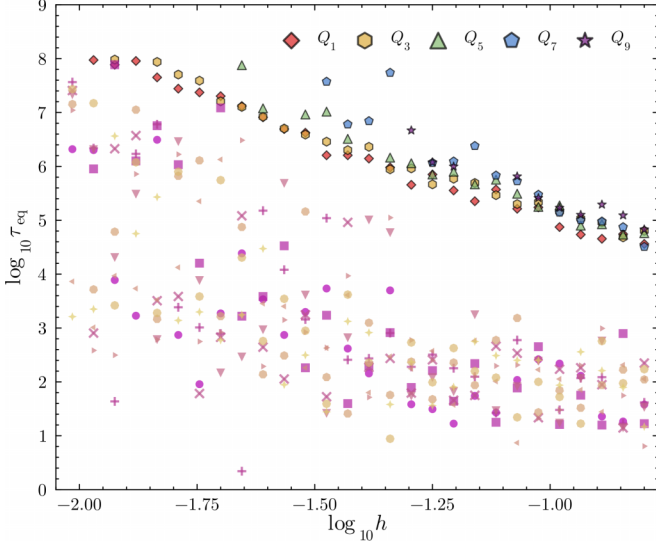
FIG. 5. Equipartition times computed by evolving the first five odd roots (rainbow colors) and 280 random initial states (tones of magenta and orange) at several different energy densities in a chain with $N = 63$ and $\alpha = 1/4$. These data were obtained from spectral entropies computed up to $t = 10^8$, such that the top left equipartition times, which are of $\mathcal{O}(10^8)$, are likely slightly overestimated by being of the order of the computation time. The equipartition time for Q_1 (red diamonds) clearly forms an upper bound that the ones computed from random initial states essentially never cross. The equipartition times computed from excited modes consistently do so, albeit slowly as a function of increasing root number.

If equipartition times reach the order of the propagation time used to compute them, namely $\mathcal{O}(10^8)$, then convergence can no longer be achieved for these parameters, which is clear in Fig. 5 by the presence of missing data for some values of energy density. However, the "disappearance" of points does not increase fully monotonically with mode excitation number, with some data displaying what could be described as "erratic behavior." Since dependence on initial conditions is tantamount to ergodicity breaking, investigating quantities that are proved to be invariant in any ergodic system will shed light into the source of the instabilities seen in Fig. 5. The most important of such quantities is the MLE, λ_1, which indicates the presence of chaotic behavior and is not only invariant with respect to diffeomorphisms but also constant for ergodic dynamical systems [28]. From this exponent one can extract the Lyapunov time, $\tau_1 = 1/\lambda_1$, which measures the time a vector tangent to a trajectory takes to align with the maximally chaotic, one-dimensional submanifold in the tangent bundle [33,44,45]. Since the Lyapunov time is computed from the divergence of two initially close trajectories, it takes a finite time to converge and it is often useful to observe its evolution as a function of time, as done in Fig. 6.

Figures 6(a) and 6(b) display the evolution of $\tau_1(t)$ extracted from propagating odd roots and random initial conditions for an energy density of $h \approx 0.085$. Not only is convergence achieved for all initial states employed, but the time-dependent portrait of both typical and atypical initial conditions is identical. Upon decreasing the energy density, however, Fig. 6(c) shows that atypical initial conditions start
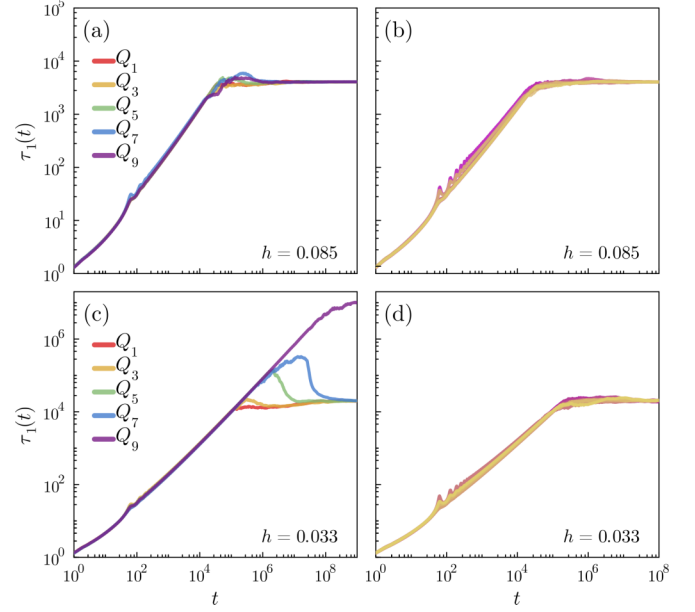

