http://www.aimspress.com/journal/mine

Mathematics in Engineering, 3(3): 1–22.
DOI:10.3934/mine.2021025
Received: 05 March 2020
Accepted: 24 May 2020
Published: 23 July 2020

Research article

Understanding the FPU state in FPU–like models†

Giancarlo Benettin* **and Antonio Ponno**

Università degli Studi di Padova, Dipartimento di Matematica "Tullio Levi-Civita", Via Trieste 63, 35121 Padova, Italy

† **This contribution is part of the Special Issue:** Modern methods in Hamiltonian perturbation theory
Guest Editors: Marco Sansottera; Ugo Locatelli
Link: www.aimspress.com/mine/article/5514/special-articles

* **Correspondence:** Email: benettin@math.unipd.it.

Abstract: Many papers investigated, in a variety of ways, the so-called "FPU state" in the Fermi-Pasta-Ulam model, namely the state, intermediate between the initial state and equipartition, that the system soon reaches if initially one or a few long-wavelength normal modes are excited. The FPU state has been observed, in particular, to obey a few characterizing scalings laws. The aim of this paper is twofold: First, reviewing and commenting the literature on the FPU state, suggesting a possible way to organize it. Second, contributing to a better understanding of the FPU state by studying the similar state in the Toda model, which provides, as is known, the closest integrable approximation to FPU. As a new tool, we analyze the dimensionality of Toda invariant tori in states corresponding to the FPU state, and observe it obeys the main scaling law characterizing the FPU state.

Keywords: FPU problem; Toda model; integrability; scaling laws; Toda actions

Foreword

This paper is dedicated to our colleague and friend Antonio Giorgilli, in occasion of his 70$^{\text{th}}$ birthday. The scientific contribution of Antonio, in different branches of Mathematical Physics, is well known and is invaluable. Concerning in particular FPU, beyond the interest of specific results, Antonio had a special propulsive role in Italy, and influenced very much our own work. "Understanding the FPU state...", which is the title of the present paper, started for us from studying [14] and the scaling law (3) there illustrated. Thanks, Antonio!

1. Introduction

This paper is devoted to the Fermi-Pasta-Ulam (FPU) problem [1], namely the problem of energy sharing in weakly nonlinear chains of oscillators. The Hamiltonian has the form

$$H(p,q) = \frac{1}{2}\sum_{i=1}^{N} p_i^2 + \sum_{i=0}^{N} V(q_{i+1} - q_i)\,, \tag{1}$$

V being some nearest-neighbors potential with single minimum in zero,

$$V(r) = \frac{r^2}{2} + \alpha\frac{r^3}{3} + \beta\frac{r^4}{4} + \cdots\,, \qquad \beta > 0\,. \tag{2}$$

In (1) the boundary conditions are still not specified; they are generally either fixed ends, i.e., $q_0 = q_{N+1} = 0$ (like in the original FPU paper), or periodic, $q_N = q_0$. The fixed-ends case can be viewed as a submanifold of the periodic case with $N' = 2(N+1)$ degrees of freedom, with symmetry $q_{N'-j} = -q_j$.

The value of α, if different from zero, is irrelevant, since a trivial rescaling reports it to any prefixed value; the effective parameters determining the dynamics are indeed $|\alpha|\sqrt{\varepsilon}$, $\varepsilon = E/N$ being the specific energy, and then $\beta/\alpha^2, \ldots$ The choice of α actually fixes the energy scale, as well as the scale of β and of possibly further coefficients in (2).

As is well known, Fermi, Pasta and Ulam aimed to investigate how the system reaches the statistical equilibrium, identified with the equipartition of energy among normal modes, if started very far from equilibrium, the whole energy being initially given to only one or two long-wavelength normal modes. With great surprise they found, within the time scale accessible to their computer,* no equilibrium at all: the system apparently reached a stationary state, different both from the initial state and from equipartition, in which only a few normal modes significantly share energy. Such a peculiar state is often called "the FPU state". Besides the lack of equipartition, the FPU state is characterized by long time recurrences, made especially evident a few years later in [3].

The literature on FPU, 65 years after the original FPU paper, is so huge, that it is not even conceivable to summarize it through a reasonable number of main references. Two collections of papers [4, 5], produced in the occasion of the 50th anniversary of FPU, may give an idea of the status of the art around 2004–2005, as well as of the widespread attitudes of different authors, or different schools, in front of such non easy subject which touches the very foundations of classical statistical mechanics.

Two main questions have been extensively studied in the literature, since the very beginning:

i. Understanding the unexpected FPU state: its characteristic features, its long time recurrences, how is it produced from initial data, how is it possible that, in spite of the nonlinearity which couples normal modes, energy is shared among only a few of them.
ii. Understanding whether, possibly on a much larger time scale, the FPU state further evolves, eventually reaching thermalization and normal statistical behavior, as in the original FPU expectation.

*The celebrated MANIAC I of the Los Alamos Laboratories, capable to process approximately 10^4 instructions per second [2]. Compare with modern computers, having hundreds or thousands of CPUs with clock at several GHz.

Both questions have been studied rather soon, independently. On the one hand, the recurrences observed in [1,3] suggested the presence of a close underlying integrable dynamics, which was initially searched, starting with [6] in 1965, among nonlinear wave equations, and later on, starting with [8] and [9], in the Toda model. On the other hand, almost simultaneously in 1966, the possibility was outlined [7] that normal statistical behavior is recovered if the specific energy $\varepsilon = E/N$ exceeds a threshold, which in turn, in [7], was conjectured to vanish for $N \to \infty$.[†]

The literature on FPU accumulated, along these lines, without much order, and it is not so easy to follow it. Even the basic distinction between formation of the FPU state and possible later evolution to normal statistical behavior, with different phenomena occurring on different time scales, was not always clear at the beginning, and started being emphasized only in 1982 in [11, 12], where a possible connection between the FPU state and metastability was outlined. Not so many papers are based on this distinction; among them, [10, 13–18].

In this paper we focus the attention on the former problem, namely understanding as much as possible the FPU state. The aim is to revisit the existing literature, tracing possible paths through it, and then to provide a further understanding of the FPU state by looking more closely at the Toda model.

Browsing the literature, we extract two main ideas (see Section 2 for references):

– The FPU state is characterized by a few simple scaling laws; the most relevant one is that the number M of normal modes effectively involved in the energy sharing, when one or a few long-wavelength normal modes are initially excited, scales as

$$M \sim N\varepsilon^{1/4} . \tag{3}$$

– The reference integrable dynamics underlying FPU, is not the linear one, made of free normal modes, rather the essentially nonlinear Toda dynamics.

Our contribution, in this background, is the following:

– We check in Toda the scaling laws known to hold for FPU, and find they are fully satisfied. Scaling laws appear thus to be proper of Toda, and FPU obeys them as long as its dynamics remains close to the Toda dynamics.
– We exploit the numerical procedure explained in [9] to compute numerically the Toda actions; the procedure allows to determine, in any given state (p, q), the number of nonvanishing actions, that is the effective dimensionality of the corresponding Toda torus. We consider states in which only one or a few low frequency modes are excited, and observe that the dimensionality $\mathcal{M}$ of the torus scales exactly as in (3),

$$\mathcal{M} \sim N\varepsilon^{1/4} . \tag{4}$$

This means: the scaling laws characterizing the FPU state are contained in the nontrivial correspondence between normal modes and Toda actions. This is a more innovative view, in which dynamics (integration of Hamilton equations) does not play a role. By the way, the accuracy of the scaling law (4) is the best we ever met (see in particular Figures 8 and 9 below).

[†]Ref. [7], as well as many later papers investigating the conditions for thermalization in FPU, deal with the so-called β model, having $\alpha = 0$. The two models in fact (together with other variants of FPU, like Lennard-Jones) quite often have been considered equivalent and interchangeable, without much criticism. Certainly, they are not for what concerns the FPU state, which is peculiar of the α-models. The long time behavior, with thermalization on a suitably long time scale, is instead rather similar; see for a comparison [10].

The paper is organized as follows: Section 2, after a few preliminaries, proposes our personal way to look at the literature on the FPU state, and also includes (Sect. 2.4) the new results concerning the scaling laws in the Toda model. Section 3 is instead devoted to the computation of the Toda actions. An Appendix follows, in which the procedure to compute the Toda actions is shortly recalled.

2. The FPU state: scaling laws from dynamics

2.1. A first look at the FPU state

Consider, to be definite, the case of fixed ends. Normal modes are then

$$Q_k = \sqrt{\frac{2}{N+1}} \sum_{i=1}^{N} q_i \, \sin \frac{\pi k i}{N+1} \,, \qquad P_k = \sqrt{\frac{2}{N+1}} \sum_{i=1}^{N} p_i \, \sin \frac{\pi k i}{N+1} \,;$$

their energies E_k and frequencies ω_k are

$$E_k = \frac{1}{2}(P_k^2 + \omega_k^2 Q_k^2) \,, \qquad \omega_k = 2 \, \sin \frac{\pi k}{2(N+1)} \,.$$

For given initial data let[‡]

$$\overline{E}_k(T) = \frac{1}{\frac{1}{3}T} \int_{\frac{2}{3}T}^{T} E_k(P(t), Q(t)) \, \mathrm{d}t \,.$$

A preliminary qualitative view of the formation of the FPU state, and of its later slow evolution to equipartition, is provided by Figure 1 (a remake of Figure 1 in [19]) which refers to $N = 1023$, $\alpha = -1$, $\beta = 2$, specific energy $\varepsilon = 10^{-4}$, and shows $\overline{E}_k(T)$ vs. k/n at selected times T (marked in the figure) in geometric progression. The energy was initially equidistributed among the lowest 10% of modes (the rectangular profile in the figure). One can see that quite soon, already at $T \simeq 10^3$, a well defined profile is formed, in which only some low frequency modes effectively take part to the energy sharing, the energies of the remaining ones decaying exponentially with k/N. The energy profile keeps its form nearly unchanged for a rather large time scale, $T \simeq 10^5$ or 10^6, definitely much larger than the time needed to form it. Afterwords, on a much larger time scale, the dynamics slowly evolves towards energy equipartition, the high-frequency modes being progressively involved into the energy-sharing game; in the above conditions, equipartition requires $T \simeq 10^{10}$. The initial phases (this will have importance, see later) were chosen randomly; moreover, to improve a little the figure reducing the fluctuations, an average of $E_k(T)$ on 24 different random choices of the phases was performed.

