Energy diffusion due to nonlinear perturbation on linear Hamiltonians

Gin-yih Tsaur and Jyhpyng Wang*
Institute of Atomic and Molecular Sciences, Academia Sinica, P.O. Box 23-166, Taipei, 10764 Taiwan
(Received 8 April 1996; revised manuscript received 29 July 1996)

In nonintegrable Hamiltonian systems, energy initially localized in a few degrees of freedom tends to disperse through nonlinear couplings. We analyze such processes in systems of many degrees of freedom. As a complement to the well-known Arnold diffusion, which describes energy diffusion by chaotic motion near separatrices, our analysis treats another universal case: coupled small oscillations near stable equilibrium points. Because we are concerned with the low-energy regime, where the nonlinearity of the unperturbed Hamiltonian is negligibly small, existing theories of Arnold diffusion cannot apply. Using probability theories we show that resonances of small detuning, which are ubiquitous in systems of many degrees of freedom, make energy diffusion possible. These resonances are the cause of energy equipartition in the low-energy limit. From our analysis, simple analytic equations that relate the energy, the degrees of freedom, the strength of nonlinear coupling, and the time scale for equipartition emerge naturally. These equations reproduce results from large-scale numerical simulations with remarkable accuracy. [S1063-651X(96)04311-5]

PACS number(s): 02.50.Fz, 02.50.Kd, 05.20.−y, 05.45.+b

I. INTRODUCTION

Since Poincaré showed that most nonlinear Hamiltonian systems are nonintegrable, i.e., they do not possess any constant of motion other than the total energy [1], it has become apparent that in treating the vast variety of nonlinear systems, studying the general characteristics can be more important than searching for particular solutions of the equations of motion. This is especially true for systems of many degrees of freedom, because in such systems the solutions are most likely too complicated to offer digestible information.

A powerful technique for extracting the general characteristics of complex dynamic systems is statistical analysis. Consider the general Hamiltonian of near-integrable systems $H(\mathbf{p},\mathbf{q})=H_0(\mathbf{p},\mathbf{q})+\epsilon H_1(\mathbf{p},\mathbf{q})$, where H_0 is integrable while H_1 not and ϵ is a parameter $\ll 1$. It is more convenient to describe the system in terms of the action and angle variables $(\mathbf{I},\boldsymbol{\theta})$ of H_0, so that H_0 depends only on the action variables $\mathbf{I}$, i.e., $H(\mathbf{I},\boldsymbol{\theta})=H_0(\mathbf{I})+\epsilon H_1(\mathbf{I},\boldsymbol{\theta})$. In analyzing the statistical behavior of a nonlinear system, one is concerned more with the action variables than the angle variables. This is because the angle variables cycle rapidly between 0 and 2π in time scales too small for much physics to happen. In contrast, each action variable, which is a one-to-one mapping of the energy of a degree of freedom, changes slowly under the influence of ϵH_1. The distribution of the action variables and its evolution in time, referred to as the energy redistribution process, is the major concern of this paper.

Before Kolmogorov, Arnold, and Moser proved the existence of invariant trajectories in near-integrable systems and that the invariant trajectories may constitute a finite measure of the phase space [2–5], it was generally assumed that energy redistribution through nonlinear couplings would eventually bring a nonintegrable system to states of approximate equipartition. A well-known numerical experiment by Fermi, Pasta, and Ulam was intended to show such scenarios [6]. Among many important facts, the Kolmogorov-Arnold-Moser (KAM) theorem shows that energy redistribution does not always occur in nonintegrable systems. Redistribution is possible only when the theorem is not valid, namely, when the nonintegrable part of the Hamiltonian is sufficiently large or when the system is sufficiently close to resonances. By investigating the conditions under which the KAM theorem breaks down, one can obtain better insight into the energy redistribution process.

Resonances are known to enhance the energy transfer among coupled degrees of freedom. Parametric oscillation in nonlinear optics [7] and Fermi resonances in molecular dynamics [8] are well-known examples. Consider again the general near-integrable Hamiltonian $H(\mathbf{I},\boldsymbol{\theta})=H_0(\mathbf{I})+\epsilon H_1(\mathbf{I},\boldsymbol{\theta})$, where $\mathbf{I}=(I_1,I_2,\ldots,I_N)$ and $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_N)$ are the action and angle variables, respectively, and N is the number of degrees of freedom. In general, H_1 can be written in terms of its Fourier components as $H_1=\Sigma_{\mathbf{m}}V_{\mathbf{m}}(\mathbf{I})\cos(\mathbf{m}\cdot\boldsymbol{\theta})$, where $\mathbf{m}=(m_1,m_2,\ldots,m_N)$ is an array of integers. A resonance exists if the angular frequencies $\boldsymbol{\omega}=(\omega_1,\omega_2,\ldots,\omega_N)$, where $\omega_i\equiv\partial H_0/\partial I_i$, satisfies $\mathbf{m}\cdot\boldsymbol{\omega}\equiv\sum_{i=1}^{N}m_i\omega_i=0$ for some $\mathbf{m}$ in the expansion of H_1. The quantity $\Delta\omega_{\mathbf{m}}\equiv\mathbf{m}\cdot\boldsymbol{\omega}$ is called the resonance detuning. If H_0 is not linear, then $\boldsymbol{\omega}$ depends on $\mathbf{I}$. Resonance conditions are satisfied for the values of $\mathbf{I}$ on the intersections of the resonance surfaces $\mathbf{m}\cdot\boldsymbol{\omega}(\mathbf{I})=0$ and the energy surface $H(\mathbf{I},\boldsymbol{\theta})=$ const. Such resonances form a weblike structure known as the Arnold web. Although the Arnold web may constitute only a small part of the phase space, it is known that for $N\geqslant 3$ initially localized energy can diffuse along the stochastic layers of the interconnected resonances. The phenomenon is known as the Arnold diffusion [9]. Under the moderate nonlinearity condition $\epsilon\ll\alpha\ll(1/\epsilon)$, where $\alpha\equiv(I/\omega)(d\omega/dI)$ represents the nonlinearity of H_0, Chirikov was able to estimate the rate of Arnold diffusion along the so-called guiding resonance by considering the effect of a dominant ''layer resonance'' and other small ''driving resonances'' [10]. One of the important

*Author to whom correspondence should be addressed.

conclusions of Ref. [10] is that the diffusion rate is proportional to $\exp(-c|\Delta\omega_{\mathbf{m}}|/\sqrt{\epsilon})$, meaning that when the perturbation is very small, the diffusion is effectively frozen. However, in a recent paper Chirikov and Vecheslavov proposed a much faster diffusion mechanism by considering the relation between high-order perturbation terms and resonance detunings [11].

In Arnold diffusion energy redistribution proceeds along the stochastic layers near the separatrices by chaotic motion. The process cannot occur in the low-energy limit because it takes a significant amount of energy to reach the separatrices. Moreover, in the low-energy limit the nonlinear part of H_0 becomes negligible. In other words, H_0 is effectively linear. In such cases Chirikov's method cannot be applied because the condition $\alpha \gg \epsilon$ is not satisfied. How energy redistribution occurs in the low-energy limit remains unclear.

Energy redistribution is intimately related to the equipartition principle in statistical mechanics. Even if one has not been bothered by the KAM theorem when thinking about the equipartition principle, there is still the problem of the time scale for equipartition. In statistical mechanics the concept of ergodicity is defined in the $t\rightarrow\infty$ limit. But in our practical world one cannot wait for $t\rightarrow\infty$. It is important to know how long it takes for a nonintegrable system to reach equipartition, if that will ever occur.

In the low-energy limit the effect of nonlinear perturbation is small, hence the only way to walk around the KAM theorem is by going through resonances. As mentioned above, in the low-energy limit H_0 is effectively linear and ω_i are constants. The chance of having exact resonances $(\mathbf{m}\cdot\boldsymbol{\omega}=0)$ is small except for those imposed by symmetry requirements. Although given any set of $\boldsymbol{\omega}$ one can always find an $\mathbf{m}$ to make $\mathbf{m}\cdot\boldsymbol{\omega}$ arbitrarily small, the components of such $\mathbf{m}$ are often so large that the corresponding perturbation $V_{\mathbf{m}}(\mathbf{I})\cos(\mathbf{m}\cdot\boldsymbol{\theta})$ does not exist or is negligibly small. Therefore one should focus the attention on how far the system is away from resonances, instead of the existence of exact resonances. As an example, one of the authors (J. W.) has taken such an approach to show that the different behaviors of intramolecular vibrational energy redistribution in CF_2Cl_2 and SF_6 can be explained by their significant difference in the number of low-order (small-m_i) resonances of small detuning [12]. One may ask the following: Given a group of dominant low-order resonances, are the resonance detunings sufficiently small to cause energy redistribution? If yes, how fast is the process? Without a detailed knowledge of the dynamics of the system, it seems difficult to answer such questions. Yet when N is sufficiently large, due to the large number of frequency combinations, the statistics of the resonance detunings dictates the typical behavior of the system. Therefore, a natural approach is to incorporate the statistics of resonance detunings into the standard perturbation theory. As we shall see, such an approach unveils not only the dynamical origin of the equipartition principle, but also its validity boundaries and the time scale for reaching equipartition.