FIG. 6. (a) Time-dependent Lyapunov times computed by evolving the first five odd modes of a chain with $N = 63$ and $\alpha = 1/4$ at a fixed energy density $h \approx 0.085$ up to $t_f = 10^9$. Clearly, at this energy density all modes behave similarly and indistinguishably from the random initial states shown in panel (b), for which the propagation time was set to $t_f = 10^8$. (c) By lowering the energy density down to $h \approx 0.033$, roots start to behave differently as a function of time when compared to the random states used in panel (d). The differences appear to grow with excitation number, resulting in high-order modes such Q_9 not converging at all within this propagation interval (which is more than enough for all other initial states).

behaving very differently not only with respect to typical ones, but also among themselves. While the Lyapunov times obtained from Q_1 and Q_3 are similar to the ones of the random initial conditions in Fig. 6(d), the data obtained from evolving Q_5 is seen to start forming a "belly" before convergence. Thus, it shows that Q_5 lies in a region of phase space that takes longer to align with the maximally chaotic direction, similarly to the trapping times described in Ref. [24]. The difference here is that we observe longer trapping times at fixed energy density as the root number increases. By evolving even higher excited roots such as Q_7, Fig. 6(c) shows that the "belly" increases when compared to the one of Q_5, such that Q_7 is trapped for a time one order of magnitude longer. Extrapolating to Q_9 this transient trapping regime, which is a manifestation of prethermalization in an invariant quantity and therefore *must* disappear for sufficiently long times, ends up being longer than the propagation time and does not converge in our computations. The reason for the longer transient times as we go up from Q_1 to Q_9 is that the higher the excitation number, the smaller the subspace of mode space to which dynamics is approximately restricted during the prethermal regime and the higher the mode energies, as discussed in Sec. III E. Indeed, Fig. 3 shows that a large fraction of the energy is still concentrated in multiples of the initial excitation number at equipartition time, which as seen by comparing Figs. 6 and 7 is always longer than Lyapunov time when dealing with atypical initial conditions. Thus, the trapping time
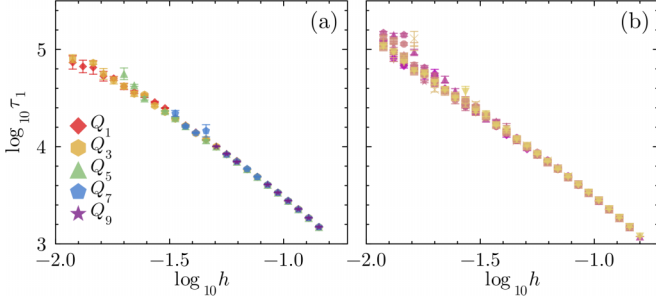
FIG. 7. (a) Lyapunov times computed by evolving the first five odd roots (rainbow colors) of a chain with $N = 63$ and $\alpha = 1/4$ at several different energy densities. Propagation time used was $t_f = 10^9$, for which τ_1's computed by evolving Q_1 converge well for all energy densities used. Other modes, however, did not converge for the whole range of energy densities employed here, requiring longer and unpractical propagation times (see text for explanation). (b) Lyapunov times computed by evolving 10 random initial states at the same energy densities as panel (a). Propagation times in this case were only $t_f = 10^7$, which is enough to obtain converged results starting from typical initial conditions. Error bars in all plots are the standard deviation of the corresponding $\tau_1(t)$ between t_f and $t_f/10$.

will increase with the excitation number, which is also a consequence of higher-excited modes lying closer to q-breathers.