If the initial energy distribution is narrower, then the formation of the FPU state is slower, but the state looks the same, as well at the later evolution; see Figure 2, left panel, where the initial distribution is $0 < k/N < 0.03$. If energy is raised, the width of the FPU state grows, see the right panel of Figure 2 where $\varepsilon = 4 \times 10^{-4}$.

We shall now try to order the literature on the FPU state, by tracing two possible paths through it.

[‡]Restricting the average to the running window $\frac{2}{3}T \le t \le T$ does not change the asymptotics and makes a little quicker the convergence: time averages from the beginning are somehow lazy.

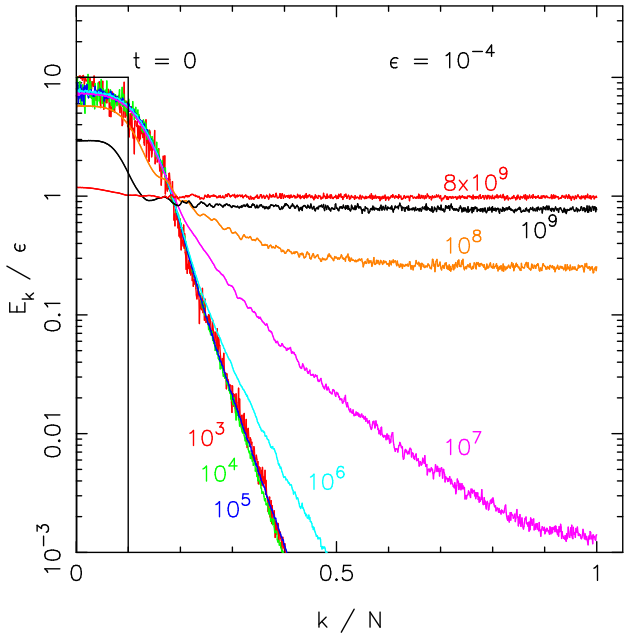

Figure 1. The shape of the averaged energy spectrum of normal modes $\overline{E}_k(T)$ plotted vs. k/N, at selected times T (marked in the figure) in geometric progression. Fixed ends FPU, $N = 1023$, $\alpha = 1$, $\beta = 2$, $\varepsilon = 10^{-4}$. Energy initially equidistributed among modes with $0 < k/N < 0.1$, see the rectangle marked $t = 0$. Each point is the average over 24 random extractions of the initial phases.

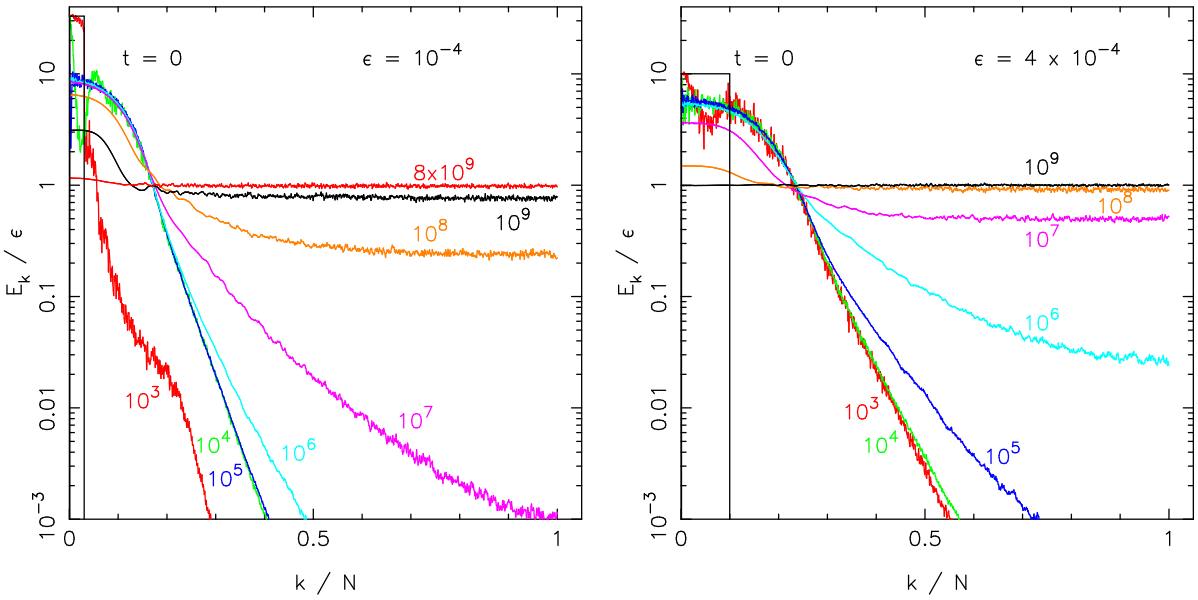

Figure 2. Left: same as in Figure 1, with narrower initial energy distribution, $0 < k/N < 0.03$. Right: same as in Figure 1, at higher energy $\varepsilon = 4 \times 10^{-4}$.

2.2. Path one: scaling laws in FPU

The FPU state needs to be characterized quantitatively, looking in particular at its width—how many modes are effectively involved—and at the time scale for its formation, to be thought of as depending on N and ε as well as on the kind of the initial excitation.

Let us start by numerical evidence. Both in [20] and [14] a convincing evidence is produced (particularly accurate in [14]) that the width of the FPU state, namely $f = M/N$ where M is the (suitably defined) number of modes effectively sharing energy, for large N depends only on the specific energy ε and not on N, and scales as

$$f \sim \varepsilon^{1/4} ; \tag{5}$$

correspondingly, M obeys (3). One should remark that [20] deals with a periodic pure α model, and considers an initial excitation flat on a fraction f_0 of modes already proportional to $\varepsilon^{1/4}$, while [14] deals with a fixed ends model with $\beta = 4\alpha^2$, and only the first mode is initially excited. The notion of "number of modes effectively shearing energy", that is the definition of M, is also different, see the original papers; it is remarkable that the scaling law (5) is the same—the phenomenon looks, so to speak, robust. Some numerical evidence of (5) was already present in [21]. Concerning the time of formation of the FPU state, in [20] we find the estimate

$$t \sim \varepsilon^{-3/4} , \tag{6}$$

holding for large N and initial data as before, that is for f_0 obeying (5).[§]

As is remarkable, the scaling laws (5) and (6) have been also justified theoretically, looking at the way normal modes interact. More precisely, (5) is deduced in [21] by adapting to the α FPU model the Chirikov method of overlapping of resonances, originally formulated for the β model [7]; both (5) and (6) are instead deduced in [13], on which [20] is based, by adapting to FPU some ideas from the theory of weak turbulence. In both cases, the reference uncoupled model is the harmonic chain.

The scaling laws entering the formation of the FPU state have been later studied in [22] in a more systematic way. The conventional definition of M—actually the same as in [20]—is as follows: let

$$\eta = -\sum_{k=1}^{N} F_k \log F_k , \qquad \text{where} \quad F_k = \frac{\overline{E}_k}{\sum_j \overline{E}_j} ,$$

denote the "entropy" of the spectrum [23]; then[¶]

$$M = e^{\eta} .$$

Consider initial data with the first lowest M_0 modes initially excited; think f to be a function of t, N, ε and $f_0 = M_0/N$. The results in [22] can be summarized in two general scaling laws for $f(t, N, \varepsilon, f_0)$, both holding for large N, depending on whether the phases of the initially excited modes are chosen randomly or are coherent.

[§]No explicit estimate of t as function of ε is provided in [14], but an indication is present, if we correctly interpret, $t \sim \varepsilon^{-1/2}$. We conjecture the authors have seen the correct law $t \sim \varepsilon^{-3/8}$ valid for fixed M_0, see below, which is close to $\varepsilon^{-1/2}$ and easily confused with it in a non systematic study.

[¶]Let $\overline{E}_k = E/M$ for $k \le M$, $\overline{E}_k = 0$ for $k > M$. Then it is $\eta = \log M$.

(a) For random initial phases, f is independent of N and scales as

$$f(t, N, \varepsilon, f_0) = \varepsilon^{1/4} \mathcal{G}(\varepsilon^{3/8} f_0^{3/2} t) \,, \tag{7}$$

where $\mathcal{G}$ is a kind of universal function of sigmoidal shape, reaching a plateau when its argument grows to infinity. According to (7), the width of the FPU state obeys (5), no matter which is f_0. The time scale for the formation of the state scales instead as

$$t \;\sim\; \varepsilon^{-3/8} \, f_0^{-3/2} \,. \tag{8}$$

The main interest, having in mind thermodynamics, is for fixed f_0, that is M_0 proportional to N (modes excited up to a given frequency). The scaling law (8) however also holds for $f_0 \sim \varepsilon^{1/4}$, as in [20], giving in this case (6), as well as down to small fixed M_0, as in [14], giving in such case times growing with N,

$$t \;\sim\; \varepsilon^{-3/8} \, N^{3/2} \,; \tag{9}$$

for fixed small M_0 the assumption that the initial phases are random is meaningless, and is in fact irrelevant to deduce (9).

(b) If instead many modes are initially excited with coherent initial phases, for example equal phases, or phases alternating in some easy pattern, then the total energy E, rather than the specific energy ε, appears to be the the relevant variable, and (7) is replaced by

$$f(t, N, \varepsilon, f_0) = E^{1/4} f_0^{1/4} \mathcal{G}'(E^{3/8} f_0^{15/8} t) \,. \tag{10}$$

For large E, in particular for fixed ε and large N, the intermediate FPU state practically confuses with equipartition among all modes: unless of course $f_0 \sim 1/N$, in which case, with interesting continuity, (9) is recovered.