In this paper we study the energy redistribution process in the low-energy limit by analyzing the role of resonance detunings. Our model system is a linear H_0 of large N perturbed by a small nonlinear ϵH_1. Physically it represents small-amplitude harmonic oscillations near stable equilibrium points perturbed by weak nonlinear couplings. The system has many correspondences in the real world.

In Sec. II we show that if the lower bound of resonances detunings $|\Delta\omega|_{\min}$ is not small, energy redistribution can be bounded to a near neighborhood of the initial point. In contrast, if $|\Delta\omega|_{\min}$ is sufficiently small, which is likely the case when N is sufficiently large, we show that the energy redistribution resembles the diffusion process. The diffusion rate is proportional to ϵ^2 times the probability density of near-zero detuning. In Sec. III we discuss how the resonance detunings scale with N. The scaling relation reveals how improbable that energy redistribution in a system of large N is bounded. The result shows that resonances of small detuning omnipresent in systems of large N are the cause of equipartition in the low-energy limit. In Sec. IV we present our analysis of the time scale for equipartition and compare our estimation with results from recent large-scale numerical simulations. Without any fitting parameter, our analytical expressions reproduce results from two independent works. The paper is concluded in Sec. V with a discussion.

II. ROLE OF RESONANCE DETUNING IN ENERGY DIFFUSION

In this section we investigate the evolution of the action variables in an ensemble of the model systems. To facilitate the discussion and the comparison with established numerical experiments, we use the Fermi-Pasta-Ulam (FPU) β model and the ϕ^4 model as examples. Because the analysis is not model dependent, the results we obtained can be extended to other Hamiltonians by scaling the relevant parameters.

The Hamiltonians of the FPU β model and the ϕ^4 model are described in detail in Appendix A. Both models represent strings made of discrete mass and spring units. Each mass-spring unit forms a nonlinear oscillator that couples to its neighbors. The Hamiltonians have the form

$$H(\mathbf{I},\boldsymbol{\theta})=\sum_{i=1}^{N}\omega_i I_i+\epsilon\sum_{\mathbf{m}}V_{\mathbf{m}}(\mathbf{I})\cos(\mathbf{m}\cdot\boldsymbol{\theta}), \tag{2.1}$$

where ω_i are constants, $\{\mathbf{m}\}=\{(m_1,\ldots,m_N): m_i\in\mathbb{Z}, \Sigma_{i=1}^{N}|m_i|\leq 4\}$, and the number of the elements in $\{\mathbf{m}\}$ is $\mathcal{N}(\{\mathbf{m}\})=2^4N^4$ (see Appendix A). The equation of motion of the action variables is

$$\dot{I}_i=-\frac{\partial H}{\partial\theta_i}=\epsilon\sum_{\mathbf{m}}m_i V_{\mathbf{m}}(\mathbf{I})\sin(\mathbf{m}\cdot\boldsymbol{\theta}). \tag{2.2}$$

Substituting $\theta_i=\omega_i t+\theta_i^0+O(\epsilon)$ into Eq. (2.2), one has

$$\dot{I}_i=\epsilon\sum_{\mathbf{m}}m_i V_{\mathbf{m}}(\mathbf{I})\sin(\Delta\omega_{\mathbf{m}}t+\theta_{\mathbf{m}}^0)+O(\epsilon^2), \tag{2.3}$$

where $\Delta\omega_{\mathbf{m}}\equiv\Sigma_{i=1}^{N}m_i\omega_i$ represent the detunings of the resonances and $\theta_{\mathbf{m}}^0\equiv\Sigma_{i=1}^{N}m_i\theta_i^0$ represent the initial conditions of the angle variables. We calculate the change of I_i during a period ΔT by integrating Eq. (2.3) from 0 to ΔT. The length of ΔT is chosen to be sufficiently short so that $V_{\mathbf{m}}(\mathbf{I})$ do not change significantly during ΔT, yet much longer than the periods of the angle variables. For sufficiently small ϵ, $\mathbf{I}$

evolves much slower than θ, such choices of ΔT are always possible. Now that the change of $V_{\mathbf{m}}(\mathbf{I})$ is negligible during ΔT,

$$\begin{aligned}\Delta I_i &\approx \int_0^{\Delta T} \epsilon \sum_{\mathbf{m}} m_i V_{\mathbf{m}}(\mathbf{I}) \sin(\Delta\omega_{\mathbf{m}} t + \theta_{\mathbf{m}}^0)\, dt \\ &\approx -\epsilon \sum_{\mathbf{m}} m_i \frac{V_{\mathbf{m}}(\mathbf{I})}{\Delta\omega_{\mathbf{m}}} \{[\cos(\Delta\omega_{\mathbf{m}}\Delta T) - 1]\cos(\theta_{\mathbf{m}}^0) \\ &\quad - \sin(\Delta\omega_{\mathbf{m}}\Delta T)\sin(\theta_{\mathbf{m}}^0)\}. \end{aligned} \tag{2.4}$$

By averaging over the initial conditions θ_i^0, one obtains $\langle \Delta I_i^2 \rangle$, the average of ΔI_i^2 in a microcanonical ensemble,

$$\begin{aligned}\langle \Delta I_i^2 \rangle &= \frac{1}{2}\epsilon^2 \sum_{\mathbf{m}} {m_i}^2 \frac{V_{\mathbf{m}}^2(\mathbf{I})}{(\Delta\omega_{\mathbf{m}})^2} \\ &\quad \times \{[\cos(\Delta\omega_{\mathbf{m}}\Delta T) - 1]^2 + \sin^2(\Delta\omega_{\mathbf{m}}\Delta T)\} \\ &= \epsilon^2 \sum_{\mathbf{m}} {m_i}^2 V_{\mathbf{m}}^2(\mathbf{I}) \frac{1 - \cos(\Delta\omega_{\mathbf{m}}\Delta T)}{(\Delta\omega_{\mathbf{m}})^2}. \end{aligned} \tag{2.5}$$

In both the FPU β model and the ϕ^4 model each resonance detuning $\Delta\omega_{\mathbf{m}}$ is a linear sum of at most four ω_i because there are at most four nonzero integers in $\mathbf{m}$. We may divide $\{\mathbf{m}\}$ into two groups: (i) $\{\mathbf{m}'\} = \{\mathbf{m}$: with four nonzero $m_i\}$, $\mathcal{N}(\{\mathbf{m}'\}) = 2^4 \times 4! \times C_4^N$ and (ii) $\{\mathbf{m}''\} = \{\mathbf{m}$: with at most three nonzero $m_i\}$, $\mathcal{N}(\{\mathbf{m}''\}) = 2^4 \times N^4 - 2^4 \times 4! \times C_4^N$. Because we are interested in systems of large N, where $\mathcal{N}(\{\mathbf{m}'\})$ is much larger than $\mathcal{N}(\{\mathbf{m}''\})$, the major contribution to $\langle \Delta I_i^2 \rangle$ is from $\{\mathbf{m}'\}$, that is,

$$\langle \Delta I_i^2 \rangle \approx \epsilon^2 \sum_{\mathbf{m} \in \{\mathbf{m}'\}} {m_i}^2 V_{\mathbf{m}}^2(\mathbf{I}) f(\Delta\omega_{\mathbf{m}}, \Delta T), \tag{2.6}$$

where

$$f(\Delta\omega_{\mathbf{m}}, \Delta T) \equiv \frac{1 - \cos(\Delta\omega_{\mathbf{m}}\Delta T)}{(\Delta\omega_{\mathbf{m}})^2}. \tag{2.7}$$

For $\mathbf{m} \in \{\mathbf{m}'\}$, $\sum_{i=1}^N m_i^2 = 4$. One has

$$\langle |\Delta\mathbf{I}|^2 \rangle \equiv \sum_{i=1}^N \langle \Delta I_i^2 \rangle \approx 4\epsilon^2 \sum_{\mathbf{m} \in \{\mathbf{m}'\}} V_{\mathbf{m}}^2(\mathbf{I}) f(\Delta\omega_{\mathbf{m}}, \Delta T). \tag{2.8}$$

In systems of many degrees of freedom, the number of possible $\mathbf{m}$ is very large, hence there are many different $\Delta\omega_{\mathbf{m}}$. Because we are interested in the general statistical behavior of such systems, we may describe Eq. (2.8) in terms of the distributions of $V_{\mathbf{m}}$ and $\Delta\omega_{\mathbf{m}}$. Because $\Delta\omega_{\mathbf{m}}$ is determined by ω_i from the unperturbed Hamiltonian H_0, whereas $V_{\mathbf{m}}$ by the nonlinear perturbation H_1, it is reasonable to assume that in general the distributions of $V_{\mathbf{m}}$ and $\Delta\omega_{\mathbf{m}}$ are uncorrelated. Letting $P(\Delta\omega_{\mathbf{m}})$ be the distribution of $\Delta\omega_{\mathbf{m}}$, one has

$$\begin{aligned}\langle |\Delta\mathbf{I}|^2 \rangle &\approx 4\epsilon^2 \left(\sum_{\mathbf{m} \in \{\mathbf{m}'\}} V_{\mathbf{m}}^2(\mathbf{I}) \right) \\ &\quad \times \int_{|\Delta\omega|_{\min} \le |x| \le |\Delta\omega|_{\max}} f(x, \Delta T) P(x)\, dx, \end{aligned} \tag{2.9}$$

where $|\Delta\omega|_{\min}, |\Delta\omega|_{\max}$ are the bounds of $|\Delta\omega_{\mathbf{m}}|$.