Figure 7, which shows the final (converged) Lyapunov times as a function of energy density for typical and atypical initial conditions, is clear evidence that the system we are dealing with is numerically ergodic: The times are the same, no matter what initial conditions are chosen (although the larger error bars and spread for $h \approx 10^{-2}$ show that $t_f = 10^9$ is barely long enough to reach numerical ergodicity for the smallest values of h used here). This does not mean, as discussed before, that all data for roots converges, with the main reason for failure being the presence of extremely long prethermal regimes as seen in Fig. 6.

C. Kolmogorov-Sinai entropies

Given the imprint prethermalization has in a dynamical invariant such as the Lyapunov time, in Fig. 8 we display the time evolution of one of the most meaningful, yet computationally expensive, invariant quantities in dynamical systems, namely the Kolmogorov-Sinai entropy, κ_{KS}. This quantity, just

like the Lyapunov time and the associated maximal exponent, displays a clear dip when computed by propagating roots, while for a typical initial condition it approaches its converged value from above. Since κ_{KS} is the sum of all positive Lyapunov exponents, what Fig. 8 shows is that the prethermal regime is present in all phase-space directions, not only the maximal one associated to the MLE. Upon decreasing the energy density we also see a tendency of a larger dip for higher-excited modes, in accordance with our expectation that the prethermal regime is more pronounced for higher excitation numbers. It is worth noting that the convergence of Kolmogorov–Sinai entropies to the same asymptotic value likely constitutes the most compelling numerical evidence of ergodicity attainable in simulations.

V. DISCUSSION

The root chosen in the original FPUT experiment and most subsequent investigations of the FPUT paradox was the fundamental mode. In essence, and as previously shown by comparisons with the nearby and integrable Toda chain [24], the proximity of the fundamental mode to a stationary state of the FPUT chain ends up trapping it in a near-regular portion of phase space for a finite time. We have shown here that higher-excited roots are trapped for even longer than the fundamental mode in such near-regular regions, even though all other parameters in the system are held constant.

Traditionally, the Lyapunov time is considered to characterize alignment along the maximally chaotic direction in a decomposition of the tangent space known as Oseledec splitting, which is covariant with respect to the Hamiltonian flow [33,45]. Evidently, such decomposition is meaningless in the case of a near-regular trajectory, and it is expected that the transient regime of $\lambda_1(t)$ will be different depending on how close the initial state is to a near-integrable region. However, the Lyapunov time itself is an invariant quantity in ergodic systems, such that at the end all initial states must provide the same value for the converged MLE [that is, the asymptotic limit of $\lambda_1(t)$]. This was verified in Figs. 6 and 7. Thus, the time needed to escape a region of near-integrability is more related to the equipartition time than to Lyapunov time, since we expect mode space to be fully covered only once a trajectory starting in a root becomes sufficiently chaotic. The Lyapunov time itself then carries absolutely no information on
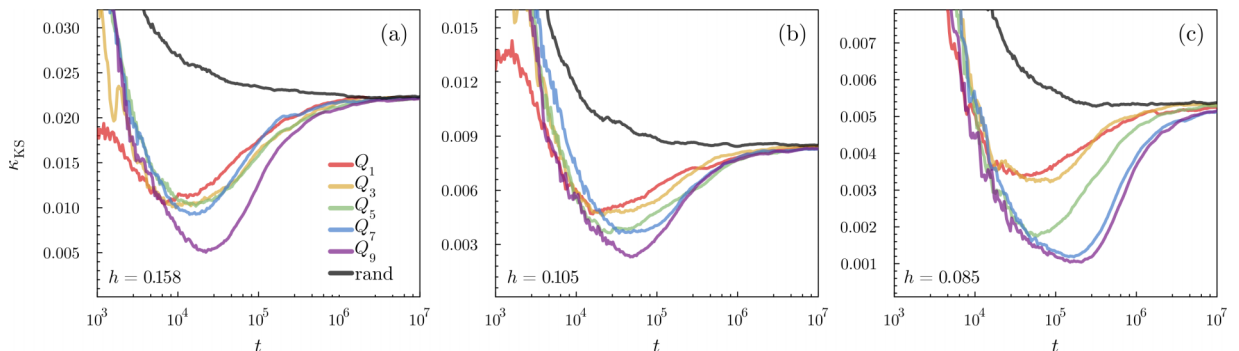