It is worthwhile to remark that the choice of equal initial phases, although not well motivated, is rather common in the literature: it occurs indeed whenever, for example, either the q_i's or the p_i's initially vanish, and correspondingly energy is purely kinetic or purely potential (the easiest initial data to be handled numerically). With such initial data, unless $f_0 \sim N^{-1}$, the FPU state for large N disappears. The lack of attention to initial phases, in our opinion, is among the reasons why it not so easy, sometimes, to compare papers.

In [22] the general scaling laws (7) and (10) are also motivated theoretically, but in a way substantially different from [21] and [13], namely using, as a normal form for FPU, the KdV nonlinear wave equation. Considering FPU as a perturbation of a nonlinear integrable system is indeed an old idea, that we shall shortly revisit in the next subsection.

2.3. Path two: FPU state and integrability

As already remarked in the Introduction, to explain the presence of long time recurrences in FPU it was soon suggested that the FPU dynamics is close to the dynamics of a nonlinear but nevertheless integrable system, and this was first searched among nonlinear wave equations. The first attempt in this direction, going back to 1965, is [6], a fundamental paper which is at the basis of modern theory of nonlinear integrable wave equations, in which FPU appears as a main motivation to study again, after years, the KdV equation,

$$u_t = \tfrac{\alpha}{\sqrt{2}} u u_x + \tfrac{1}{24} u_{xxx} \,.$$

The progress in the field is then rapid: a couple of years later the method of inverse scattering [24] and Lax pairs [25] are introduced, and the presence of infinitely many constants of motion is established [26]. Finally, in 1971, KdV is shown to be an infinite dimensional completely integrable Hamiltonian system [27].

In fact, the nonlinear wave equation which is more immediately related to FPU, if one searches for a continuum limit in which the first nonlinear and the first dispersive terms beyond the wave equation are kept, is the Boussinesq equation; a possible form is

$$u_{tt} = u_{xx} + 2\alpha u_x u_{xx} + \frac{1}{12} u_{xxxx} \, .$$

From Boussinesq it is possible to deduce, in a convenient nontrivial limit, KdV, but the Boussinesq equation itself was soon proved, in 1973, to be integrable [28]. In [28] the connection between FPU and Boussinesq is quite emphasized, the lack of integrability of FPU being explained as due to the discretization.

In the meanwhile, statistical physicists become interested in the Toda model [29–32]. As is well known, this is a Hamiltonian system with the same form as (1), V being the Toda exponential potential

$$V_T(r) = \frac{1}{\lambda^2}(e^{\lambda r} - 1 - \lambda r) \, .$$

In 1974 the Toda model was proved to be completely integrable, remarkably in three independent papers [8, 33, 34]. Reference [8] is particularly important for FPU, because the connection with FPU is there emphasized. Indeed, for $\lambda = 2\alpha$ it is

$$V_T(r) = \frac{1}{2} r^2 + \frac{1}{3}\alpha r^3 + \frac{1}{4}\beta_T r^4 + \frac{1}{5}\gamma_T r^5 + \cdots \, ,$$

with

$$\beta_T = \tfrac{2}{3}\alpha^2 \, , \qquad \gamma_T = \tfrac{1}{3}\alpha^3 \, , \qquad \ldots \, ,$$

so the model is tangent to FPU up to the third order included and provides an integrable approximation to FPU better than the harmonic chain. In [8] the slow stochastization of FPU is no more attributed to the discretization with respect to an integrable wave equation, but to the small difference, with dominating term $\frac{1}{4}(\beta - \beta_T) r^4$ in the potential, between FPU and Toda. The reference to Toda as the best integrable approximation for FPU, is a considerable change of paradigm in the FPU problem.

The connection between FPU and Toda was proposed again, and emphasized, in 1982, in [9]. In this paper, on the one hand a striking evidence is provided that the dynamics of FPU, in the FPU state before the slow drift towards equipartition, is hardly distinguishable from the Toda dynamics; on the other hand, the integrability of Toda is studied very constructively, and an algorithm is proposed to compute numerically the Toda actions. We shall come back to this point in Section 3.

2.4. Scaling laws in the Toda model

The obvious conjecture, at this point, is that the scaling laws (7) and (10), characterizing the FPU state and implying in particular (5) and (6), are not proper of FPU but come from the underlying Toda integrable dynamics. This is what we shall see in this section, where we revisit and remake for Toda some numerical work originally done in [22] for FPU.

The scaling laws (7) and (10), reducing the variables from four to two, are obtained in [22] as combinations of three independent homogeneity relations characterizing the function $f(t, N, \varepsilon, f_0)$ for large N. Two of them do not depend on the choice of the phases of the excited modes and are

$$f(\lambda^{-3/4}t, \lambda^{-1/4}N, \lambda\varepsilon, \lambda^{1/4}f_0) = \lambda^{1/4} f(t, N, \varepsilon, f_0) \tag{11}$$

$$f(\mu^{3/2}t, \mu N, \varepsilon, \mu^{-1}f_0) = f(t, N, \varepsilon, f_0)\,. \tag{12}$$

We cannot enter here the way (11) and (12), which might appear obscure, are (heuristically) deduced from KdV, demanding for this to [22].

Figure 3 provides a test of the scaling law (11). The left panel shows f vs. t for eight different choices of N, ε and f_0 obeying (11), namely $N = 2896 - 9742$ with step $2^{-1/4}$, $\varepsilon = 10^{-3} - 7.8 \times 10^{-6}$ with step $1/2$, and $f_0 = M_0/N$ with constant $M_0 = 102$. The right panel shows the same curves rescaled according to (11), namely rescaled horizontally by a factor $(\varepsilon/\varepsilon_*)^{-3/4}$ and vertically by a factor $(\varepsilon/\varepsilon_*)^{1/4}$, with $\varepsilon_* = 2.5 \times 10^{-4}$ (the value of ε of the third curve from above, which does not move). One can see the rescaled curves superimpose almost exactly. The phases were chosen randomly, and each curve is in fact the average over 24 different random choices of the phases.

Figure 4 provides similarly a test of the scaling law (12). Here $\varepsilon = 2.4 \times 10^{-4}$ is kept fixed, while N ranges between 1023 and 16383; $f_0 = M_0/N$ with fixed $M_0 = 26$. The left panel shows the curves as they are, while in the right panel t is scaled by a factor $(N/N_*)^{-3/2}$, with $N_* = 4095$ (corresponding to the central curve, which stays unchanged). As before, each curve is the average of 24 different random choices of the initial phases.

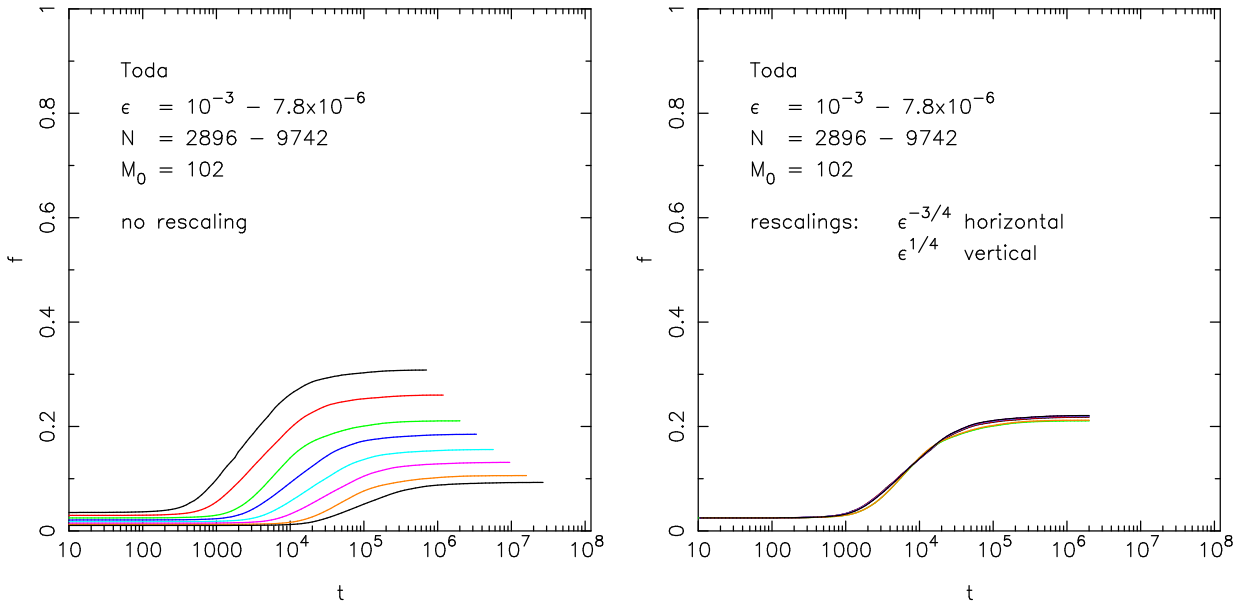

Figure 3. A test of the scaling law (11) in the Toda model. Parameters: $N = 2896 - 9742$, step $2^{-1/4}$; $\varepsilon = 10^{-3} - 7.8 \times 10^{-6}$, step $1/2$; $M_0 = 102$. Random phases (24 choices). The plotted quantity is f vs. t as it is (left) and rescaled (right).