In what follows we shall further reduce Eq. (2.9) into a form in which the role of $\Delta\omega_{\mathbf{m}}$ in the energy redistribution process is clearly displayed. By analyzing the integral in Eq. (2.9), we shall show that resonances with small detunings play the dominant roles in the energy redistribution process. Let P and Ω be the height and width of $P(x)$, respectively, and assume that $P(x)$ is well behaved such that

$$|P^{(k)}(x)| \le P\left(\frac{c_1}{\Omega}\right)^k \tag{2.10}$$

holds for some constant $c_1 \approx O(1)$, where $P^{(k)}(x)$ is the kth derivative of $P(x)$. As an example, $c_1 = 1/\sqrt{2}$ for the Gaussian distribution $(1/\sqrt{2\pi}\Omega)\exp(-x^2/2\Omega^2)$. To simplify the integral in Eq. (2.9) let us separate the range of x into two parts: $|\Delta\omega|_{\min} \le |x| \le \Omega/c_1$ and $\Omega/c_1 \le |x| \le |\Delta\omega|_{\max}$. For $|x| \le \Omega/c_1$, expanding $P(x)$ in series and noting that $f(x, \Delta T)$ is an even function of x, one obtains

$$\begin{aligned}&\int_{|\Delta\omega|_{\min} \le |x| \le |\Delta\omega|_{\max}} f(x,\Delta T) P(x)\, dx \\ &= \int_{|\Delta\omega|_{\min} \le |x| \le \Omega/c_1} f(x,\Delta T) \left[\sum_{k=0}^{\infty} \frac{P^{(k)}(0)}{k!} x^k \right] \\ &\quad + \int_{\Omega/c_1 \le |x| \le |\Delta\omega|_{\max}} f(x,\Delta T) P(x)\, dx \\ &= 2\int_{|\Delta\omega|_{\min}}^{\Omega/c_1} f(x,\Delta T) \left[P(0) + \sum_{k=1}^{\infty} \frac{P^{(2k)}(0)}{(2k)!} x^{2k} \right] dx \\ &\quad + \int_{\Omega/c_1 \le |x| \le |\Delta\omega|_{\max}} f(x,\Delta T) P(x)\, dx \\ &= 2\int_{|\Delta\omega|_{\min}}^{\infty} f(x,\Delta T) P(0)\, dx - 2\int_{\Omega/c_1}^{\infty} f(x,\Delta T) P(0)\, dx \\ &\quad + 2\int_{|\Delta\omega|_{\min}}^{\Omega/c_1} f(x,\Delta T) \left[\sum_{k=1}^{\infty} \frac{P^{(2k)}(0)}{(2k)!} x^{2k} \right] dx \\ &\quad + \int_{\Omega/c_1 \le |x| \le |\Delta\omega|_{\max}} f(x,\Delta T) P(x)\, dx. \end{aligned} \tag{2.11}$$

Letting $\mathcal{F}$ represent the first term and $\mathcal{G}$ the sum of the last three terms in Eq. (2.11), Eq. (2.9) becomes

$$\langle |\Delta\mathbf{I}|^2 \rangle \approx 4\left(\epsilon^2 \sum_{\mathbf{m} \in \{\mathbf{m}'\}} V_{\mathbf{m}}^2(\mathbf{I}) \right) (\mathcal{F} + \mathcal{G}). \tag{2.12}$$

Define

$$F(x) \equiv x \int_x^{\infty} \frac{1 - \cos u}{u^2}\, du. \tag{2.13}$$

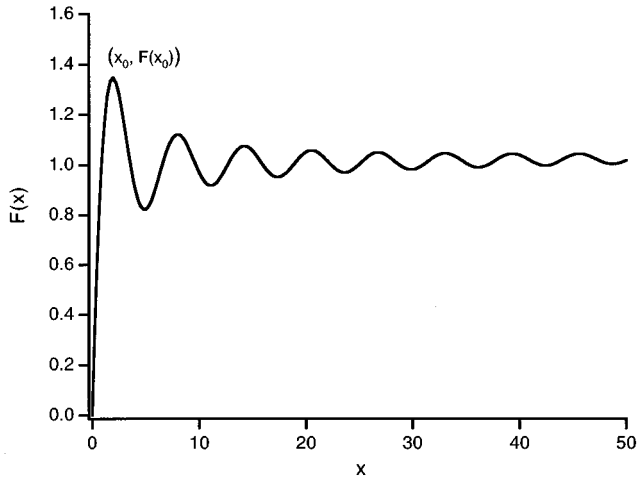

FIG. 1. Shape of $F(x)$.

Then

$$\mathcal{F}=2P(0)\Delta T\int_{|\Delta\omega|_{\min}\Delta T}^{\infty}\frac{1-\cos u}{u^2}du$$

$$=\frac{2P(0)}{|\Delta\omega|_{\min}}F(|\Delta\omega|_{\min}\Delta T). \tag{2.14}$$

In Fig. 1 we plot the function $F(x)$. It can be seen that $F(x)$ is bounded by its maximum $F(x_0)<1.4$ and for $x<x_0$ the function increases monotonically with x. When $x\ll 1$

$$F(x)\approx x\int_0^{\infty}\frac{1-\cos u}{u^2}du=\frac{\pi}{2}x. \tag{2.15}$$

Using Eq. (2.10), one can show that

$$|\mathcal{G}|\leqslant 12c_1P/\Omega. \tag{2.16}$$

The proof is given in Appendix B. These properties of $\mathcal{F}$ and $\mathcal{G}$ will be used in the following analysis.

If $|\Delta\omega|_{\min}$ is not small, the energy could be bounded to a near neighborhood of the initial point $\mathbf{I}$. The condition for energy localization is $\langle|\Delta\mathbf{I}|\rangle/|\mathbf{I}|\ll 1$ for all time, where $\langle|\Delta\mathbf{I}|\rangle$ is a simplified notation of $\sqrt{\langle|\Delta\mathbf{I}|^2\rangle}$. From Eqs. (2.14) and (2.16)

$$|\mathcal{F}|+|\mathcal{G}|\leqslant\frac{2PF(x_0)}{|\Delta\omega|_{\min}}+\frac{12c_1P}{\Omega}, \tag{2.17}$$

where $F(x_0)<1.4$ and $1/\Omega<1/|\Delta\omega|_{\min}$. From Eqs. (2.12) and (2.17) one can see there exists a constant c_2 of order 10 such that $\langle|\Delta\mathbf{I}|\rangle/|\mathbf{I}|$ is bounded by

$$B=\frac{\sqrt{c_2\left(\epsilon^2\sum_{\mathbf{m}\in\{\mathbf{m}'\}}V_{\mathbf{m}}^2(\mathbf{I})\right)P/|\Delta\omega|_{\min}}}{|\mathbf{I}|}. \tag{2.18}$$

Although Eq. (2.12) holds only for a period ΔT during which the change of $V_{\mathbf{m}}(\mathbf{I})$ is negligible, if $|\Delta\omega|_{\min}$ is sufficiently large or the perturbation $\epsilon^2\Sigma V_{\mathbf{m}}^2$ is sufficiently small, such that $B\ll 1$, then the average fractional change of $V_{\mathbf{m}}(\mathbf{I})$, which is proportional to $\langle|\Delta\mathbf{I}|\rangle/|\mathbf{I}|$, stays negligible for arbitrarily long ΔT. In this case the energy is localized, as expected from the KAM theorem.