FIG. 8. Time-dependent Kolmogorov-Sinai entropy in an α-FPUT chain with $N = 63$ and $\alpha = 1/4$ computed from roots (rainbow colors) and from a random state (black). Each panel is computed for initial states with decreasing energy densities, h, shown in the left corner.
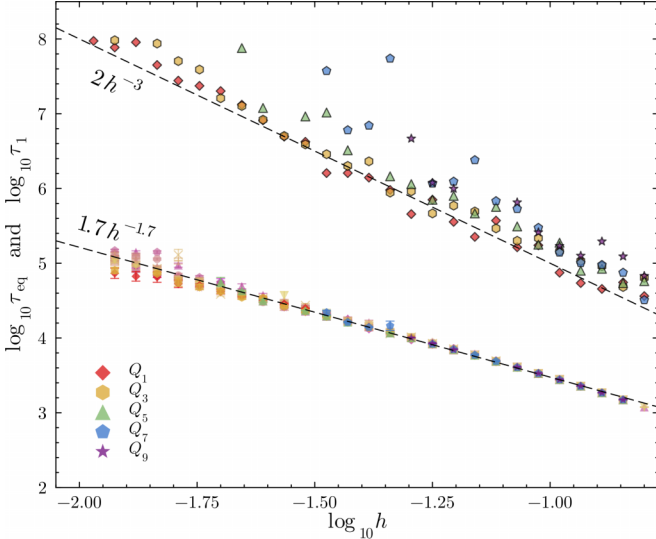
FIG. 9. Comparison of Lyapunov times in Fig. 7, where here both panels are superposed, and its corresponding linear fit. The equipartition times for root modes in Fig. 5 are also displayed, together with a linear fit for those of Q_1, showing that they are orders of magnitude longer than the Lyapunov times for all energy densities.

prethermalization, although its time-evolution certainly does. This leads to a reinterpretation of τ_1 as the time a *typical* initial condition takes to align with the maximally chaotic covariant direction in tangent space, since atypical ones might not align with it at all. This results in the Lyapunov time being always shorter than the equipartition time of a root, as seen by plotting them together in Fig. 9.

The only timescale longer than the equipartition time of roots appears to be the convergence time of their time-dependent Kolmogorov-Sinai entropies, $\kappa_{KS}(t)$. Like $\lambda_1(t)$, this quantity will only converge once the trajectory leaves a near-regular region, but now alignment with respect to a large fraction of covariant Lyapunov directions is required instead of only the maximal one. Since several directions are associated to small Lyapunov exponents it is not necessary to align with *all* of them to have a well-converged result, but Fig. 8(c) shows that the convergence times are, nevertheless, longer than the equipartition times shown in Fig. 2. Thus, the inescapable conclusion is that equipartition is reached before the trajectory has fully explored the chaotic sea.

It would be interesting to attempt a computation of the Lyapunov time of a submanifold, i.e., calculate τ_1 starting from a nonthermal mode, e.g., Q_2, and see how this exponent relates to the true τ_1 obtained from a thermal initial condition. Unfortunately this type of numerical investigation is hard, if not impossible, to carry on. This is due to the fact that the bushes associated to nonthermal modes in α-FPUT chains are unstable and will not be preserved in long-time evolution no matter which numerical integrator is chosen [27]. Indeed, if the chosen root is nonthermal, then the time evolution of the mode energies reveals that the initial excitation eventually escapes the bush and spreads throughout mode space, as numerical integration does not preserve the discrete symmetries of bushes. This might be behind the observation that the evolution of nonthermal modes results in approximately the same
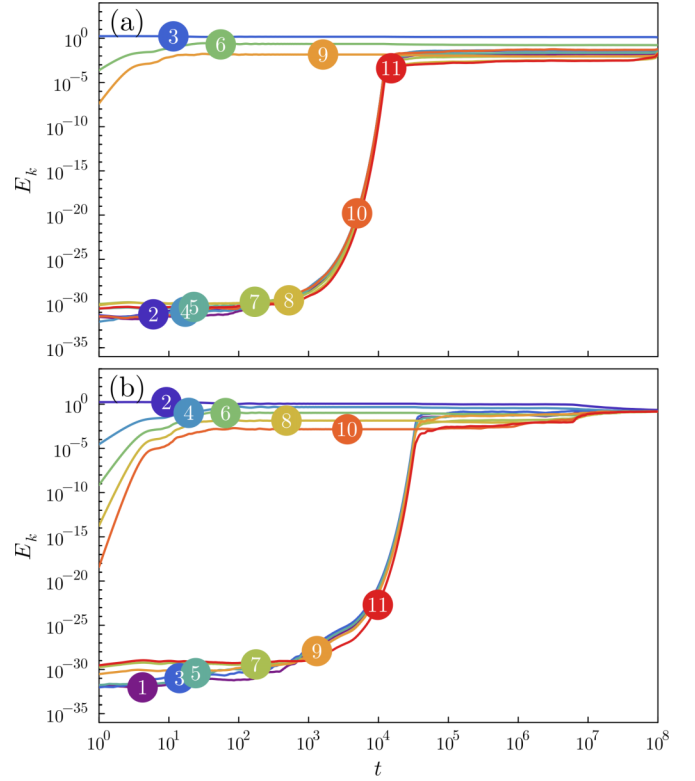

