The Fermi–Pasta–Ulam problem and its underlying integrable dynamics

G. Benettin*, H. Christodoulidi†, and A. Ponno‡

Università degli Studi di Padova, Dipartimento di Matematica,
Via Trieste 63, 35121 Padova (Italy)

Abstract

This paper is devoted to a numerical study of the familiar $\alpha+\beta$ FPU model. Precisely, we here discuss, revisit and combine together two main ideas on the subject: (i) In the system, at small specific energy $\varepsilon = E/N$, two well separated time–scales are present: in the former one a kind of metastable state is produced, while in the second much larger one, such an intermediate state evolves and reaches statistical equilibrium. (ii) FPU should be interpreted as a perturbed Toda model, rather than (as is typical) as a linear model perturbed by nonlinear terms. In the view we here present and support, the former time scale is the one in which FPU is essentially integrable, its dynamics being almost indistinguishable from the Toda dynamics: the Toda actions stay constant for FPU too (while the usual linear normal modes do not), the angles fill their almost invariant torus, and nothing else happens. The second time scale is instead the one in which the Toda actions significantly evolve, and statistical equilibrium is possible. We study both FPU–like initial states, in which only a few degrees of freedom are excited, and generic initial states extracted randomly from an (approximated) microcanonical distribution. The study is based on a close comparison between the behavior of FPU and Toda in various situations. The main technical novelty is the study of the correlation functions of the Toda constants of motion in the FPU dynamics; such a study allows us to provide a good definition of the equilibrium time τ, i.e. of the second time scale, for generic initial data. Our investigation shows that τ is stable in the thermodynamic limit, i.e. the limit of large N at fixed ε, and that by reducing ε (ideally, the temperature), τ approximately grows following a power law $\tau \sim \varepsilon^{-a}$, with $a = 5/2$.

*E-mail: benettin@math.unipd.it

†Present address: University of Patras, Department of Mathematics, GR-26500, Patras, Greece; e-mail: hchrist@master.math.upatras.gr

‡E-mail: ponno@math.unipd.it

1 Introduction

1.1 Our purpose

This paper is devoted to the Fermi–Pasta–Ulam (FPU) model [1], more precisely to the so–called "α+β" model with Hamiltonian

$$H(p,q) = \frac{1}{2}\sum_{i=1}^{N} p_i^2 + \sum_{i=0}^{N} V(q_{i+1}-q_i) \ , \qquad q_0 = q_{N+1} = 0 \ , \tag{1}$$

where

$$V(r) = \frac{r^2}{2} + \alpha\frac{r^3}{3} + \beta\frac{r^4}{4} \ . \tag{2}$$

The nonlinearity depends of course on α and β, but also on the energy; more precisely, as is shown by elementary scaling arguments, it is given by the two quantities

$$\alpha\sqrt{\varepsilon} \ , \qquad \beta\varepsilon \ ,$$

$\varepsilon = E/N$ being the energy per degree of freedom. The value of α (if different from zero) is irrelevant and simply fixes the energy scale; throughout the paper we shall set $\alpha = 1$. Due to nonlinearity, energy is expected, and was expected by FPU,[1] to circulate freely among normal modes, in agreement with the principle of equipartition of energy of statistical mechanics. As is well known, the paradox discovered by FPU is the fact that, for small nonlinearity, energy sharing among all modes does not occur at all: in particular, if energy is initially given to the lowest frequency mode (as they did), for long times—the longest accessible in those years—only a few modes do share it; moreover the dynamics, contrary to expectation, appeared to be not mixing but quasi–periodic (a striking evidence of quasi periodicity was later produced in [2]).

Dozens and dozens of papers have been written on the subject, after the original FPU paper. Quite different ideas, methods and purposes have been pursued in a variety of papers, among which it is hard to extract any order (a partial account of such abundant literature can be found in [3, 4]; the variety of perspectives, not fitting a single picture and occasionally contradictory, is striking).