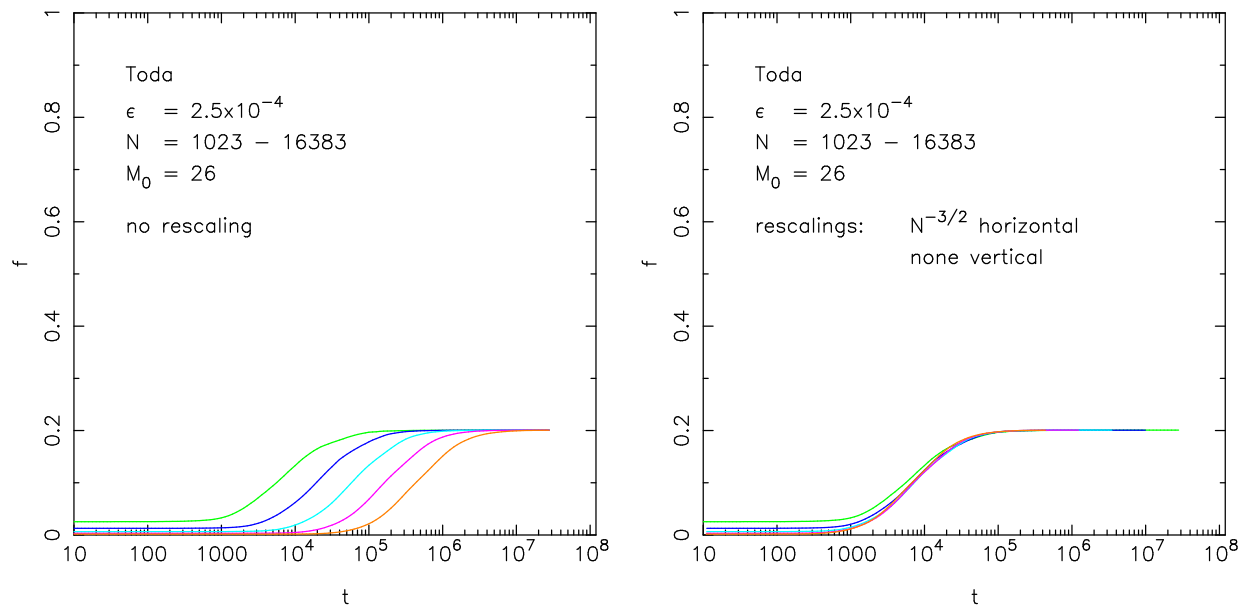

Figure 4. A test of the scaling law (12) in the Toda model. Parameters: $N = 1023 - 16383$, $\varepsilon = 2.5 \times 10^{-4}$, $M_0 = 26$. Random phases (24 choices). The plotted quantity is f vs. t as it is (left) and rescaled (right).

Putting together (11) and (12) one gets

$$f(t, N, \varepsilon, f_0) = \varepsilon^{1/4} \mathcal{F}(\varepsilon^{3/8} N^{-3/2} t, N f_0) \,, \tag{13}$$

with a convenient $\mathcal{F}$ depending only on two variables. Such a relation can be further tested by considering combinations of N, ε and f_0 different from the previous ones, for example one can keep N constant and move ε; the corresponding scaling factors are in this case $\varepsilon^{-3/8}$ on the horizontal axis, $\varepsilon^{1/4}$ on the vertical axis. Figure 5 shows the result for $N = 1023$, seven values of ε in the range $10^{-3} - 1.56 \times 10^{-5}$, $f_0 = M_0/N$ with small $M_0 = 3$. As before the left panel shows f vs. t as it is, while the right panel reports the rescaled curves.

The third relation which completes (11) and (12) depends instead in an essential way on the choice of the phases. For random phases, it simply tells that, for large N and given f_0, f is independent of N. This means $\mathcal{F}(x, y) = \mathcal{G}(xy^{3/2})$, and (7) is deduced. A test of the independence of f on N for large N at fixed f_0 is provided in Figure 6, where ε is either 10^{-3} or 2.5×10^{-4}, while N assumes five different values in the range $1023 - 16383$; $f_0 = 0.025$ is fixed. The left panel displays f vs. t; quite clearly, f depends on ε but not on N. The further rescaling of the vertical axis by a factor $\varepsilon^{1/4}$, shown in the right panel, makes all curves superimpose.

For coherent phases instead the additional scaling law is

$$f(t, \rho N, \rho^{-1} \varepsilon, f_0) = f(t, N, \varepsilon, f_0) \,; \tag{14}$$

that is, f is independent of N at fixed f_0 only if the total energy $E = N\varepsilon$ is kept fixed. From (14) and (9), (10) is deduced. A test of (14) is exhibited in Figure 7, left panel, where $E = 0.265$ and $f_0 = 0.1$ are kept fixed, while N assumes six values in the range $511 - 16383$. The right panel shows that if ε, instead of E, is fixed, then the independence of N is lost and for large N the FPU state does not exist

any more. The phases are here constant, namely $\varphi_k = \pi/2$ for all excited modes (purely potential initial energy), but a qualitative similar picture is obtained by choosing for example $\varphi_k = 0$, or $\varphi_k = k\pi/2$, or similar regular patterns.

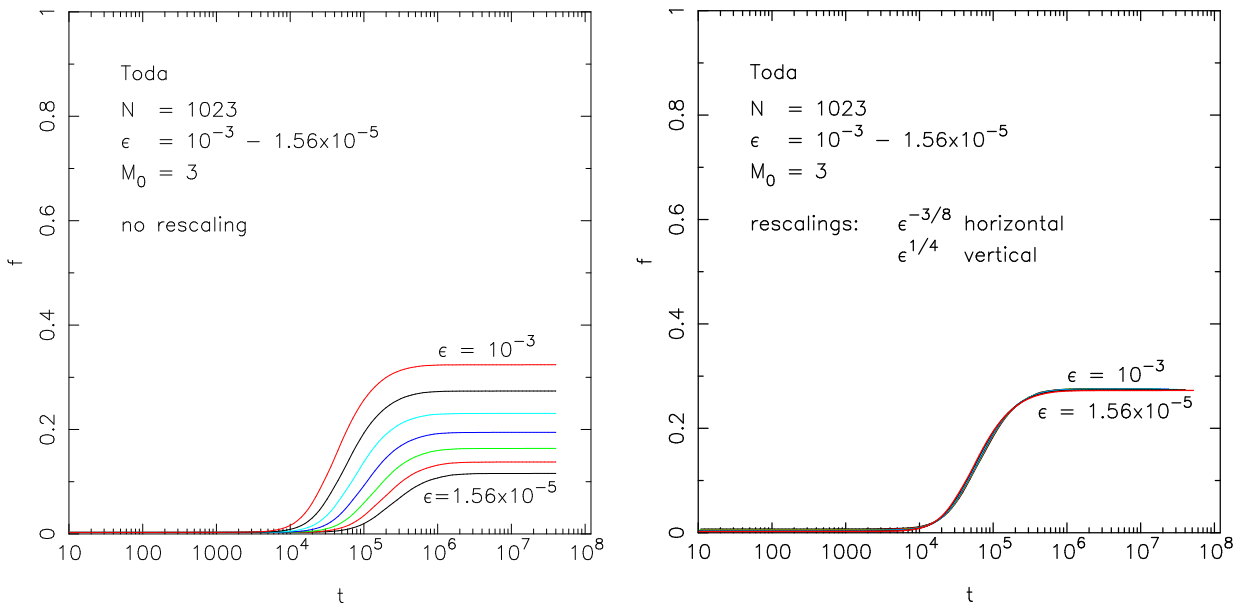

Figure 5. A test of the scaling law (13) in the Toda model. Parameters: $N = 1023$, $\varepsilon = 10^{-3} - 1.56 \times 10^{-5}$, $M_0 = 3$. The plotted quantity is f vs. t as it is (left) and rescaled (right).

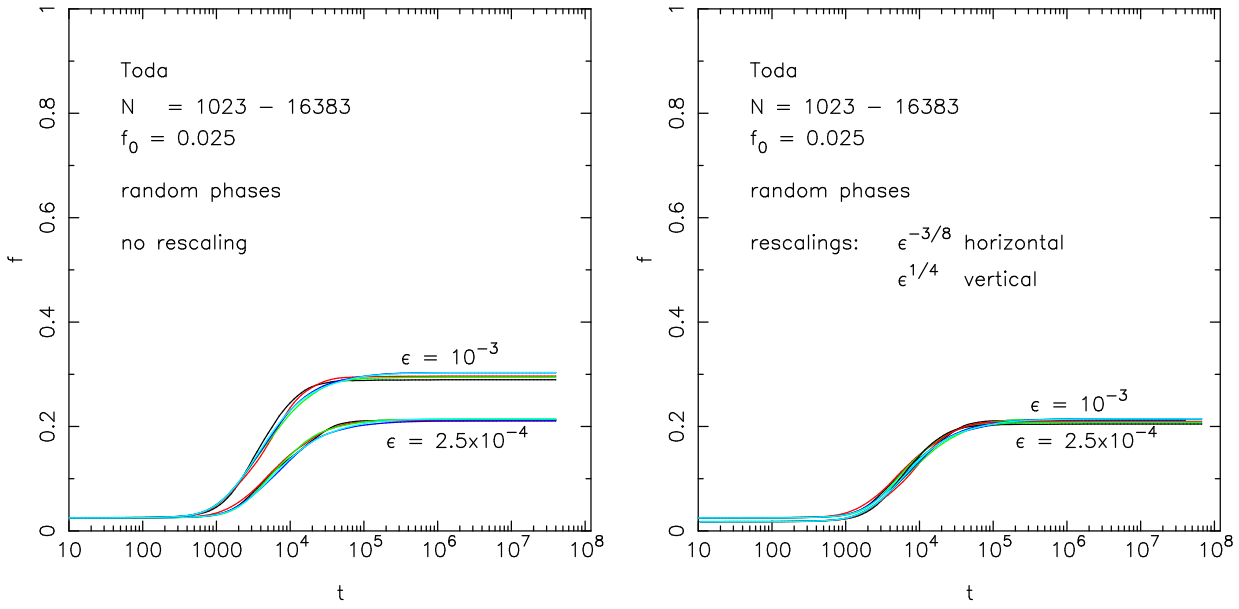

Figure 6. A test of the independence of f on N, at fixed f_0, for random initial phases. Parameters: ε either 10^{-3} or 2.5×10^{-4}, five different values of N in the range 1023 – 16383, $f_0 = 0.025$. Left panel: f vs. t as it is; right panel: vertical rescaling proportional to $\varepsilon^{-1/4}$.