FIG. 10. Examples of error-induced equipartition due to numerical instability of bushes in a chain with $N = 11$, $h \approx 0.16$, and $\alpha = 1/4$. (a) Same as Fig. 1(a) but for longer times. (b) The evolution of Q_2, which as all even roots is nonthermal. In both panels, states that are proven to be nonthermal end up reaching equipartition due to numerical errors. Panels (a) and (b) use different symplectic integrators (second and fourth orders), showing that the phenomenon happens independently of the solver, although it happens at different times depending on the time step (attesting for its numerical source).

Lyapunov times as thermal ones [24], which is surprising given that bush dynamics takes place only in a submanifold of mode space. Nevertheless, it is unclear how ergodicity in mode space relates to that of phase space, such that it is possible that a partial covering of mode space by a large enough bush is enough to resolve the "true" MLE. Besides, it might even happen that dynamics in phase space is ergodic while that in mode space is not, since ergodicity always depends on the observables chosen [46] unless it is tracked by means of invariant quantities [47]. At present, the authors are unaware of a special type of integrator that is capable of preserving bushes of excitations, such that studies regarding dynamical properties of chaotic subspaces cannot be performed in a numerically meaningful fashion. A consequence of this is that the energy in Fig. 1(a), which is correctly localized in the bush associated to Q_3, will eventually spread to all modes for long enough times. Nevertheless, the way "equipartition" takes place in this case is pathological and clearly traceable to numerical errors, as can be seen in the Appendix.

VI. CONCLUSION

We have extended previous studies on the cubic Fermi-Pasta-Ulam-Tsingou (α-FPUT) model by investigating the

consequences of evolving higher-excited normal modes instead of the fundamental. The evolution of such atypical initial conditions was also compared to that obtained from typical (random) ones. We provided an explicit and simple way of ordering excited modes perturbatively in α when starting from a single normal mode (the root), and showed that higher-excited roots take longer to cover mode space. Thorough numerical investigations based on computing Lyapunov exponents and Kolmogorov-Sinai entropies showed that even invariant quantities carry imprints of the atypical initial states used to compute them in their transient regime, despite converging to the correct invariant value for long times. The prethermal dynamics taking place before convergence is confirmed to be a consequence of some select roots lying close to stationary states also in the α-FPUT system, and therefore being trapped in near-regular regions. The escape from such regions is what puts an end to the paradigmatic metastable plateaus in mode energies, whose end marks the moment invariant quantities computed from roots start to converge.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX: NUMERICAL INSTABILITY OF BUSHES

The bushes associated with nonthermal roots in α-FPUT chains obey discrete symmetries that are not respected by the integrators, being therefore unstable in numerical simulations. The main consequence of this instability is that energy, which should only be shared among modes in the bush and never leave for any finite time, ends up exciting other modes due to numerical errors. In Fig. 10 we provide two examples of this behavior when evolving provenly nonthermal modes: the roots Q_3 and Q_2 in a $N = 11$ chain. The first case is essentially a continuation of Fig. 1 for much longer times, until one can see an anomalous and simultaneous jump in the energy of modes that should not be excited starting from Q_3; the root Q_2, however, is even and therefore always nonthermal, undergoing a very similar pathological jump and reaching equipartition in Fig. 10 exclusively due to numerical errors.