In this paper we shall reconsider, combine together and illustrate two main ideas: old ones—actually going back to the early 80's of last century—but, in our opinion, simple and clear, actually the cornerstones to understand FPU:

a) For small ε, two well separated time scales enter the problem. In the shorter one—the only one observed by FPU—the energy, if initially given to one or a few modes, is practically shared only by a small subset of them; in the second much larger one, energy equipartition among all modes is reached. The two time scales have been introduced in [5, 6]; their presence and relevance has been stressed in a sequence of papers by the Milano team, starting from [7, 8] (see, for a review, [9] in [4]).

b) The phenomenon observed by FPU (the partial energy sharing) is essentially integrability, *the reference integrable model being not the linear chain, but the Toda model.* The idea that integrability explains in some way the FPU paradox goes back to the pioneering paper [10], where the connection between FPU and the KdV equation—limited to states where only long–wavelength modes are excited—was first established. After the discovery of the integrability

[1] From ref. [1]: ... *It is, therefore, very hard to observe the rate of 'thermalization' or mixing in our problem, and this was the initial purpose of the calculation.*

of the Toda model [11, 12], the strong connection between the dynamics of FPU and Toda was put in evidence in [13]. Unfortunately, such a fundamental paper was almost forgotten in the subsequent literature, although with interesting exceptions like [14, 15, 16, 17, 18]. The present paper is deeply influenced by [13].

The Toda model, we recall, has potential

$$V_T(r) = V_0(e^{\lambda r} - 1 - \lambda r) \ , \qquad V_0,\, \lambda \quad \text{free parameters}\,. \tag{3}$$

For the particular choice

$$V_0 = \tfrac{1}{4}\alpha^{-2} \ , \qquad \lambda = 2\alpha \ ,$$

the Toda potential is tangent to the FPU one up to third order, namely

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

with

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

Roughly speaking, the distance between the FPU model and the linear model is $\alpha\sqrt{\varepsilon}$, while the distance between FPU and Toda is $|\beta - \beta_T|\varepsilon$, much smaller at low ε.

The Toda model is integrable and so there exist action–angle variables (I, φ), such that the actions stay constant while the angles advance linearly in time and (generically) fill an N–dimensional invariant torus. The view that will progressively emerge in the paper is very simple: the short time scale is the one in which the FPU and Toda model are almost indistinguishable; this means that in FPU too the Toda actions stay almost constant and only the angles evolve, producing in this way the partial quasiperiodic energy sharing occurring in the FPU dynamics. The larger time scale instead is the one in which the Toda actions do evolve, and statistical equilibrium gets eventually possible. Unfortunately, the transformation to action–angle variables is not explicitly known, and so we can produce only an indirect evidence of the above picture, although, we hope, a rather convincing one.

In the remaining part of this Introduction we shall preliminarily illustrate the above ideas by revisiting, with improvements, some results we produced in previous papers [19, 20]. Such results concern FPU–like initial data, namely initial data very far from statistical equilibrium in which only a small fraction of the normal modes is initially excited. Sections 2 and 3 deal instead with generic initial data in the sense of the microcanonical measure, and contain genuine new results. Till Section 2, as in most papers on FPU, we shall use the linear normal modes as the basic variables to be analyzed. In Section 3 instead we shall study (in the FPU dynamics) the Toda constants of motion, which are functions of the actions and, at variance with the actions, are explicit. A study of the first nontrivial Toda integral in FPU has been performed in [20, 21].