In Sec. III we show that $|\Delta\omega|_{\min}$ decreases rapidly with increasing N. Hence, for large N, the energy will not be bounded in a small neighborhood of the initial point. If $|\Delta\omega|_{\min}\Delta T\ll 1$, from Eq. (2.15)

$$\mathcal{F}\approx\frac{2P(0)}{|\Delta\omega|_{\min}}\frac{\pi}{2}(|\Delta\omega|_{\min}\Delta T)=\pi P(0)\Delta T. \tag{2.19}$$

In Appendix B we show that $\pi P(0)\geqslant P$, whereas $|\mathcal{G}|\leqslant 12c_1P/\Omega$. If ϵ and $|\Delta\omega|_{\min}$ are both sufficiently small we may choose $\Delta T\gg 12c_1/\Omega$ without violating the conditions that the change of $V_{\mathbf{m}}(\mathbf{I})$ is negligible during ΔT and $|\Delta\omega|_{\min}\Delta T\ll 1$. Then $\mathcal{G}$ in Eq. (2.12) can be neglected and

$$\langle|\Delta\mathbf{I}|^2\rangle\approx 4\pi\left(\epsilon^2\sum_{\mathbf{m}\in\{\mathbf{m}'\}}V_{\mathbf{m}}^2(\mathbf{I})\right)P(0)\Delta T. \tag{2.20}$$

Equation (2.20) shows that when the number of degrees of freedom N is large, under suitable ''coarse graining'' (choice of ΔT), the energy redistribution resembles a diffusion (random walk) process. The $\mathbf{I}$-dependent ''diffusion rate''

$$D=4\pi\left(\epsilon^2\sum_{\mathbf{m}\in\{\mathbf{m}'\}}V_{\mathbf{m}}^2(\mathbf{I})\right)P(0) \tag{2.21}$$

is independent of $|\Delta\omega|_{\min}$. From the form of $V_{\mathbf{m}}(\mathbf{I})$ described in Eq. (A25) it can be seen that the diffusion rate is never zero except at the phase space boundary where $I_i=0$ for some i. From Eq. (2.3) it can be seen that at this boundary $\dot{I}_i=0$ the diffusion becomes limited in the directions perpendicular to the i axis in the phase space.

Equation (2.20) holds under the condition that during ΔT the change of $V_{\mathbf{m}}(\mathbf{I})$ is negligible or, equivalently, $\langle|\Delta\mathbf{I}|\rangle/|\mathbf{I}|\ll 1$. This condition can be used to estimate the upper bound of $|\epsilon H_1|/H_0$ for Eq. (2.20) to hold. For simplicity let us require that $\langle|\Delta\mathbf{I}|\rangle/|\mathbf{I}|\leqslant c_3$, where c_3 is of the order 0.1. Because $\Delta T\gg 12c_1/\Omega$ and $\langle|\Delta\mathbf{I}|^2\rangle=D\Delta T$, where D is the diffusion rate in Eq. (2.21),

$$\sqrt{\frac{12c_1}{\Omega}}\ll\sqrt{\Delta T}\leqslant\frac{c_3|\mathbf{I}|}{\langle|\Delta\mathbf{I}|\rangle}\sqrt{\Delta T}\approx\frac{c_3|\mathbf{I}|}{\sqrt{D}}. \tag{2.22}$$

To carry out the estimation, we replace $|\mathbf{I}|$ by its average $\overline{|\mathbf{I}|}$, which is approximately $\sqrt{NI_{\text{av}}^2}$, where I_{av} is the average action per degree of freedom, and relate $\overline{|\mathbf{I}|}$ to H_0 and D to H_1. Because H_0 is approximately $N\omega_{\text{av}}I_{\text{av}}$, where ω_{av} is the average value of ω_i,

$$\overline{|\mathbf{I}|}\approx\frac{H_0}{\sqrt{N}\omega_{\text{av}}}. \tag{2.23}$$

Because $\epsilon^2\Sigma V_{\mathbf{m}}^2$ is of the order $2\epsilon^2H_1^2$,

$$D\approx 8\pi\epsilon^2H_1^2P(0). \tag{2.24}$$

With $c_1 \approx O(1)$, $c_3 \approx O(0.1)$, $\pi P(0) \geqslant P$, and the probability normalization condition $P\Omega \approx 1$, Eqs. (2.22)–(2.24) yield

$$\frac{|\epsilon H_1|}{H_0} \ll \frac{0.01\Omega}{\omega_{\rm av}\sqrt{N}}. \tag{2.25}$$

This condition defines how weak the perturbation strength $|\epsilon H_1|/H_0$ has to be for Eq. (2.20) to be valid.

III. SCALING RELATIONS BETWEEN $|\Delta\omega|_{\rm min}$ AND N

It was shown in the preceding section that the nature of the energy redistribution process in systems of large N is determined by the lower bound of resonance detuning $|\Delta\omega|_{\rm min}$. As $|\Delta\omega|_{\rm min}$ decreases, energy redistribution changes from being bounded to diffusion. In systems of large N, $|\Delta\omega|_{\rm min}$ is likely to be small because there are many combinations of $\Delta\omega_{\mathbf{m}}$. Therefore diffusion is likely to dominate the energy redistribution process. However, because the diffusion rate is independent of $|\Delta\omega|_{\rm min}$, once $|\Delta\omega|_{\rm min}$ is sufficiently small for diffusion to occur, it does not make much difference to have an even larger N. Therefore, for a given perturbation ϵH_1, in a rough sense there exists a threshold number of degrees of freedom $N_{\rm th}$ such that the energy redistribution changes its nature from being bounded to diffusion as N crosses $N_{\rm th}$. To clarify the role of N we shall now discuss the relationship between $|\Delta\omega|_{\rm min}$ and N by calculating the expectation value of $|\Delta\omega|_{\rm min}$.

For any particular $\mathbf{m} \in \{\mathbf{m}'\}$ the probability of $|\Delta\omega_{\mathbf{m}}| > y$ is

$$1 - \int_{|x| \leqslant y} P(x)dx. \tag{3.1}$$

Because $\mathcal{N}(\{\mathbf{m}'\}) = 2^4 \times 4! \times C_4^N$ and every 4! of the $\Delta\omega_{\mathbf{m}}$ (hence every $2 \times 4!$ of the $|\Delta\omega_{\mathbf{m}}|$) has only one value, the probability of $|\Delta\omega|_{\rm min} \geqslant y$, defined as $G(y)$, is

$$G(y) = \left[1 - \int_{|x| \leqslant y} P(x)dx\right]^{2^3 C_4^N}. \tag{3.2}$$

Therefore the probability density for $|\Delta\omega|_{\rm min} = y$ is $-G'(y)$ and the expectation value of $|\Delta\omega|_{\rm min}$ is

$$\int_0^\infty y[-G'(y)]dy = -yG(y)|_0^\infty + \int_0^\infty G(y)dy = \int_0^\infty G(y)dy. \tag{3.3}$$

Let c be sufficiently small such that when $y \leqslant c$,

$$\begin{aligned} \ln G(y) &= 2^3 C_4^N \ln\left[1 - \int_{|x| \leqslant y} P(x)dx\right] \\ &\approx 2^3 C_4^N \left[-\int_{|x| \leqslant y} P(x)dx\right] \\ &\approx 2^3 C_4^N \left[-\int_{|x| \leqslant y} P(0)dx\right] \\ &= -2^4 C_4^N P(0) y. \end{aligned} \tag{3.4}$$

From Eq. (3.4) one has

$$\begin{aligned} \int_0^\infty G(y)dy &\approx \int_0^c \exp[-2^4 C_4^N P(0) y]dy + \int_c^\infty G(y)dy \\ &= \frac{1}{2^4 C_4^N P(0)}\{1 - \exp[-2^4 C_4^N P(0) c]\} \\ &\quad + \int_c^\infty G(y)dy. \end{aligned} \tag{3.5}$$

For systems of large N, $\exp[-2^4 C_4^N P(0) c] \ll 1$ can be neglected. Define $Q(y) \equiv 1 - \int_{|x| \leqslant y} P(x)dx$. Then

$$\begin{aligned} \int_c^\infty G(y)dy &= \int_c^\infty [Q(y)]^{2^3 C_4^N} dy \\ &\leqslant [Q(c)]^{2^3 C_4^N - 1} \int_0^\infty Q(y)dy, \end{aligned} \tag{3.6}$$

where $\int_0^\infty Q(y)dy$ is the expectation value of $|\Delta\omega_{\mathbf{m}}|$, which is finite and independent of N. Because $Q(c) < 1$, $Q(c)^M \ll 1/M$ for sufficiently large M. Hence $\int_c^\infty G(y)dy \ll 1/[2^4 C_4^N P(0)]$ for sufficiently large N. For systems of large N the expectation value of $|\Delta\omega|_{\rm min}$ is then approximately

$$|\Delta\omega|_{\rm min} \approx \frac{1}{2^4 C_4^N P(0)}. \tag{3.7}$$

This relation shows how $|\Delta\omega|_{\rm min}$ scales with N. It is approximately equal to the inverse probability density $1/P(0)$ divided by the number of frequency combinations $2^4 C_4^N$.