[1] E. Fermi, J. Pasta, S. Ulam, and M. Tsingou, Studies of nonlinear problems. I, Technical Report LA-1940 (Los Alamos Scientific Laboratory, Los Alamos, NM, 1955).

[2] J. Ford, The Fermi-Pasta-Ulam problem: Paradox turns discovery, Phys. Rep. **213**, 271 (1992).

[3] T. P. Weissert, *The Genesis of Simulation in Dynamics: Pursuing the Fermi-Pasta-Ulam Problem* (Springer-Verlag, New York, NY, 1997).

[4] G. Gallavotti, *The Fermi-Pasta-Ulam Problem: A Status Report*, Vol. 728 (Springer Science & Business Media, Cham, 2007).

[5] A. A. N. Kolmogorov, Preservation of conditionally periodic movements with small change in the Hamilton function, in *Stochastic Behavior in Classical and Quantum Hamiltonian Systems*, edited by G. Casati and J. Ford, Lecture Notes in Physics, Vol. 93 (Springer, Berlin, Heidelberg, 1979).

[6] V. I. Arnold, Small denominators. I. Mapping of the circumference onto itself, *Collected Works: Representations of Functions, Celestial Mechanics and KAM Theory, 1957–1965 Vladimir I. Arnold—Collected Works* (Springer-Verlag Berlin, Heidelberg, 2009), pp. 152–223.

[7] J. Moser, On invariant curves of area-preserving mappings of an annulus, Nachr. Akad. Wiss. Göttingen Math. Phys. Kl. **2**, 1 (1962).

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

[9] G. D. Birkhoff, On the periodic motions of dynamical systems, Acta Math. **50**, 359 (1927).

[10] F. G. Gustavson, On constructing formal integrals of a Hamiltonian system near an equilibrium point, Astron. J. **71**, 670 (1966).

[11] A. M. O. de Almeida, *Hamiltonian Systems: Chaos and Quantization*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, UK, 1990).

[12] J. A. Murdock, *Normal Forms and Unfoldings for Local Dynamical Systems*, Applied Mathematical Sciences (Springer, New York, NY, 2003), Vol. 182.

[13] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, 2nd ed., Texts in Applied Mathematics, Vol. 2 (Springer, New York, NY, 2003).

[14] S. D. Pace and D. K. Campbell, Behavior and breakdown of higher-order Fermi-Pasta-Ulam-Tsingou recurrences, Chaos: Interdisc. J. Nonlin. Sci. **29**, 023132 (2019).

[15] P. Poggi and S. Ruffo, Exact solutions in the FPU oscillator chain, Physica D **103**, 251 (1997).

[16] G. Berman and F. Izrailev, The Fermi–Pasta–Ulam problem: Fifty years of progress, Chaos **15**, 015104 (2005).

[17] D. K. Campbell, P. Rosenau, and G. M. Zaslavsky, Introduction: The Fermi–Pasta–Ulam problem—The first fifty years, Chaos **15**, 015101 (2005).

[18] G. Benettin, A. Carati, L. Galgani, and A. Giorgilli, The Fermi-Pasta-Ulam problem and the metastability perspective, in *The Fermi-Pasta-Ulam Problem*, edited by G. Gallavotti, Lecture Notes in Physics, Vol. 728 (Springer, Berlin, Heidelberg, 2007), pp. 151–189.

[19] F. M. Izrailev and B. V. Chirikov, Statistical properties of a nonlinear string, Dokl. Akad. Nauk SSSR **166**, 57 (1966) [Sov. Phys. Dokl. **11**, 30 (1966)].

[20] T. Nishida, A note on an existence of conditionally periodic oscillation in a one-dimensional anharmonic lattice, Mem. Fac. Eng. Univ. Kyoto **33**, 27 (1971).

[21] B. Rink and F. Verhulst, Near-integrability of periodic FPU-chains, Physica A **285**, 467 (2000).

[22] B. Rink, Symmetry and resonance in periodic FPU chains, Commun. Math. Phys. **218**, 665 (2001).

[23] F. Verhulst, Variations on the Fermi-Pasta-Ulam chain, a survey, in *Proceedings of the Chaotic Modeling and Simulation International Conference* (Springer, Berlin, 2020), pp. 1025–1042.