1.2 An evidence of the two time scales

Let

$$Q_k = \sqrt{\frac{2}{N+1}} \sum_{i=1}^{N} q_i \, \sin\frac{\pi k i}{N+1} \ , \qquad P_k = \sqrt{\frac{2}{N+1}} \sum_{i=1}^{N} p_i \, \sin\frac{\pi k i}{N+1} \ , \tag{4}$$

$k = 1, \ldots, N$, denote the usual harmonic normal modes; their energy E_k and frequency ω_k are respectively

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

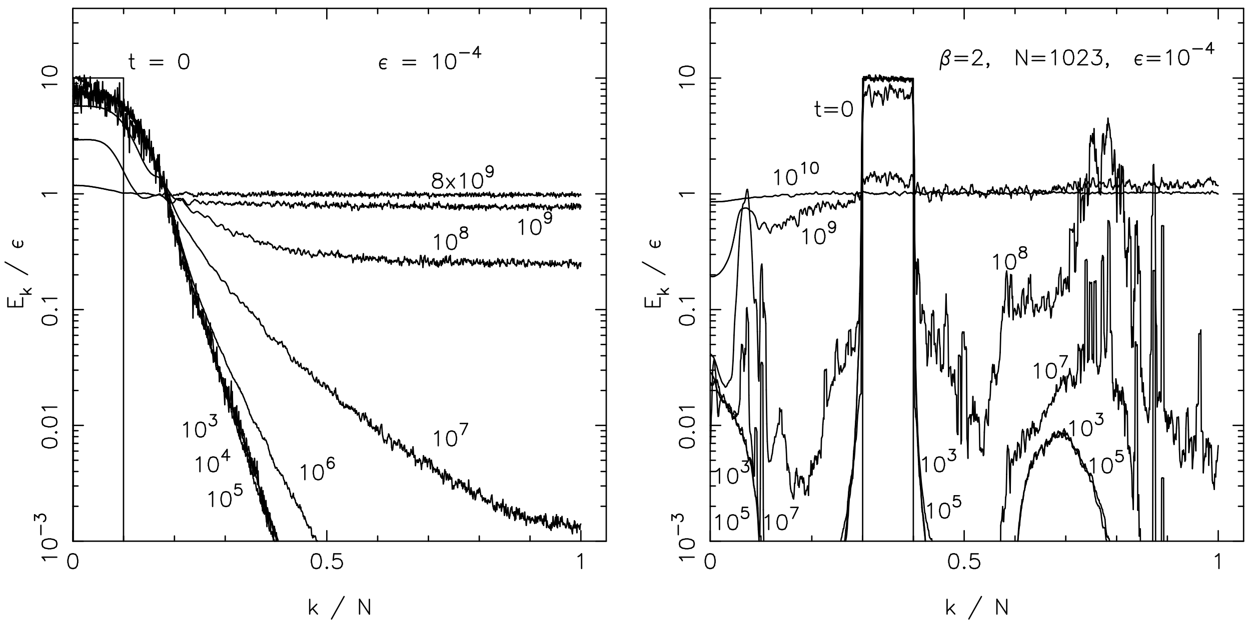


Figure 1: *The shape of the averaged energy spectrum of normal modes* $\overline{E}_k(T)$ *plotted vs.* k/N*, at selected times* T *(marked in the figure) in geometric progression. Energy initially equidistributed among modes with* $0 < k/N < 0.1$ *(left) and with* $0.3 < k/N < 0.4$ *(right), see the rectangle marked* $t = 0$*. Each point is the average over 24 random extractions of the initial phases. Parameters:* $N = 1023$*,* $\alpha = 1$*,* $\beta = 2$*,* $\varepsilon = 10^{-4}$*.*

According to the principle of energy equipartition, one expects

$$\overline{E}_k(T) \stackrel{T\to\infty}{\longrightarrow} \langle E_k \rangle \simeq \varepsilon \ , \tag{5}$$

where $\overline{E}_k(T)$ denotes the time average of E_k up to time T. As a minor improvement, we found convenient to use in numerical computations, in place of the familiar time averages starting from $t = 0$, averages over the last part of the trajectory, more precisely in the running window $\frac{2}{3}T \le t \le T$:

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

similar averages have the same asymptotics as the averages from $t = 0$, but are more prompt to follow the evolution of the system and reveal the asymptotics earlier. (The choice of $\frac{2}{3}T$ as the lower extreme of the interval is fairly irrelevant; $\frac{1}{2}T$ is good as well.)