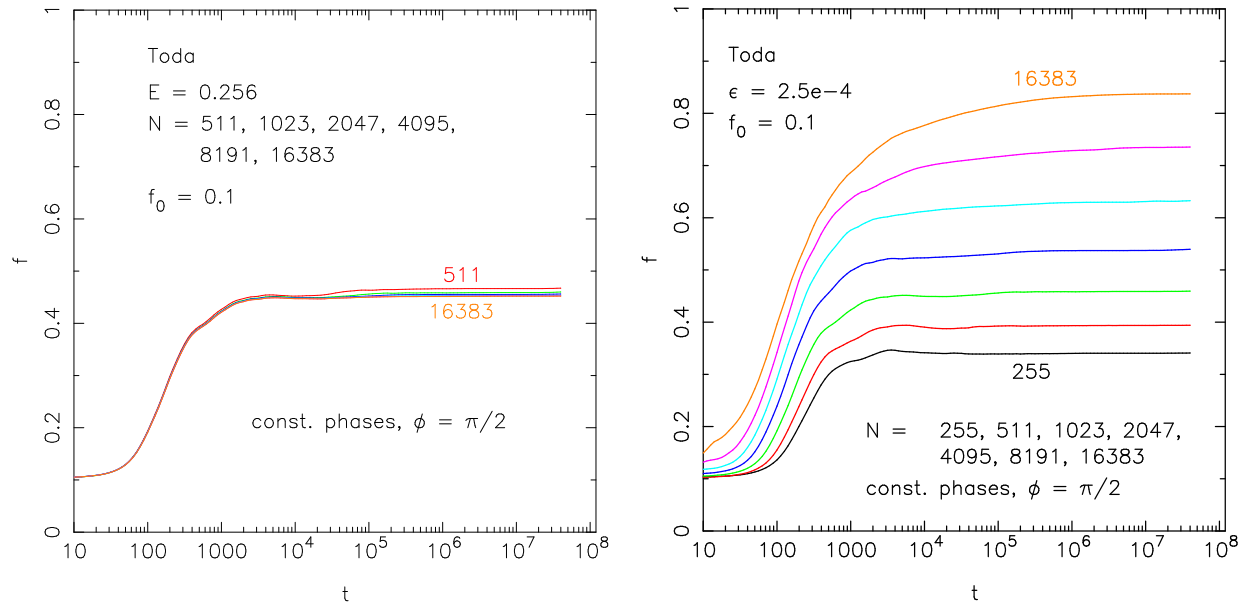

Figure 7. Left: $f(t, N, \varepsilon, f_0)$ vs.t for seven values of N in the range $255 - 16383$, fixed $E = 0.256$ and $\varepsilon = E/N$, fixed $f_0 = 0.1$. Right: fixed $\varepsilon = 2.4 \times 10^{-4}$ instead of fixed E.

3. More inside Toda: scaling laws from Toda actions

In this section we shall investigate the basic scaling law (5) following a substantially different road, namely looking at the correspondence between linear normal modes and action variables in the Toda model. The aim is understanding the scaling law (5) governing the FPU state in Toda by looking only at the initial state—how many actions are excited, which is the effective dimensionality of the torus supporting the dynamics—with no need to integrate the equations of motion.

Let us remark that computing the Toda actions for large N is a rather nontrivial work, facing delicate questions of numerical precision. We present here only a few preliminary results, without entering in details, with the only aim to show this is a promising way (see also [35]). We shall come back soon to the subject, in a wider devoted paper.

The appropriate definition of Toda actions, as well as the numerical procedure to compute them, are provided in [9]; for the convenience of the unfamiliar reader, however, the essentials are reported in the Appendix. The natural model to look at, in this connection, is a periodic chain with N particles, that is, eliminating the barycenter, with $N - 1$ degrees of freedom. Let us recall that for the corresponding harmonic periodic chain, normal modes are not stationary waves but traveling waves; a convenient choice is

$$q_i(t) = \sum_{k=1}^{N-1} A_k \, q_i^{(k)}(t) \,, \qquad i = 1, \ldots, N \,,$$

where

$$q_i^{(k)}(t) = \sqrt{\frac{2}{N}} \, \frac{1}{\omega_k} \, \sin\Big(\frac{2\pi i k}{N} + \omega_k t + \theta_k\Big) \,, \qquad \omega_k = 2 \sin \frac{\pi k}{N}$$

(similarly for $p_i(t)$). The harmonic energies and harmonic actions are respectively

$$E_k = A_k^2 \,, \qquad I_k = E_k/\omega_k \,.$$

Let $\mathcal{I}_1, \ldots, \mathcal{I}_{N-1}$ denote the Toda actions, as defined in [9]. They turn out to be perturbations of the harmonic actions $I_1, \ldots, I_n$, and in particular, if a single harmonic action I_k is different from zero, then in the limit of vanishing energy E, it is

$$\frac{\mathcal{I}_k}{I_k} \to 1 \,, \qquad \frac{\mathcal{I}_{k'}}{I_k} \to 0 \quad \text{for} \quad k' \neq k \,. \tag{15}$$

The limit however, see [36–38], is highly non uniform in N, and for small k, to get convergence, EN^3, equivalently εN^4, needs to be small. The perturbative approach instead fails, and the correspondence between normal modes and actions gets out of theoretical control, for large N at fixed ε, that is in the thermodynamic limit.

Having in mind the FPU state and the scaling law (5) it satisfies, we naturally expect that for large N, if one or a small number M_0 of waves of low k are present, then the number of Toda actions that are substantially different from zero scales as $N\varepsilon^{1/4}$. This is precisely what we found in our numerical computations.

To define and compute the effective number $\mathcal{M}$ of non vanishing actions in any given state of the system, we found convenient to resort to the analog of spectral entropy, namely

$$\mathcal{H} = -\textstyle\sum_k \mathcal{F}_k \log \mathcal{F}_k \,, \qquad \text{where} \quad \mathcal{F}_k = \dfrac{\mathcal{E}_k}{\sum_j \mathcal{E}_j} \,, \quad \mathcal{E}_k = \omega_k \mathcal{I}_k \,;$$

then we let, in complete analogy, $\mathcal{M} = e^{\mathcal{H}}$.

Figure 8 reports $\mathcal{M}$ vs. ε in the wide range $10^{-15} \leq \varepsilon \leq 1$, for several values of N between 32 and 8192; the state is a single traveling wave with $k = 1$. The solid lines drawn in the figure are obtained by a least square fit, and all slopes, denoted a and reported in the figure aside N, turn out to be very close to $1/4$. So, the behavior is as expected: in agreement with (15), all curves converge to $\mathcal{M} = 1$ for $\varepsilon \to 0$, while for larger ε, with crossover depending on N, $\mathcal{M}$ grows proportionally to $\varepsilon^{1/4}$.

Figure 9, left panel, is the same as Figure 8, but the fraction $f = \mathcal{M}/(N-1)$ of nonvanishing actions, instead of $\mathcal{M}$, is there reported vs. ε. Quite clearly all curves flatten, by growing ε, on the same line. The solid line drawn in the figure interpolates the data for $N = 8192$ and has slope $a = 0.25$. An overall check that the scaling law is

$$\mathcal{M} \;\sim\; (N-1)\,\varepsilon^{1/4} \tag{16}$$

is provided by the right panel of Figure 9, where $\mathcal{M}$ is reported vs. $(N-1)\varepsilon^{1/4}$. Here too, the line interpolates data for $N = 8192$; its computed slope is 0.998. In conclusion, in the state of single traveling wave with $k = 1$, the expected scaling law is satisfied with quite great accuracy.

Such an accuracy was indeed a nice surprise, and was not really expected, the numerical procedure to compute the actions being new for us and somehow problematic (see the Appendix, Section C). It is hard to say how reliable are the values of the slopes reported in Figure 8. The only easy analysis is the uncertainty of the least-squares fit giving a. Figure 10 reports the computed slopes vs. N, the error bars corresponding, as is typical, to ± 3 std. deviations; the numerical results are clearly compatible with $a = 0.25$. Such an analysis of the accuracy does not take into consideration other possible sources of

error: At the moment we are not able to draw error bars around the symbols of Figure 8, and the choice of the intervals of ε where performing the best-fit analysis, after the crossover with the horizontal line $N = 1$, is also somehow arbitrary. Anyhow, the overall precision looks remarkable.

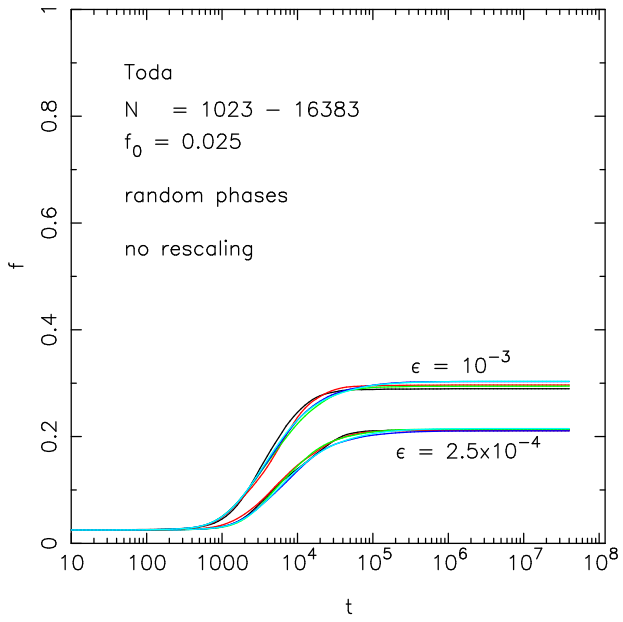

Figure 8. The effective number $\mathcal{M}$ of nonvanishing actions vs. ε, for several values of N in the range $32 - 8192$. Single traveling wave, $k = 1$. Solid lines are best fit interpolations, slopes are reported in the figure aside N.