Now that the dependence of $|\Delta\omega|_{\rm min}$ on N is known, we are in a good position to estimate $N_{\rm th}$. Substituting Eq. (3.7) into Eq. (2.18), the condition for energy localization is

$$B = \frac{\sqrt{c_2\left(\epsilon^2 \sum_{\mathbf{m} \in \{\mathbf{m}'\}} V_{\mathbf{m}}^2(\mathbf{I})\right)[2^4 C_4^N P P(0)]}}{|\mathbf{I}|} \ll 1. \tag{3.8}$$

Again, to carry out the estimation we replace $|\mathbf{I}|$ by $\overline{|\mathbf{I}|} \approx H_0/(\sqrt{N}\omega_{\rm av})$, and noting that $c_2 \approx O(10)$, $P(0) \geqslant P/\pi$, $P \approx 1/\Omega$, and $\epsilon^2 \Sigma V_{\mathbf{m}}^2 \approx 2\epsilon^2 H_1^2$, Eq. (3.8) yields

$$N^{5/2}\left(\frac{\omega_{\rm av}}{\Omega}\right)\left(\frac{|\epsilon H_1|}{H_0}\right) \ll 1. \tag{3.9}$$

For simplicity let us require that the left-hand side of Eq. (3.9) equals c_4, where c_4 is of the order 0.1. Then we have an estimation of $N_{\rm th}$,

$$N_{\rm th} = c_4^{2/5}\left(\frac{\Omega}{\omega_{\rm av}}\frac{H_0}{|\epsilon H_1|}\right)^{2/5} \approx 0.4\left(\frac{\Omega}{\omega_{\rm av}}\frac{H_0}{|\epsilon H_1|}\right)^{2/5}. \tag{3.10}$$

IV. TIME SCALE FOR EQUIPARTITION

The foundation of the equipartition principle in dynamics is an old problem. The emergence of the KAM theorem has made it more puzzling. In the high-energy limit where the motion is highly chaotic, it is relatively easy to appreciate the equipartition principle. In the intermediate energy range,

TABLE I. Comparison with numerical simulations from Ref. [14]. $\ln\tau_R$, numerical results, $\ln T_e$, calculated values.

	$N=128$						$N=256$			$N=512$	
$\ln\mathcal{E}$	-6.5	-5.9	-5.2	-4.6	-4.2	-3.8	-6.2	-5.5	-4.4	-5.6	-5.0
$\ln\tau_R$	14.3	13.0	12.3	10.9	10.5	9.6	14.2	12.3	10.6	12.4	11.5
$\ln T_e$	15.1	13.9	12.5	11.3	10.5	9.7	13.8	12.4	10.2	11.9	10.7

where the motion is partially chaotic, Arnold diffusion provides a possible mechanism for equipartition. In the low-energy limit where regular motion is expected, it is more difficult to relate the equipartition principle to dynamics. The problem actually lies in the fact that our intuition of dynamics often comes from systems of small N. In Secs. II and III we showed that if $N>N_{\rm th}$, energy diffusion is likely uninhibited. From the fact that in an isolated system the simplest steady-state solution of the diffusion equation is a constant, one expects that given sufficient time, equipartition will be reached. In other words, if the action distribution will ever reach a steady state, the state should be an equipartition state.

In experiment, Fermi, Pasta, and Ulam employed one of the earliest electronic digital computers to study the problem [6]. Although they did not find any evidence of energy equipartition, their results stimulated many subsequent works in theory as well as in numerical experiment [13]. In early numerical works the lengths of integration were limited by the speed of the computers; hence, if equipartition was not found, one could always suspect that the result may be different with a longer integration time. Regardless of the true reason behind the original FPU results, recent simulations by Pettini and Landolfi [14], Goedde *et al.* [15], and Kantz *et al.* [16] have revealed strong evidence of near equipartition. Using a CRAY-XMP supercomputer, Pettini and Landolfi studied energy redistribution in the FPU β model and the ϕ^4 model with wide ranges of parameters. They were able to observe near equipartition in all the cases they studied as long as the integration time is sufficiently long.

From the discussion above it is clear that the important question is actually how long one has to wait before the system reaches equipartition. Will that occur while we are still interested? Armed with Eq. (2.20), one can estimate the time scale for reaching equipartition. Consider the typical distance R between two arbitrary points in the action space. $R\approx\overline{|\mathbf{I}|}$, where $\overline{|\mathbf{I}|}$ is the average value of $|\mathbf{I}|$. To ensure that diffusion has filled up approximately the whole action space, let us define the time scale for equipartition T_e to be the time it takes for the diffusion to reach twice of the typical distance; then $T_e\approx(2\overline{|\mathbf{I}|})^2/D$, where D is the diffusion rate in Eq. (2.21). From Eqs. (2.23) and (2.24) one obtains

$$T_e\approx\frac{1}{2\pi N\omega_{\rm av}^2P(0)}\frac{H_0^2}{\epsilon^2H_1^2}. \tag{4.1}$$

Note that in Eq. (4.1) the N dependence is not as simple as it appears because the ratio $H_0/|\epsilon H_1|$ may also depend on N. For example, consider the potential energy of the FPU β Hamiltonian [see Eq. (A6)]

$$U=\sum_{i=1}^{N}\frac{1}{2}(\Delta q_i)^2+\frac{1}{4}\beta\sum_{i=1}^{N}(\Delta q_i)^4. \tag{4.2}$$

If the total energy of the system is E, then $\Delta q_{\rm av}^2$, the average value of Δq_i^2, is approximately E/N and $\Delta q_{\rm av}^4\approx E^2/N^2$. Hence

$$\frac{H_0}{|\epsilon H_1|}\approx\frac{E/N}{(\beta/4)(E^2/N^2)}=\frac{4N}{\beta E}. \tag{4.3}$$

If we keep E/N fixed, $T_e\propto 1/N$. But if we keep E fixed, $T_e\propto N$. This is exactly the scaling law Kantz *et al.* discovered in their simulation of the FPU β model with fixed E and β [16].

Not only the scaling law $T_e\propto N$, but Eq. (4.1) actually reproduces the time scale for equipartition observed in Ref. [16]. Before we proceed further with the comparison, let us evaluate $\omega_{\rm av}$ and $P(0)$ first. In Ref. [16] $\omega_i=2\sin(\pi i/N)$. The average value $\omega_{\rm av}$ is approximately 1.27 and $P(0)$ is approximately 0.15, as estimated in Appendix B. Substituting $\omega_{\rm av}$, $P(0)$, $\beta=0.1$, and $E=10$, the step width $0.04/\sqrt{E}$ as stated in Ref. [16], into Eq. (4.1), with the scaling relation Eq. (4.3) one has

$$\frac{t}{N}=\frac{T_e}{(0.04/\sqrt{10})N}\approx\frac{\{1/[2\pi N\omega_{\rm av}^2P(0)]\}[4N/(\beta E)]^2}{(0.04/\sqrt{10})N}\approx 832. \tag{4.4}$$

This agrees well with the time scale shown in Fig. 2 of Ref. [16].

Pettini and Landolfi have studied the dependence of T_e on the average energy per degree of freedom $\mathcal{E}=E/N$ (ϵ in their notation) for different N [14]. They obtained the time scale for equipartition (τ_R in their notation) from the decay curve of the ''spectral entropy'' η, which is a natural indicator of the degree of equipartition. The simulations were carried out for both the FPU and the ϕ^4 model with a wide range of $\mathcal{E}$. The same formula Eq. (4.1) reproduces all the data in Ref. [14] for which $\mathcal{E}$ satisfies the condition of weak perturbation Eq. (2.25). The comparison is shown in Table I. There is an intrinsic uncertainty in determining τ_R due to the slow decay of η near equipartition. Taking into account this intrinsic uncertainty, the agreement between τ_R and T_e is remarkable.

Goedde *et al.* have also investigated the time scale for equipartition for the ϕ^4 model with numerical simulation [15]. However, in their simulation $|\epsilon H_1|/H_0\approx 1$, which is too large for our analysis to be valid. Therefore we are not able to make a similar comparison with their numerical results.

It should be noted that the comparisons above are done without any fitting parameter. The remarkable accuracy of Eq. (4.1) strongly supports our thesis, namely, small resonance detunings in systems of large N is the basis of equipartition in the low-energy limit.