Figure 1, left panel, shows the shape of the energy spectrum, more precisely $\overline{E}_k(T)/\varepsilon$ vs. k/N (logarithmic vertical axis), at different values of T in geometric progression. The figure refers to a model with $N = 1023$, $\beta = 2$, and rather small $\varepsilon = 10^{-4}$; energy was initially equidistributed among the lowest 10% of modes (see the rectangular profile in the figure). The figure shows 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 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, definitely much larger than the time needed to form it; only on *much* larger times, say $T = 10^9$ or 10^{10}, the system evolves towards energy equipartition, the high–frequency modes being progressively involved into the energy sharing. If the phases of the initially excited modes are chosen randomly (this is crucial, see [22]), then the behavior, for large N at fixed small ε, is independent of N, i.e. it persists in the thermodynamic limit. The initial phases, for the above numerical computation, were in fact chosen randomly; moreover, to improve a little the figure and reduce the fluctuations, a further average of $E_k(T)$ on 24 different random choices of the phases was introduced.

The right panel of figure 1 shows the same phenomenon for a different initial excitation, namely energy equidistributed among modes with $0.3 < k/N < 0.4$; the intermediate state is quite different and more complicated—two bumps of modes around $k = 0$ and k/N doubled ($0.6 < k/N < 0.8$) are produced—but the essence of the phenomenon is the same: in the short time scale $T = 10^3$ some intermediate state is formed; on the much larger time scale 10^9 or 10^{10} equipartition is reached.

Let us notice that observing the second time scale, with the evolution to equipartition, requires a considerable computational power. In the original FPU paper, and later on for decades, only the intermediate state could be observed; see the remarks at the end of Section 3 for further comments.

1.3 A first evidence of the underlying Toda dynamics

An easy way to get convinced that on the short time scale FPU behaves as an integrable system, the underlying dynamics being the Toda dynamics, is to repeat the computations leading to figure 1 using, in place of FPU, the tangent Toda model. Figure 2, upper panels, shows the result; for an easier comparison, the lower panels show again the behavior of FPU, but only up to $T = 10^5$, before the tail appreciably raises. Quite clearly, on the short time scale FPU and Toda are hardly distinguishable, and get progressively different only on larger times, when FPU slowly evolves towards energy equipartition while Toda does not evolve at all.

Remark: as is well known, when only long waves are significantly excited, all chains of oscillators with leading cubic nonlinearity, including both FPU and Toda, are well modeled by the KdV wave equation [10, 23, 22]; in [22] in particular, using KdV as a normal form for such models, it was possible to understand in detail the formation of the intermediate state (shape, formation time, dependence on ε, on the width of the initial excitation, on the initial phases, with exact correspondence between analytic and numeric results). In particular, KdV explains the left panels of figures 1 and 2, in the short time scale. *The similarity of the right panels of the figures shows, however, that the similarity of FPU and Toda persists also when the standard continuum approximation does not hold any more and the use of KdV is not legitimate.*

Besides the spectrum, even looking at the instantaneous values of the mode energies E_k without averaging, the similarity between the FPU and the Toda dynamics clearly appears. Figure 3 shows the behaviour of a selected number of modes ($k = 1$, 30, 65, 100) up to $t = 10^4$, both for the FPU model (left) and for the tangent Toda model (right); same parameters and initial conditions as in figure 2. The similarity is rather impressive. For smaller models, for example for $N = 32$ as in the original FPU study, the similarity is even more precise, see figure 4. Let us stress that such a striking similarity had been noticed already in [13]: figure 4 is indeed a remake of figures 25, 26 of [13].