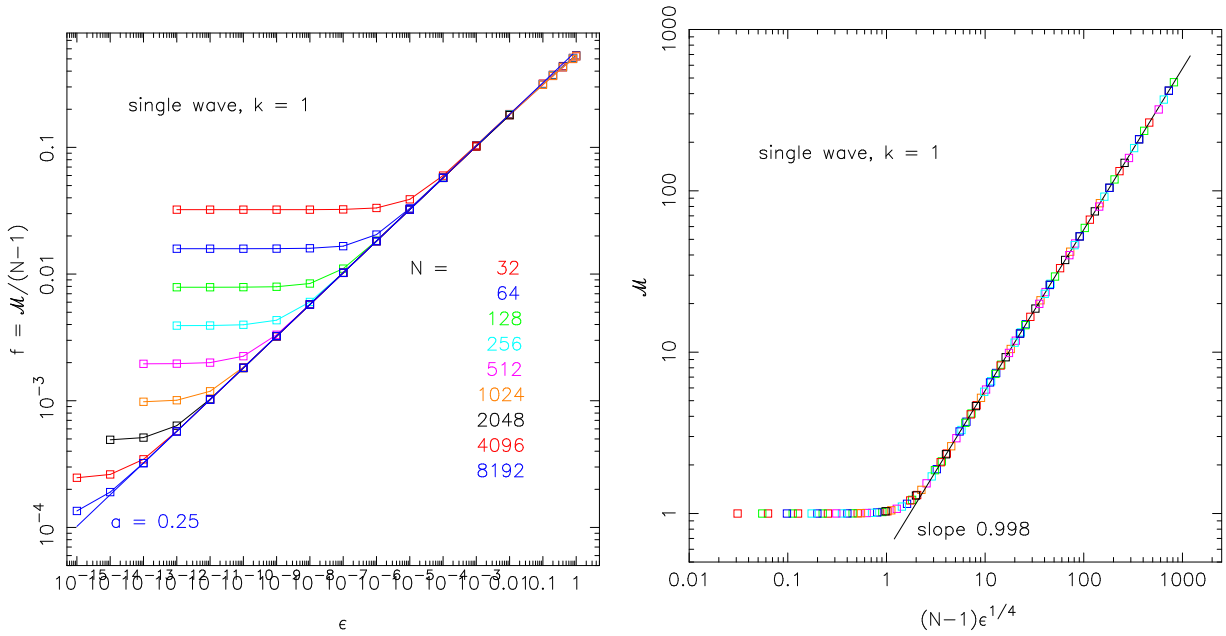

Figure 9. Left: as in Figure 8, $f = \mathcal{M}/(N-1)$ in place of $\mathcal{M}$. Right: as in Figure 8, $(N-1)\varepsilon^{1/4}$ in the abscissa.

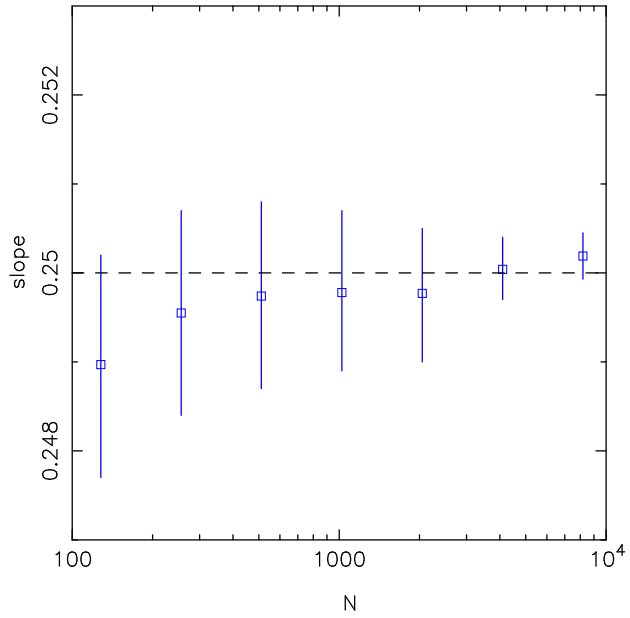

Figure 10. The slopes of the lines appearing in Figure 8 vs. N, with error bars given by the best-fit analysis (±3 std. deviations).

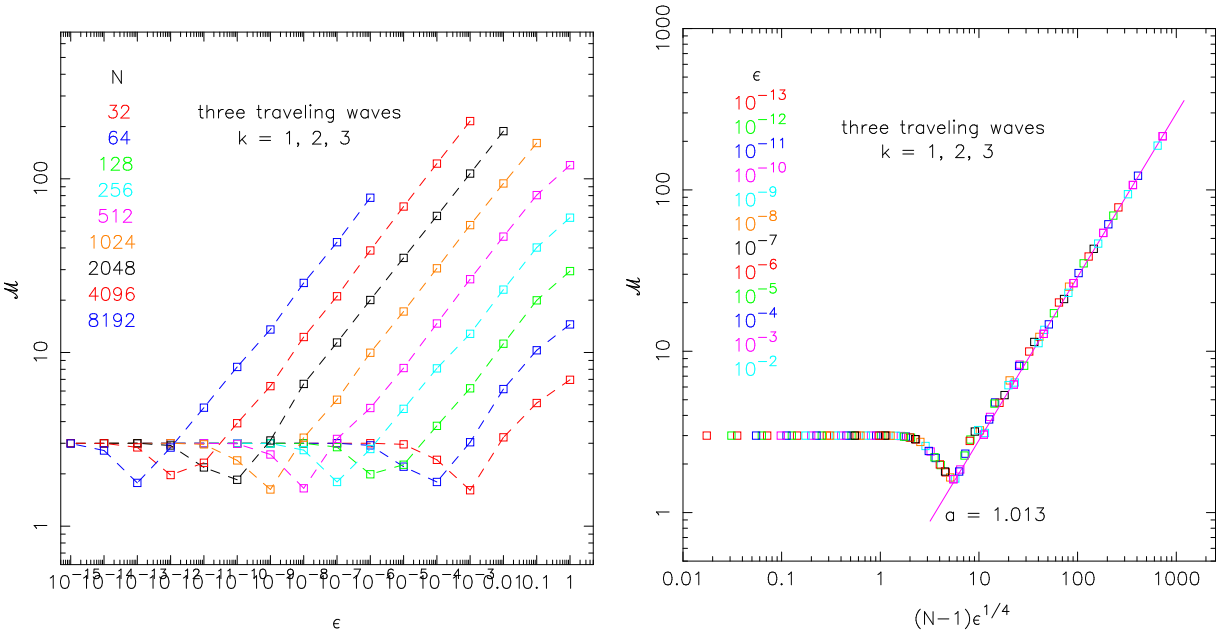

Figure 11. Left: same as Figure 8, for three traveling waves with $k = 1, 2, 3$. Right: same as Figure 9-left.

Nothing essentially changes if, in place of a single traveling wave with $k = 1$, one considers, for example, three waves with $k = 1, 2, 3$, having the same $\mathcal{E}_k$. Figure 11 shows, in the left panel, the analog of Figure 8, that is $\mathcal{M}$ vs. ε, and in the right panel the same as Figure 9, right panel, that is $\mathcal{M}$ vs. $N\varepsilon^{1/4}$. Now the limit for $N\varepsilon^{1/4} \to 0$ is of course $\mathcal{M} = 3$. The figure shows a curious phenomenon,

that we did not try to investigate: before growing, here too as $\varepsilon^{1/4}$, $\mathcal{M}$ decreases from 3 to about 2, and this for any N. The accuracy of the scaling law is not as good as for a single wave, but remains satisfactory.

Considering long waves with small k is essential to produce a FPU state obeying the scaling law (16). If instead we consider a single short wave excitation with, for example, $k = 0.4N$ (recall the maximal frequency is at $k = N/2$), then the limit $\mathcal{M} \to 1$ for $\varepsilon \to 0$ looks uniform in N, and up to rather large $\varepsilon \simeq 0.1$, see Figure 12, only one action is appreciably different from zero. Note that the vertical scale is quite enlarged with respect to Figure 8: passing from $\mathcal{M} = 1$ to 2 requires $\varepsilon \simeq 1$.

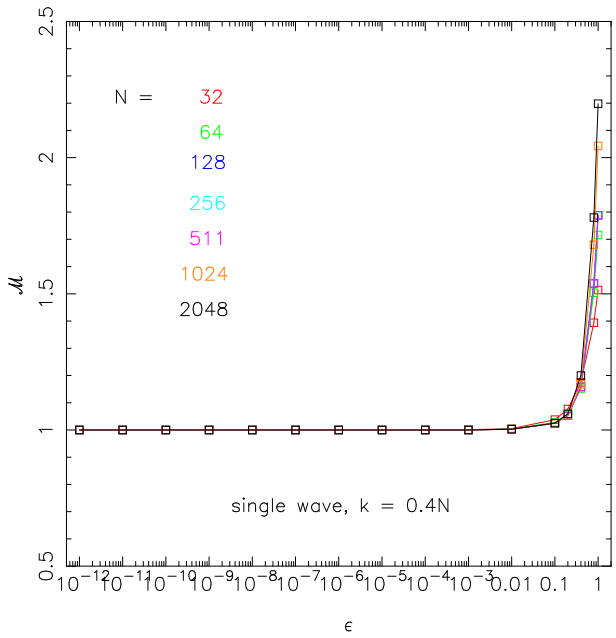

Figure 12. Same as Figure 8, for a single traveling wave with $k = 0.4N$.

Acknowledgements

We feel indebted to Martino Cecchetto (Padova) for helping us in understanding the computation of the Toda actions.

Conflict of interest

The authors declare no conflict of interest.

References

1. Fermi E, Pasta J, Ulam S (1955) *Studies of Non Linear Problems,* Los-Alamos Internal Report, Document LA-1940.
2. Lazarus RB, Voorhees EA, Wells MB, et al. (1978) *Computing at LASL in the 1949s and 1950s*, Los Alamos internal note LA-6943-H, part III.