Confusion regarding the limit of $N\rightarrow\infty$ may arise. Does equipartition always happen in systems of large N? If yes, how should one look at the regular motion of a nonlinear string, such as the continuum limit of the FPU β model and the ϕ^4 model? To answer such questions, one must distinguish the continuum limit from the thermodynamic limit. In the continuum limit let $N\rightarrow\infty$ while keeping the energy E constant. In such cases $H_0/|\epsilon H_1|$ is of order N and $T_e\propto N$. Therefore, in practice, equipartition will never occur. This can be traced back to the diffusion rate D in Eq. (2.21). Because $D\propto\epsilon^2 H_1^2$, as E/N approaches zero, D also approaches zero. Even though diffusion is allowed because $|\Delta\omega|_{\min}\rightarrow 0$, the diffusion rate is too small to bear any practical significance. On the contrary, in the thermodynamic limit E/N is kept constant as $N\rightarrow\infty$. In such cases $H_0/\epsilon H_1$ remains roughly constant and $T_e\propto 1/N$. Equipartition can occur in practical time scales.

V. DISCUSSION

In Sec. II it is seen that in the low-energy limit the energy redistribution process shifts from bounded motion to diffusion as $|\Delta\omega|_{\min}$ approaches zero. Because $|\Delta\omega|_{\min}$ decreases rapidly with increasing N, as shown in Sec. III, we expect that in systems of large N energy diffusion always occurs, even in the low-energy limit where from casual inspection of the Hamiltonian the nonlinear couplings look unimportant. The conclusion above is not a surprise. Long ago Ford and Lunsford had given the following Hamiltonian as an example, in which the erratic dynamics is independent of the strength of perturbation γ. Only the time scale for observing the erratic dynamics changes with γ [17]:

$$H=J_1+2J_2+3J_3+\gamma[\alpha J_1 J_2^{1/2}\cos(2\theta_1-\theta_2)+\beta(J_1J_2J_3)^{1/2}\cos(\theta_1+\theta_2-\theta_3)]. \tag{5.1}$$

The example above is somewhat unnatural because the chance of having three frequencies of the exact ratio 1:2:3 is very small. Nevertheless, it carries some spirit of the basic idea presented in this paper: Small detunings make weak perturbation important. By replacing the unnatural requirement of exact resonance with a statistical distribution of detunings, the role of resonances in the energy redistribution becomes clear and relations among N, $|\epsilon H_1|/H_0$, $|\Delta\omega|_{\min}$, D, and T_e also become apparent.

We also noted in particular the importance of T_e when discussing equipartition. We showed that even though small $|\Delta\omega|_{\min}$ makes energy diffusion possible, it does not guarantee that equipartition will be reached in practical time scales. In particular, we used the FPU β model and the ϕ^4 model to show that low energy elastic waves in continuous media do not contribute to equipartition because energy diffusion caused by such motions is too slow to be of practical significance.

Our analysis can be readily generalized to other Hamiltonians. For example, if ϕ^6 instead of ϕ^4 was the dominant interaction, we would have $\Sigma_{i=1}^{N}|m_i|\leqslant 6$ in Eq. (2.1) and $|\Delta\omega|_{\min}$ becomes proportional to $1/C_6^N$. Consequently, $N_{\rm th}$ scales as $(H_0/|\epsilon H_1|)^{2/7}$, while the general form of D and T_e remains the same.

It should be noted that even though our analysis is meant to be used on nonintegrable Hamiltonians, questions may arise. For example, what would happen if the Hamiltonian only appears to be nonintegrable? There are classes of nonlinear Hamiltonians that are actually integrable. If we treat the nonlinear part of those Hamiltonians as perturbation, it is still possible to see ''apparent'' diffusion. Yet in such cases the motion of different degrees of freedom is correlated in such ways that with proper sets of variables the static nature of the motion can be revealed. Nevertheless, if one wishes to describe the motion in terms of linear mode expansions, the perturbation picture is still valid. It bears a close analogy to the interaction picture in quantum mechanics.

The theory presented in this paper can be readily extended to quantum mechanics. Let c_i be the probability amplitude of a system in state i. Then $|c_i|^2$ corresponds to the classical action variable I_i and the phase angle of c_i corresponds to the angle variable θ_i. The evolution of $|c_i|^2$ is governed by an equation similar to Eq. (2.4); hence using our method, one can analyze probability diffusion in quantum systems in a similar way. Such an analysis can be used to elucidate the dynamic basis of quantum microcanonical ensemble theories. We plan to present such an analysis in future papers.

ACKNOWLEDGMENT

We thank the National Science Council of Taiwan for financial support (Grant No. NSC85-2113-M-001-043).

APPENDIX A

The FPU β model is the discrete version of the nonlinear partial differential equation

$$\frac{\partial^2 y}{\partial t^2}-\frac{\partial^2 y}{\partial x^2}\left[1+3\beta\left(\frac{\partial y}{\partial x}\right)^2\right]=0. \tag{A1}$$

It is one of the nonlinear models Fermi, Pasta, and Ulam studied with an early digital computer. Discretizing the continuous function $y(x)$ in Eq. (A1) into a vector $(y_1,\ldots,y_N)$ and substituting in

$$3\left(\frac{\partial y}{\partial x}\right)^2=\frac{(y_{i+1}-y_i)^2}{(\Delta x)^2}+\frac{(y_i-y_{i-1})^2}{(\Delta x)^2}+\frac{(y_{i+1}-y_i)(y_i-y_{i-1})}{(\Delta x)^2}, \tag{A2}$$

$$\frac{\partial^2 y}{\partial x^2}=\frac{y_{i+1}-2y_i+y_{i-1}}{\Delta x^2}, \tag{A3}$$

where $\Delta x=L/N$ and L is the length of the string, Eq. (A1) becomes

$$\ddot{y}_i=\frac{y_{i+1}-2y_i+y_{i-1}}{(\Delta x)^2}\left\{1+\beta\left[\frac{(y_{i+1}-y_i)^2}{(\Delta x)^2}+\frac{(y_i-y_{i-1})^2}{(\Delta x)^2}+\frac{(y_{i+1}-y_i)(y_i-y_{i-1})}{(\Delta x)^2}\right]\right\}$$
$$=\frac{y_{i+1}-2y_i+y_{i-1}}{(\Delta x)^2}+\beta\left[\frac{(y_{i+1}-y_i)^3}{(\Delta x)^4}-\frac{(y_i-y_{i-1})^3}{(\Delta x)^4}\right]. \tag{A4}$$

Letting $q_i=y_i$ and $p_i=\dot{y}_i$, one has

$$\dot{q}_i=p_i,$$
$$\dot{p}_i=\frac{1}{(\Delta x)^2}(q_{i+1}-2q_i+q_{i-1})+\frac{\beta}{(\Delta x)^4}[(q_{i+1}-q_i)^3-(q_i-q_{i-1})^3]. \tag{A5}$$

From the above equations one obtains the Hamiltonian

$$H=\sum_{i=1}^{N}\left[\frac{1}{2}p_i^2+\frac{1}{2}\frac{1}{(\Delta x)^2}(q_{i+1}-q_i)^2+\frac{1}{4}\frac{\beta}{(\Delta x)^4}(q_{i+1}-q_i)^4\right] \tag{A6}$$
$$=\frac{1}{2}\sum_{i=1}^{N}p_i^2+\frac{1}{2}\sum_{i,j=1}^{N}A_{ij}q_iq_j+\frac{1}{4}\frac{\beta}{(\Delta x)^4}\sum_{i=1}^{N}(q_{i+1}-q_i)^4, \tag{A7}$$

where the periodic boundary condition $q_{n+1}=q_1$ is used, $A_{ij}=(2\delta_{i,j}-\delta_{i,j+1}-\delta_{i,j-1})/\Delta x^2$, and $\delta_{i,j}$ is the Kronecker delta. Setting $S_{ij}=1/\sqrt{N}[\cos(2\pi ij/N)+\sin(2\pi ij/N)]$, we can diagonalize the matrix $\mathcal{A}=(A_{ij})$ with $\mathcal{S}=(S_{ij})$ such that

$$\mathcal{S}\mathcal{A}\mathcal{S}^{-1}=\begin{pmatrix}\omega_1^2 & & \\ & \ddots & \\ & & \omega_N^2\end{pmatrix}, \tag{A8}$$

where all but the diagonal matrix elements are zero, $\omega_j^2=4\sin^2[(1/2)k_j]/\Delta x^2$, $k_j=2\pi j/N$, and $\mathcal{S}=\mathcal{S}^{-1}$. Making the canonical transform $\mathbf{P}=\mathcal{S}\mathbf{p}$ and $\mathbf{Q}=\mathcal{S}\mathbf{q}$,

$$H=\frac{1}{2}\sum_{j=1}^{N}P_j^2+\frac{1}{2}\sum_{j=1}^{N}\omega_j^2Q_j^2+\frac{1}{4}\frac{\beta}{(\Delta x)^4}\sum_{i=1}^{N}\left[\sum_{j=1}^{N}(S_{i+1,j}-S_{i,j})Q_j\right]^4. \tag{A9}$$