1.4 FPU–like initial data: selected results on the large time–scale

In this section we shortly recall a few results from our previous paper [19]. Results concern the second time scale for FPU–like initial data, more precisely initial data in which only low–frequency modes, with $0 < k/N < 0.1$, are equally excited, with random initial phases. The signature of the underlying Toda dynamics appears there clearly, and such results will be useful, for comparison, when we shall later investigate generic equilibrium initial data.

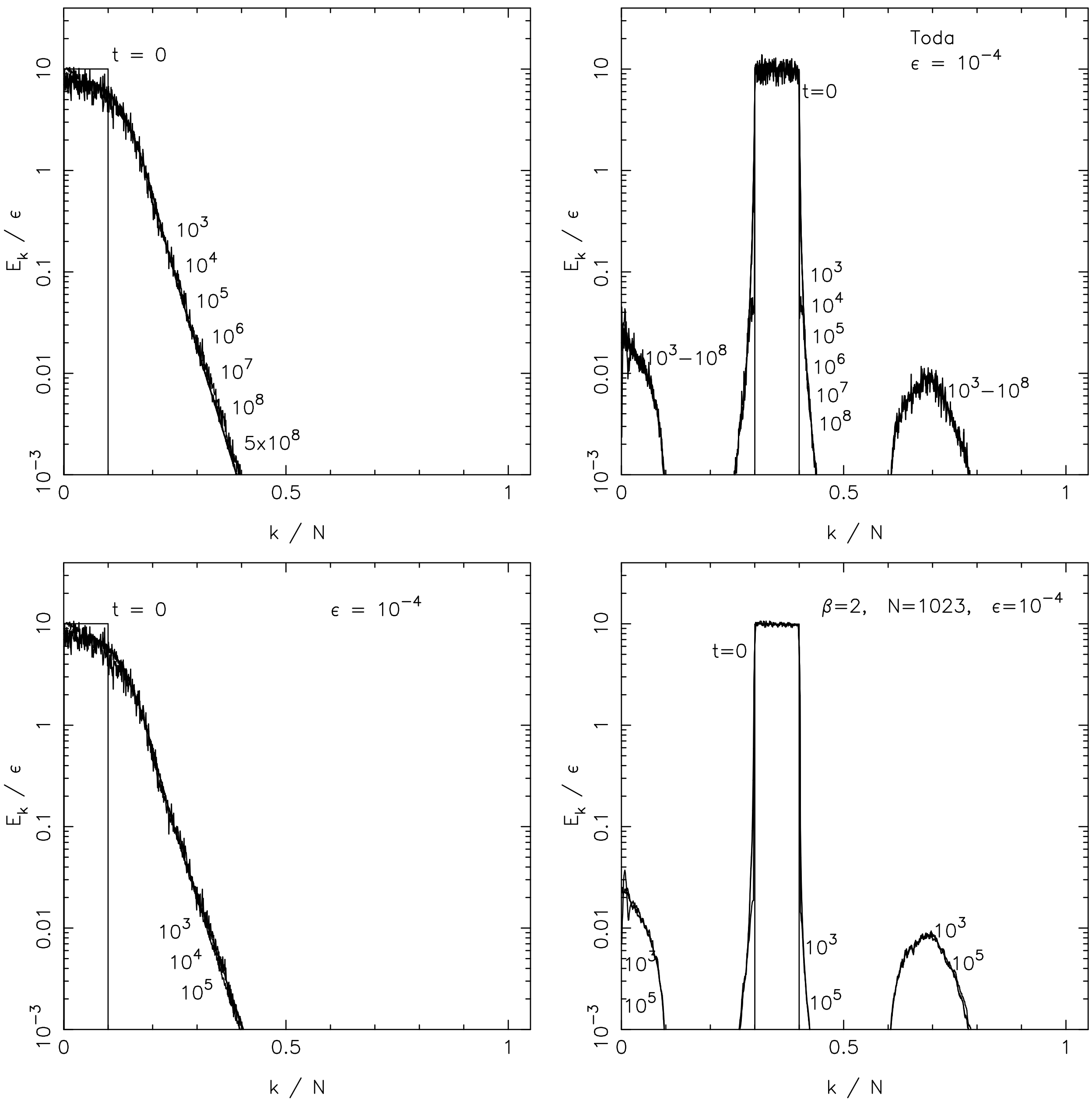


Figure 2: *Upper panels: same as in figure 1, but for the tangent Toda model (same N, ε). Lower panels: for comparison, the FPU model as in figure 1, limited to the short time scale $T = 10^5$.*