3. Tuck JL, Menzell MT (1972) The superperiod of the nonlinear weighted string (FPU) problem. *Adv Math* 9: 399–407.
4. Campbell DK, Rosenau P, Zaslavsky GM (2005) Introduction: The "Fermi–Pasta–Ulam" problem—the first 50 years. *Chaos* 15: 015101.
5. Gallavotti G (2008) *The Fermi-Pasta-Ulam Problem: A Status Report*, Berlin-Heidelberg: Springer.
6. Zabusky NJ, Kruskal MD (1965) Interaction of solitons in a collisionless plasma and the recurrence of initial states. *Phys Rev Lett* 15: 240–245.
7. Izrailev FM, Chirikov BV (1966) Statistical properties of a nonlinear string. *Sov Phys Dokl* 11: 30–34.
8. Manakov SV (1974) Complete integrability and stochastization of discrete dynamical systems. *Sov Phys JEPT* 40: 269–274.
9. Ferguson WE, Flaschka H, McLaughlin DW (1982) Nonlinear Toda modes for the Toda chain. *J Comput Phys* 45: 157–209.
10. Benettin G, Ponno A (2011) Time–scales to equipartition in the Fermi–Pasta–Ulam problem: Finite-size effects and thermodynamic limit. *J Stat Phys* 144: 793–812.
11. Fucito E, Marchesoni F, Marinari E, et al. (1982) Approach to equilibrium in a chain of nonlinear oscillators. *J Phys* 43: 707–713.
12. Livi R, Pettini M, Ruffo S, et al. (1983) Relaxation to different stationary states in the Fermi-Pasta-Ulam model. *Phys Rev A* 28: 3544–3552.
13. Kramer PR, Biello JA, L'vov YV (2003) Application of weak turbulence theory to FPU model, In: *Proceedings of the Fourth International Conference on Dynamical Systems and Differential Equations (May 24–27, 2002, Wilmington, NC, USA)*, AIMS Conference Publications, 482–491.
14. Berchialla L, Galgani L, Giorgilli A (2004) Localization of energy in FPU chains. *Discrete Cont Dyn-A* 11: 855–866.
15. Bambusi D, Ponno A (2006) On metastability in FPU. *Commun Math Phys* 264: 539–561.
16. Benettin G, Carati A, Galgani L, et al. The Fermi–Pasta–Ulam problem and the metastability perspective, In: *The Fermi-Pasta-Ulam Problem*, Berlin: Springer, 151–189.
17. Carati A, Galgani L, Giorgilli A, et al. (2007) FPU phenomenon for generic initial data. *Phys Rev E* 76: 022104/1–4.
18. Carati A, Galgani L, Giorgilli A (2004) The Fermi–Pasta–Ulam problem as a challenge for the foundations of physics. *Chaos* 15: 015105.
19. Benettin G, Christodoulidi H, Ponno A (2013), The Fermi–Pasta–Ulam problem and its underlying integrable dynamics. *J Stat Phys* 152: 195–212.
20. Biello JA, Kramer PR, L'vov YV (2003) Stages of energy transfer in the FPU model, In: *Proceedings of the Fourth International Conference on Dynamical Systems and Differential Equations (May 24–27, 2002, Wilmington, NC, USA)*, AIMS Conference Publications, 113–122.
21. Shepelyansky DL (1997) Low-energy chaos in the Fermi-Pasta-Ulam Problem. *Nonlinearity* 10: 1331–1338.

22. Benettin G, Livi R, Ponno A (2009) The Fermi-Pasta-Ulam problem: scaling laws vs. initial conditions. *J Stat Phys* 135: 873–893.

23. Livi R, Pettini M, Ruffo S, et al. (1985) Equipartition threshold in nonlinear large Hamiltonian systems: The Fermi-Pasta-Ulam model. *Phys Rev A* 31: 1039–1045.

24. Gardner CS, Green JM, Kruskal MD (1967) Method for solving the Korteweg-de Vries equation. *Phys Rev Lett* 19: 1095–1097.

25. Lax PD (1968) Integrals of nonlinear equations of evolution and solitary waves. *Commun Pure Appl Math* 21: 467–490.

26. Miura RM, Gardner CS, Kruskal MD (1968) Korteweg-de Vries equation and generalization, II. Existence of conservation laws and constants of motion. *J Math Phys* 9: 1204–1209.

27. Zakharov VE, Feddeev LD (1971) Korteweg-de Vries equation: A completely integrable Hamiltonian system. *Funct Anal Appl* 5: 280–286.

28. Zakharov VE (1973) On stochastization of one dimensional chains of nonlinear oscillators. *Sov Phys JETP* 38: 108–110.

29. Toda M (1967) Vibration of a chain with nonlinear interaction. *J Phys Soc Jpn* 22: 431–436.

30. Toda M (1967) Wave propagation in anharmonic lattices. *J Phys Soc Jpn* 23: 501–506.

31. Toda M (1969) Mechanics and statistical mechanics of nonlinear chains. *J Phys Soc Jpn* 26: 109–111.

32. Toda M (1970) Waves in nonlinear lattice. *Prog Theor Phys* 45: 174–200.

33. Hénon M (1974) Integrals of the Toda lattice. *Phys Rev B* 9: 1921–1923.

34. Flaschka H (1974) The Toda lattice. II. existence of integrals. *Phys Rev B* 9: 1924–1925.

35. Cecchetto M (2015) *Normal modes and actions in the Toda Model*, Master thesis of University of Padua, Dept. of Mathematics "Tullio Levi-Civita".

36. Henrici A, Kappeler T (2008) Global action–angle variables for the periodic Toda lattice. *Int Math Res Not* 2008: 1–52.

37. Henrici A, Kappeler T (2008) Global Birkhoff coordinates for the periodic Toda lattice. *Nonlinearity* 21: 2731–2758.

38. Bambusi D, Maspero A (2016) Birkhoff coordinates for the Toda Lattice in the limit of infinitely many particles with an application to FPU. *J Funct Anal* 270: 1818–1887.

The Appendix: the procedure to compute the actions

We recall here the procedure to compute the actions proposed in [9]. Let us stress that the procedure is very clear in principle, but not so easy to implement in practice for large N. To understand, ref. [35] was very helpful to us.

A. *The theoretical frame.* Consider a periodic Toda chain with N particles, assume the barycenter is at rest. Denote

$$a_i = e^{\lambda(q_{i+1}-q_i)/2}\,, \qquad b_i = \lambda p_i\,,$$

so that

$$\prod_i a_i = 1\,, \qquad \sum_i b_i = 0\,,$$

and let

$$L^{\pm} = \begin{pmatrix} b_1 & a_1 & & & & \pm a_N \\ a_1 & b_2 & a_2 & & & \vdots \\ \vdots & a_2 & b_3 & a_3 & & \vdots \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & b_{N-1} & a_{N-1} \\ \pm a_N & & & & a_{N-1} & b_N \end{pmatrix} \quad ;$$

L^+ is the Lax matrix associated to the system, used to prove integrability in [8, 34], L^- is an auxiliary matrix. Following [9], consider the characteristic polynomials

$$P^{\pm}(x) = \det(x\, I - L^{\pm})$$

of $L^{\pm}$, which turn out to satisfy, for any x,

$$P^-(x) - P^+(x) = 4\,.$$

Though the matrices change during the evolution, the characteristic polynomials do not (this is indeed the essence of Lax method to prove integrability). In particular, the eigenvalues $\lambda_1^{\pm}, \dots, \lambda_N^{\pm}$ of $L^{\pm}$ stay constant, and for either choice of the sign they provide an independent set of N constants of motion. The eigenvalues will be supposed to be ordered decreasingly.

To define the actions it is convenient to introduce the discriminant Δ, defined as

$$\Delta(x) = P^+(x) + 2 = P^-(x) - 2\,,$$

which then satisfies

$$\Delta(\lambda_i^+) = 2, \qquad \Delta(\lambda_i^-) = -2\,.$$

We recall here some basic properties of Δ, for details and discussion see [9]. To be definite we assume N is even.

(a) At mechanical equilibrium ($a_i = 1$, $b_i = 0$), the shape of Δ is as in Figure 13, left panel, which refers to $N = 16$; Δ oscillates exactly between ± 2 and all eigenvalues $\lambda_i^{\pm}$, but the extremal ones λ_1^+ and λ_N^+, are double. All eigenvalues stay between[‖] ± 2.

(b) For nonvanishing energy, one or more pairs of degenerate eigenvalues $\lambda_k^{\pm}, \lambda_{k+1}^{\pm}$ open, producing a "gap", and correspondingly there are maxima of Δ exceeding 2 or minima exceeding -2. The shape of Δ looks as in the right panel of Figure 13, blue curve. Denote by $J_1, \dots, J_{N-1}$, right to left, the intervals between nearby eigenvalues:

$$J_k = \begin{cases} (\lambda_{k+1}^-, \lambda_k^-) & k \text{ odd} \\ (\lambda_{k+1}^+, \lambda_k^+) & k \text{ even}\,. \end{cases}$$

[‖]Between ± 1 in [9]; for a certain convenience, our notation differs from [9] by a few sparse factors 2.

To each nonvanishing interval J_k corresponds a positive action $\mathcal{I}_k$, defined by

$$\mathcal{I}_k = \frac{1}{\pi} \int_{J_k} \rho(x)\, \mathrm{d}x\,, \qquad \rho(x) = \operatorname{acosh} \frac{|\Delta(x)|}{2}\,.$$

Following [9], one can trace inside each open gap an arc $\rho'(x) = \pm(2 + \rho(x))$, see the red curves of Figure 13. The actions are precisely, up to a trivial factor, the areas between $\rho'(x)$ and the lines ± 2. Notice that if $|\Delta|$ is large, then $\rho \simeq \log |\Delta|$.