Change the variables $(\mathbf{Q},\mathbf{P})$ to the action angle variables $(\mathbf{I},\boldsymbol{\theta})$ with

$$\frac{1}{2}P_j^2+\frac{1}{2}\omega_j^2Q_j^2\equiv E_j, \tag{A10}$$

$$I_j\equiv\frac{1}{2\pi}\oint P_jdQ_j=\frac{2}{\pi}\int_0^{Q_{\max}}(2E_j-\omega_j{}^2Q_j^2)^{1/2}dQ_j=\frac{E_j}{\omega_j}, \tag{A11}$$

where $Q_{\max}=(2E_j)^{1/2}/\omega_j$ is the value of Q_j when $P_j=0$. The generating function $F(\mathbf{Q},\mathbf{I})$ can be obtained from

$$P_j=\frac{\partial}{\partial Q_j}F(\mathbf{Q},\mathbf{I}) \tag{A12}$$

to be

$$F(\mathbf{Q},\mathbf{I})=\sum_{j=1}^{N}\int_0^{Q_j}P_jdQ_j=\sum_{j=1}^{N}\int_0^{Q_j}(2E_j-\omega_j{}^2Q_j^2)^{1/2}dQ_j=\sum_{j=1}^{N}\int_0^{Q_j}(2\omega_jI_j-\omega_j{}^2Q_j^2)^{1/2}dQ_j. \tag{A13}$$

From the generating function one obtains the angle variables

$$\theta_j=\frac{\partial}{\partial I_j}F(\mathbf{Q},\mathbf{I})=\omega_j\int_0^{Q_j}(2\omega_jI_j-\omega_j{}^2Q_j^2)^{-1/2}dQ_j=\arcsin\left[\left(\frac{\omega_j}{2I_j}\right)^{1/2}Q_j\right]. \tag{A14}$$

Substituting

$$\frac{1}{2}P_j^2+\frac{1}{2}\omega_j^2Q_j^2=E_j=\omega_jI_j, \tag{A15}$$

$$Q_j=\left(\frac{2I_j}{\omega_j}\right)^{1/2}\sin\theta_j \tag{A16}$$

into the Hamiltonian, one obtains

$$H=\sum_{j=1}^{N}\omega_jI_j+\frac{\beta}{(\Delta x)^4}\sum_{i=1}^{N}\left[\sum_{j=1}^{N}(S_{i+1,j}-S_{i,j})\left(\frac{I_j}{\omega_j}\right)^{1/2}\sin\theta_j\right]^4. \tag{A17}$$

Essentially the same procedure can be applied to discretize the sine-Gordon equation in the low-energy limit for the ϕ^4 model. Substituting Eq. (A3) into

$$\frac{\partial^2y}{\partial t^2}-\frac{\partial^2y}{\partial x^2}+\gamma\sin y=0 \tag{A18}$$

and setting $q_i=y_i$ and $p_i=\dot{y_i}$, one obtains the Hamiltonian

$$H=\frac{1}{2}\sum_{i=1}^{N}p_i^2+\frac{1}{2}\sum_{i,j=1}^{N}A_{ij}q_iq_j+\gamma\sum_{i=1}^{N}(1-\cos q_i), \tag{A19}$$

with A_{ij} defined as before. In the low-energy limit q_i is very small, hence

$$H \approx \frac{1}{2}\sum_{i=1}^{N} p_i^2 + \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} q_i q_j + \frac{\gamma}{2}\sum_{i=1}^{N} q_i^2 - \frac{\gamma}{4!}\sum_{i=1}^{N} q_i^4$$
$$= \frac{1}{2}\sum_{i=1}^{N} p_i^2 + \frac{1}{2}\sum_{i,j=1}^{N} \widetilde{A}_{ij} q_i q_j - \frac{\gamma}{4!}\sum_{i=1}^{N} q_i^4, \tag{A20}$$

where $\widetilde{A}_{ij} = A_{ij} + \gamma\delta_{i,j}$. Similarly, we can diagonalize the matrix $\mathcal{A} = (\widetilde{A}_{ij})$ with $\mathcal{S}$ defined before:

$$\mathcal{S}\mathcal{A}\mathcal{S}^{-1} = \begin{pmatrix} \omega_1^2 & & \\ & \ddots & \\ & & \omega_N^2 \end{pmatrix}, \tag{A21}$$

where $\omega_j^2 = 4\sin^2[(1/2)k_j]/\Delta x^2 + \gamma$ and $k_j = 2\pi j/N$. Changing the variables to $\mathbf{P} = \mathcal{S}\mathbf{p}$ and $\mathbf{Q} = \mathcal{S}\mathbf{q}$, and then changing them to the action angle variables $(\mathbf{I}, \boldsymbol{\theta})$ as done before for the FPU β model, one has

$$H = \sum_{j=1}^{N} \omega_j I_j - \frac{\gamma}{3!}\sum_{i=1}^{N}\left[\sum_{j=1}^{N} S_{i,j}\left(\frac{I_j}{\omega_j}\right)^{1/2} \sin\theta_j\right]^4. \tag{A22}$$

Because

$$\sin\theta_j \sin\theta_k \sin\theta_l \sin\theta_m = \sum_{n_1,\cdots,n_4=\pm 1} \frac{(-1)^{n_1\cdots n_4}}{16} \times \cos(n_1\theta_j + n_2\theta_k + n_3\theta_l + n_4\theta_m), \tag{A23}$$

both the FPU β Hamiltonian and the ϕ^4 Hamiltonian can be written in the form

$$H = \sum_{j=1}^{N} \omega_j I_j + \epsilon \sum_{1\le j,k,l,m\le N} \sum_{n_1,\ldots,n_4=\pm 1} V^{jklm}_{n_1\cdots n_4}(\mathbf{I}) \times \cos(n_1\theta_j + n_2\theta_k + n_3\theta_l + n_4\theta_m), \tag{A24}$$

where we have introduced the dimensionless parameter $\epsilon = \beta/L^2$ for the FPU β model and $\epsilon = \gamma L^2/3!$ for the ϕ^4 model, and

$$V^{jklm}_{n_1\cdots n_4}(\mathbf{I}) = \frac{(-1)^{n_1\cdots n_4}}{16}\left(\sum_{i=1}^{N} C_j^i C_k^i C_l^i C_m^i\right)\sqrt{I_j}\sqrt{I_k}\sqrt{I_l}\sqrt{I_m}, \tag{A25}$$

with $C_j^i = N(S_{i+1,j} - S_{i,j})/\sqrt{\omega_j L}$ for the FPU β model and $C_j^i = S_{i,j}/\sqrt{\omega_j L}$ for the ϕ^4 model. The perturbation part $V^{jklm}_{n_1\cdots n_4}(\mathbf{I})$ contains $2^4\times N^4$ terms. With $\mathbf{m} \equiv (m_1, \ldots, m_N)$, $\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_N)$ and $(\mathbf{m}\cdot\boldsymbol{\theta}) \equiv \Sigma_{i=1}^{N} m_i\theta_i$ one has

$$H = \sum_{j=1}^{N} \omega_j I_j + \epsilon \sum_{\mathbf{m}} V_{\mathbf{m}}(I)\cos(\mathbf{m}\cdot\boldsymbol{\theta}). \tag{A26}$$

From Eqs. (A22)–(A24) it is apparent that $\{\mathbf{m}\}$ consists of all integer arrays $(m_1, \ldots, m_N)$ with $\Sigma_{i=1}^{N}|m_i| \le 4$.

APPENDIX B

Lemma 1. $|\mathcal{G}| \le 12c_1P/\Omega$.

Proof. Because $0 \le f(x,\Delta T) = [1-\cos(x\Delta T)]/x^2 \le 2/x^2$ and $0 \le P(x) \le P$, the second term in Eq. (2.11) is bounded by

$$\left|2\int_{\Omega/c_1}^{\infty} f(x,\Delta T)P(0)dx\right| \le 4P\int_{\Omega/c_1}^{\infty}\frac{1}{x^2}dx = 4\frac{c_1P}{\Omega}. \tag{B1}$$

Similarly, the last term in Eq. (2.11) is bounded by

$$\left|\int_{\Omega/c_1\le|x|\le|\Delta\omega|_{\max}} f(x,\Delta T)P(x)dx\right| \le 2P\int_{\Omega/c_1\le|x|\le\infty}\frac{1}{x^2}dx = 4P\int_{\Omega/c_1}^{\infty}\frac{1}{x^2}dx = 4\frac{c_1P}{\Omega}. \tag{B2}$$

From Eq. (2.10), $|P^{(2k)}(0)| \le P(c_1/\Omega)^{2k}$, the third term in Eq. (2.11), is bounded by

$$\left|2\int_{|\Delta\omega|_{\min}}^{\Omega/c_1} f(x,\Delta T)\left[\sum_{k=1}^{\infty}\frac{P^{(2k)}(0)}{(2k)!}x^{2k}\right]dx\right| \le 4\int_0^{\Omega/c_1}\frac{1}{x^2}\left[\sum_{k=1}^{\infty}\frac{P}{(2k)!}\left(\frac{c_1}{\Omega}\right)^{2k}x^{2k}\right]dx = 4\frac{c_1P}{\Omega}\sum_{k=1}^{\infty}\frac{1}{(2k)!(2k-1)} \le 4\frac{c_1P}{\Omega}. \tag{B3}$$

Thus the sum of the last three terms in Eq. (2.11), which is called $\mathcal{G}$, has the absolute value less than or equal to $12c_1P/\Omega$.