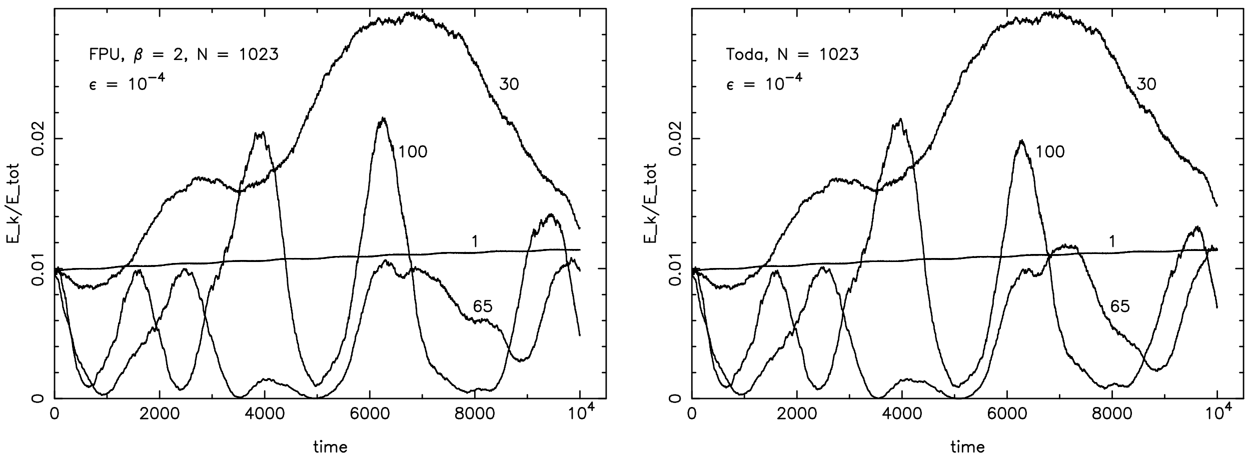


Figure 3: *The instantaneous values $E_k(t)$ vs. t, up to $t = 10^4$, for a selected set of modes. Left: the FPU model; right: the tangent Toda model. Parameters: $N = 1023$, $\alpha = 1$, $\varepsilon = 10^{-4}$ and (for FPU) $\beta = 2$. Initial data as in figure 1–left.*