[24] L. Casetti, M. Cerruti-Sola, M. Pettini, and E. G. D. Cohen, The Fermi-Pasta-Ulam problem revisited: Stochasticity thresholds in nonlinear Hamiltonian systems, Phys. Rev. E **55**, 6566 (1997).

[25] S. Flach, M. V. Ivanchenko, and O. I. Kanakov, Q-breathers and the Fermi-Pasta-Ulam problem, Phys. Rev. Lett. **95**, 064102 (2005).

[26] S. Flach, M. V. Ivanchenko, and O. I. Kanakov, Q-breathers in Fermi-Pasta-Ulam chains: Existence, localization, and stability, Phys. Rev. E **73**, 036618 (2006).

[27] G. Chechin, N. Novikova, and A. Abramenko, Bushes of vibrational modes for Fermi-Pasta-Ulam chains, Physica D **166**, 208 (2002).

[28] R. Eichhorn, S. J. Linz, and P. Hänggi, Transformation invariance of Lyapunov exponents, Chaos, Solitons Fract. **12**, 1377 (2001).

[29] S. Flach and A. Ponno, The Fermi-Pasta-Ulam problem: Periodic orbits, normal forms and resonance overlap criteria, Physica D **237**, 908 (2008).

[30] T. Cretegny, T. Dauxois, S. Ruffo, and A. Torcini, Localization and equipartition of energy in the β-FPU chain: Chaotic breathers, Physica D **121**, 109 (1998).

[31] C. Danieli, D. K. Campbell, and S. Flach, Intermittent many-body dynamics at equilibrium, Phys. Rev. E **95**, 060202(R) (2017).

[32] C. G. Goedde, A. J. Lichtenberg, and M. A. Lieberman, Chaos and the approach to equilibrium in a discrete sine-Gordon equation, Physica D **59**, 200 (1992).

[33] V. I. Oseledec, A multiplicative ergodic theorem. Lyapunov characteristic number for dynamical systems, Trans. Moscow Math. Soc. **19**, 197 (1968).

[34] Y. B. Pesin, Characteristic Lyapunov exponents and smooth ergodic theory, Russian Math. Surveys **32**, 55 (1977).

[35] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: Theory, Meccanica **15**, 9 (1980).

[36] K. Geist, U. Parlitz, and W. Lauterborn, Comparison of different methods for computing Lyapunov exponents, Prog. Theor. Phys. **83**, 875 (1990).

[37] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin/Cummings Publishing Co., Inc., Advanced Book Program, Reading, MA, 1978), Vol. 364.

[38] J. J. Palis and W. De Melo, *Geometric Theory of Dynamical Systems: An Introduction* (Springer Science & Business Media, Cham, 2012).

[39] R. Bivins, N. Metropolis, and J. R. Pasta, Nonlinear coupled oscillators: Modal equation approach, J. Comput. Phys. **12**, 65 (1973).

[40] D. Sholl, Modal coupling in one-dimensional anharmonic lattices, Phys. Lett. A **149**, 253 (1990).

[41] D. S. Sholl and B. I. Henry, Recurrence times in cubic and quartic Fermi-Pasta-Ulam chains: A shifted-frequency perturbation treatment, Phys. Rev. A **44**, 6364 (1991).

[42] R. I. McLachlan and P. Atela, The accuracy of symplectic integrators, Nonlinearity **5**, 541 (1992).

[43] H. J. Matsuyama and T. Konishi, Multistage slow relaxation in a Hamiltonian system: The Fermi-Pasta-Ulam model, Phys. Rev. E **92**, 022917 (2015).

[44] J.-P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. **57**, 617 (1985).

[45] F. Ginelli, P. Poggi, A. Turchi, H. Chaté, R. Livi, and A. Politi, Characterizing dynamics with covariant Lyapunov vectors, Phys. Rev. Lett. **99**, 130601 (2007).

[46] M. Baldovin, A. Vulpiani, and G. Gradenigo, Statistical mechanics of an integrable system, J. Stat. Phys. **183**, 41 (2021).

[47] G. M. Lando and S. Flach, Thermalization slowing down in multidimensional Josephson junction networks, Phys. Rev. E **108**, L062301 (2023).