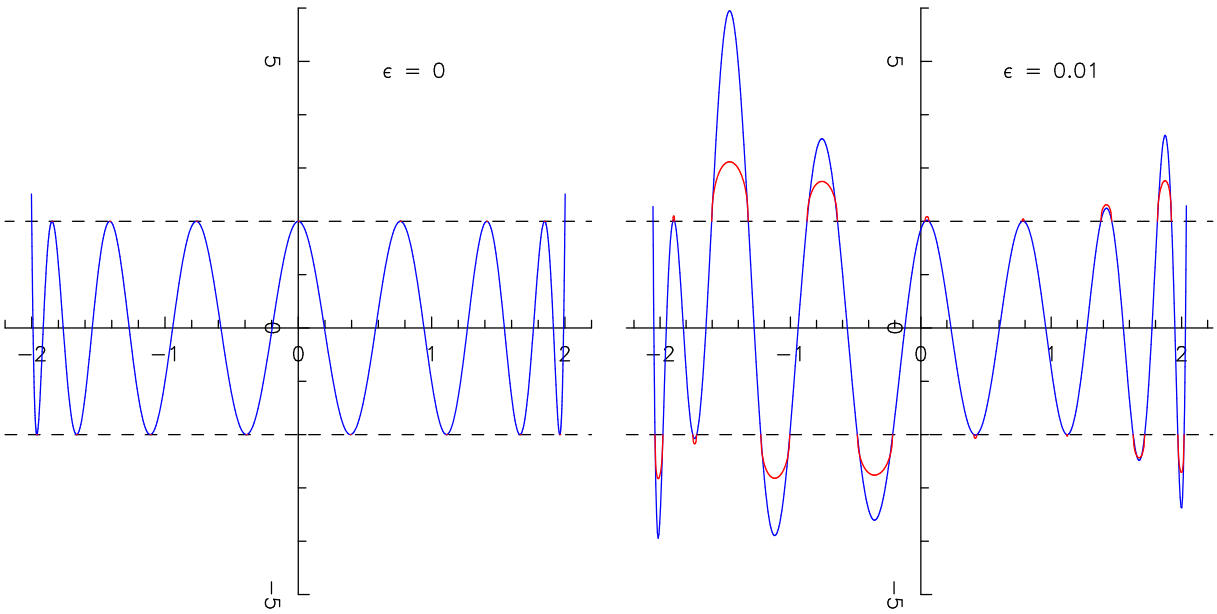

Figure 13. Left: the discriminant for $N = 16$, at rest. Right: the discriminant (blue) and the arcs $\pm(2 + \rho)$ (red), for $\varepsilon = 10^{-2}$; several traveling waves are present.

B. *An easy example.* To exemplify the computation in an easy case, consider a state with a single sinusoidal wave with not small k. Figure 14 refers to $N = 32$, $k = 12 \simeq 0.4N$, ε in the range $10^{-4} - 10^{-1}$. Quite evidently, the action $\mathcal{I}_{12}$ is dominant, although it is clear that by increasing ε other actions start to be appreciable. The peak of Δ, already for $\varepsilon = 0.01$, is large, approximately 130, and grows to approximately 4×10^6 for $\varepsilon = 0.1$.

By increasing N the peaks of Δ grow explosively, and this is the main problem in the numerical computation of the actions. Table 1 shows the peaks for some values of N up to 1024, ε between 10^{-4} and 10^{-1}. High peaks should not surprise: in a state with a single traveling wave it is $I_k = E_k/\omega_k = \varepsilon N/\omega_k$; should it be $\mathcal{I}_k \simeq I_k$, then (assuming the gap, i.e., the base of the lobe formed by $\rho'(x)$, is large) ρ is of order εN and Δ is its exponential.**

**In [9], when peaks as large as 10^{15} are met, the comment is: *The value of the discriminant at its maximum* (10^{15}) *is entirely typical, and perhaps surprisingly large for someone who has done inverse spectral theory with paper and pencil only.*

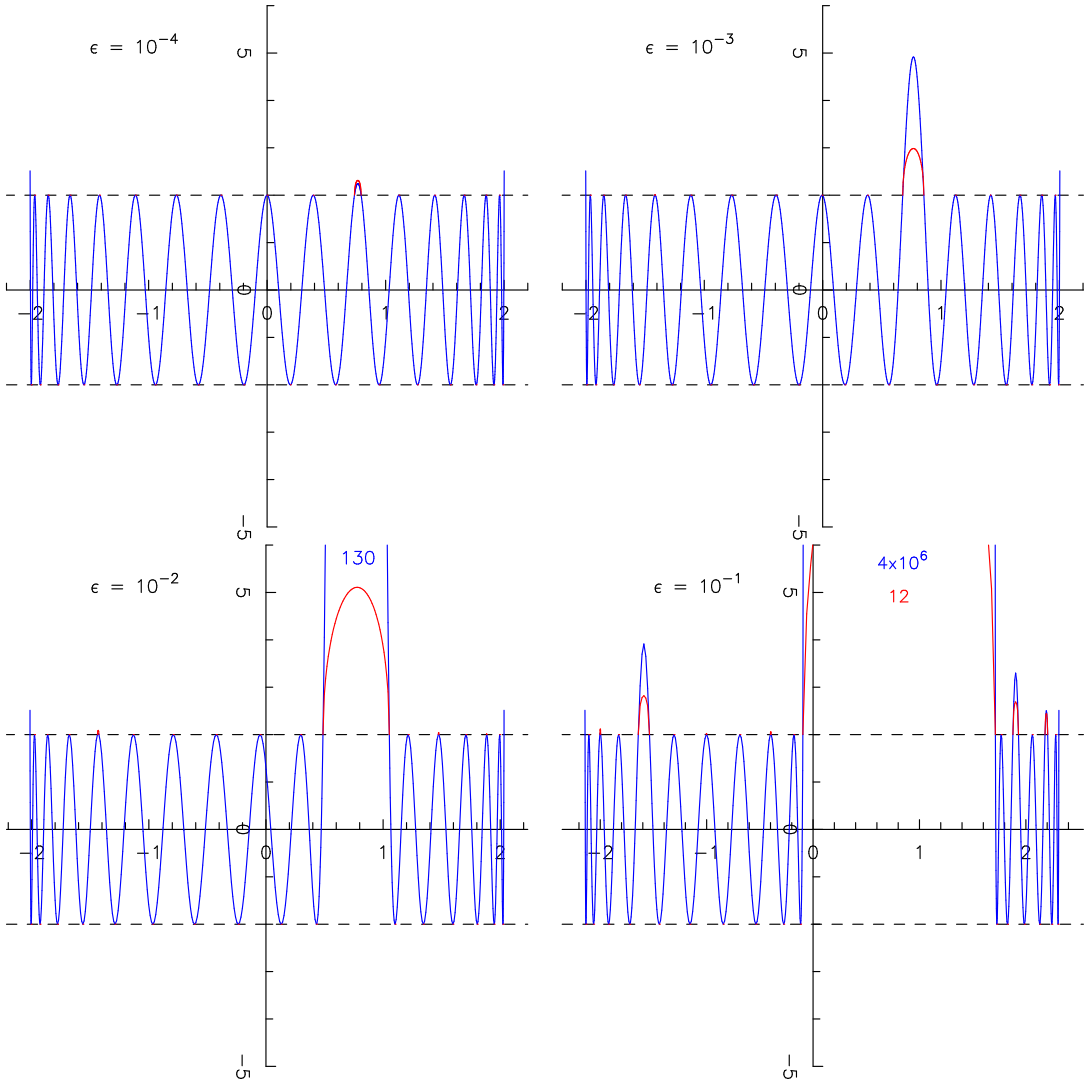

Figure 14. The discriminant (blue) and the arcs $\pm(2+\rho)$ (red) for $N = 32$, $\varepsilon = 10^{-4}$, 10^{-3}, 10^{-2}, 10^{-1}; single traveling wave, $k = 12$.

Table 1. *The maximum of* $|\Delta|$ *for different values of* N *and* ε*; single left traveling wave,* $k \simeq 0.4N$.

	$\varepsilon = 10^{-4}$	10^{-3}	10^{-2}	10^{-1}
$N = 32$	2.5	4.9	1.3×10^2	3.5×10^6
64	3.0	21	1.4×10^4	7.3×10^{12}
128	6.9	4.1×10^2	1.8×10^8	3.4×10^{25}
256	45	1.5×10^5	3.2×10^{16}	1.2×10^{51}
512	2.0×10^3	2.9×10^{10}	1.0×10^{33}	1.4×10^{102}
1024	4.1×10^6	8.3×10^{20}	9.3×10^{65}	1.2×10^{204}

C. *Technical questions.* Problems of theoretical precision, in presence of such large numbers, are serious. The results reported in Section 3 should be consider as highly tentative, subject to criticism and justified mainly by their overall coherence.

All computations have been made in the so-called quadruple precision, that is representing numbers with approximately 33 decimal digits. This is fairly enough for small N, but questions arise for, say, $N = 1024$ or larger, unless ε is very small.

To compute the actions one needs to compute with the best possible precision the eigenvalues $\lambda_i^{\pm}$. Since this is done only once per each state, computer time can be invested. The procedure we found convenient is:

- Using standard routines, namely routines (in double precision, 16 decimal digits) designed for sparse symmetric matrices, to compute the eigenvalues of $L^{\pm}$. This provides a first approximation of $\lambda_i^{\pm}$, with errors of the order, say, 10^{-14}.
- Improving the precision by searching the eigenvalues as zeros of $P^{\pm}(x)$, with the Newton method. Easy recursive expressions are available for the determinant of three-diagonal periodic matrices, that is for $P^{\pm}(x)$, as well as for the derivative. A little problem arises when the zeros of $P^{\pm}(x)$ are tangent or nearly tangent, but this is solved by approximating $P^{\pm}(x)$ with a parabola, and moreover, the corresponding gaps J_k and actions $\mathcal{I}_k$ in this case are small, so one should not worry too much. The serious numerical problem instead is the usual loss of precision, when a zero results from a huge sum of large numbers. Here the only way out seems to us to be the use of multiprecision arithmetics, as we plan to do in a nearby future.

Once the $\lambda_i^{\pm}$ are obtained, the characteristic polynomials can be computed in the standard way,

$$P^{\pm}(x) = \textstyle\prod_i (x - \lambda_i^{\pm}) \, ;$$

$\Delta(x)$, $\rho(x)$ and the actions are immediately obtained. When $|P^{\pm}(x)|$ starts to be large, one can directly compute $\rho(x) \simeq \log|\Delta(x)| \simeq \log|P^{\pm}(x)|$,

$$\rho(x) \simeq \textstyle\sum_i \log|x - \lambda_i^{\pm}| \, .$$

Everything was somehow unusual in our experience, and a little painful .

AIMS Press