Lemma 2. $\pi P(0) \ge P$.

Proof. Defining $p(x)dx$ to be the probability of any particular ω_i falling in the interval $(x-\frac{1}{2}dx, x+\frac{1}{2}dx)$, i.e., $p(x)$ is the distribution of ω_i, then $P(\Delta\omega_{\mathbf{m}})$ can be derived from $p(x)$ by convolution. Because the convolution depends on the numbers of m_i that are $+1$ and -1, respectively, for convenience of discussion we separate $\{\mathbf{m}'\}$ further into five subsets: $\{\mathbf{m}'\} = \cup_{s=0}^{4}\{\mathbf{m}_s\}$, with $\{\mathbf{m}_s\} = \{\mathbf{m}: s \text{ of the } m_i \text{ are } +1,\ 4-s \text{ are } -1, \text{ and the rest are } 0\}$. Letting $P_s(\Delta\omega_{\mathbf{m}})$ be the distribution of $\Delta\omega_{\mathbf{m}}$ for $\mathbf{m}\in\{\mathbf{m}_s\}$ because $\mathcal{N}(\{\mathbf{m}_s\}) = C_s^4\times 4!\times C_4^N$ and $\mathcal{N}(\{\mathbf{m}'\}) = \Sigma_{s=0}{}^4\mathcal{N}(\{\mathbf{m}_s\}) = 2^4\times 4!\times C_4^N$,

$$P(\Delta\omega_{\mathbf{m}}) = \sum_{s=0}^{4}\frac{C_s^4}{2^4}P_s(\Delta\omega_{\mathbf{m}}). \tag{B4}$$

To avoid any possible misunderstanding of the meaning of $P_s(\Delta\omega_{\mathbf{m}})$, let us write out $P_2(\Delta\omega_{\mathbf{m}})$ explicitly. For $\mathbf{m}\in\{\mathbf{m}_2\}$, two m_i are $+1$, two are -1, and the rest are 0, i.e., $\Delta\omega_{\mathbf{m}} = \omega_1 + \omega_2 - \omega_3 - \omega_4$. The probability distribution of $\Delta\omega_{\mathbf{m}}$ in $\{\mathbf{m}_2\}$ is

$$P_2(\Delta\omega_{\mathbf{m}})=\int\int\int p(\omega_1+\omega_2-\omega_3-\Delta\omega_{\mathbf{m}}) \times p(\omega_1)p(\omega_2)p(\omega_3)d\omega_1 d\omega_2 d\omega_3. \quad \text{(B5)}$$

Other $P_s(\Delta\omega_{\mathbf{m}})$ can be derived by convolution of $p(x)$ in similar ways. Because the convolution involves four ω_i, from the central limit theorem in probability theory [18], the widths of all the $P_s(x)$ are approximately $\sqrt{4}w$, where w is the width of $p(x)$. Because the product of the height and the width of a probability distribution is approximately 1, the heights of all the $P_s(x)$ also have approximately the same value. As we shall see in the following paragraph, the value is just $P_2(0)$.

Defining

$$g(y)\equiv\int p(y+\omega)p(\omega)d\omega, \quad \text{(B6)}$$

then

$$g(-y)=\int p(-y+\omega)p(\omega)d\omega = \int p(\omega')p(y+\omega')d\omega' = g(y). \quad \text{(B7)}$$

Setting $y=\omega_3-\omega_2$, one has

$$\begin{aligned} P_2(x) &= \int\int\int p(\omega_1+\omega_2-\omega_3-x) \times p(\omega_1)p(\omega_2)p(\omega_3)d\omega_1 d\omega_2 d\omega_3 \\ &= \int\int\int p(\omega_1-y-x)p(\omega_1)p(\omega_2) \times p(y+\omega_2)d\omega_1 d\omega_2 dy \\ &= \int g(-y-x)g(y)dy = \int g(y+x)g(y)dy. \end{aligned} \quad \text{(B8)}$$

Hence

$$P_2(0)=\int g^2(y)dy \quad \text{(B9)}$$

and

$$P_2(x)\leqslant\left(\int g^2(y+x)dy\right)^{1/2}\left(\int g^2(y)dy\right)^{1/2} = \int g^2(y)dy = P_2(0). \quad \text{(B10)}$$

Equation (B10) shows that $P_2(0)$ is the maximum (the height) of $P_2(x)$; hence in a crude sense $P_s(x)\leqslant P_2(0)$ for all s. Because

$$P(x)=\sum_{s=0}^{4}\frac{C_s^4}{2^4}P_s(x)\leqslant\sum_{s=0}^{4}\frac{C_s^4}{2^4}P_2(0)=P_2(0), \quad \text{(B11)}$$

the height of $P(x)$ is smaller than $P_2(0)$, i.e., $P\leqslant P_2(0)$. Therefore

$$\pi P(0)=\pi\sum_{s=0}^{4}\frac{C_s^4}{2^4}P_s(0)\geqslant\frac{\pi C_2^4}{2^4}P_2(0)\geqslant P. \quad \text{(B12)}$$

Lemma 3. For $\omega_i=2\sin(\pi i/N)$, $P(0)\approx 0.15$.

Proof. The average value of ω_i is $\omega_{\mathrm{av}}\approx 1.27$ and the standard deviation of $p(x)$ (the distribution of ω_i) is $\sigma\approx 0.62$. Hence the width of $p(x)$ is $w\approx 2\times 0.62=1.24$. We know from Lemma 2 that for all s the width of $P_s(x)$ is approximately $\sqrt{4}w$ and the height of $P_s(x)$ is approximately $P_2(0)$. Because the product of the height and the width of a probability distribution is approximately 1, $P_2(0)\approx 1/\sqrt{4}w\approx 0.4$. Similar to the above proof that the maxima of $P_2(x)$ happens at $x=0$, one can show that the maximums of $P_0(x)$, $P_1(x)$, $P_3(x)$, and $P_4(x)$ happen at $x=-4\omega_{\mathrm{av}}$, $-2\omega_{\mathrm{av}}$, $2\omega_{\mathrm{av}}$, and $4\omega_{\mathrm{av}}$, respectively. Since the standard deviations of the $P_s(x)$ are all approximately $\sqrt{4}\sigma=1.24\approx\omega_{\mathrm{av}}$, $x=0$ is at least two standard deviations away from the peak for $s\neq 2$. Therefore $P_s(0)$ for $s\neq 2$ are very small. We have

$$P(0)=\sum_{s=0}^{4}\frac{C_s^4}{2^4}P_s(0)\approx\frac{C_2^4}{2^4}P_2(0)=0.15. \quad \text{(B13)}$$

[1] (a) H. Poincaré, *Méthodes Nouvelles de la Mécanique Céleste* (Dover, New York, 1957), Vols. 1–3. (b) A brief description of Poincaré's theorem on nonintegrability can be found in J. Ford, Phys. Rep. **213**, 271 (1992).

[2] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **98**, 527 (1954).

[3] V. I. Arnold, Usp. Mat. Nauk **18**, 91 (1963).

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

[5] J. Moser, Nachr. Akad. Wiss. Göttingen Math. Phys. Kl. **1**, 1 (1962).

[6] E. Fermi, *Collected Papers* (University of Chicago Press, Chicago, 1965), Vol. 2, p. 978.

[7] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P.S. Pershan, Phys. Rev. **127**, 1918 (1962).

[8] E. Fermi, Z. Phys. **71**, 250 (1931).

[9] V. I. Arnold, Sov. Math. Dokl. **3**, 136 (1963).

[10] B. V. Chirikov, Phys. Rep. **52**, 263 (1979).

[11] B. V. Chirikov and V. V. Vecheslavov, J. Stat. Phys. **71**, 243 (1993).

[12] Jyhpyng Wang, Ph.D. thesis, Harvard University, 1988 (unpublished).

[13] A comprehensive review can be found in Ref. [1(b)].

[14] M. Pettini and M. Landolfi, Phys. Rev. A **41**, 768 (1990).

[15] C. G. Goedde, A. J. Lichtenberg, and M. A. Lieberman, Physica D **59**, 200 (1992).

[16] H. Kantz, R. Livi, and S. Ruffo, J. Stat. Phys. **76**, 627 (1994).

[17] J. Ford and G. H. Lunsford, Phys. Rev. A **1**, 59 (1970).

[18] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1971).