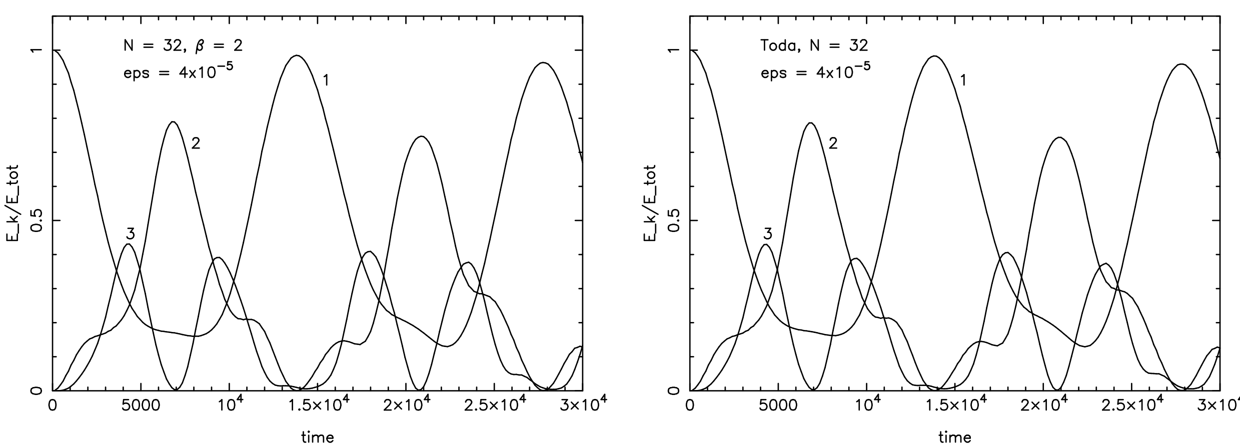


Figure 4: *The instantaneous values $E_k(t)$ vs. t, $k = 1, 2, 3$, up to $t = 10^4$, for smaller FPU (left) and Toda (right) models; $N = 32$ as in the original FPU paper, $\alpha = 1$, $\varepsilon = 4 \cdot 10^{-5}$ and (for FPU) $\beta = 2$. Only mode $k = 1$ initially excited.*

In order to characterize the second time scale, i.e. the growth of the tail, the quantity studied in [19] is (up to a minor correction, for which we defer to [19])

$$\mathcal{E}(t) = \frac{\text{energy of the tail}}{\text{due amount}} = \frac{\sum_{k>N/2} \overline{E}_k(t)}{\frac{1}{2}\sum_{k=1}^{N} \overline{E}_k(t)} \ .$$

As a matter of fact $\mathcal{E}(t)$, on a sufficiently large time scale, grows from 0 to 1, with a sigmoidal profile. Pragmatically, the long time scale $T(N,\varepsilon,\beta)$ is defined as the time at which $\mathcal{E}(t) = 0.5$. A sequence of accurate "experiments" shows that $T(N,\varepsilon,\beta)$ behaves as follows:

i) For large N at fixed ε and β, $T(N,\varepsilon,\beta)$ converges to a limit curve $T_\infty(\varepsilon,\beta)$, ideally representing the behavior of the system in the thermodynamic limit.[2]

ii) For each β different from the Toda value $\beta_T = \frac{2}{3}\alpha^2$, $T_\infty(\varepsilon,\beta)$ grows, by reducing ε, as a power of $1/\varepsilon$:
$$T_\infty(\varepsilon,\beta) \simeq C(\beta)\,\varepsilon^{-a} \ , \qquad a = \tfrac{9}{4} \ . \tag{6}$$
Changing $\beta \neq \beta_T$ changes the constant in front but not the exponent a.

iii) For $\beta = \beta_T$, i.e. when the FPU and Toda Hamiltonians get tangent at order four, the exponent of the power law (6) raises to $a = 3$. Adding terms of order five to the FPU Hamiltonian, as long as their coefficient γ is different from the Toda value $\gamma_T = \frac{1}{3}\alpha^3$, does not change a, while for $\gamma = \gamma_T$, i.e. by further raising the order of tangency, the exponent further grows to $a = 4$, and likely the process goes on. So, *FPU models appear to be divided into "universality classes", according to their tangency to Toda.*

iv) The constant $C(\beta)$ depends on β approximately as[3]
$$C(\beta) \sim (\beta - \beta_T)^{-2} \ . \tag{7}$$

For figures illustrating properties (i)–(iv) and for further comments we defer to [19].

Defining $T(N,\varepsilon,\beta)$ by looking only at growth of the tail is obviously open to criticism; other definitions are possible. But the main reason of criticism is that all of this concerns extremely particular regions of the phase space: even when $\mathcal{E}(t)$ reaches the threshold value 0.5, the system is still very, very far from statistical equilibrium. Most papers on FPU, starting from the original one, suffer of a similar limitation. As we shall show in the rest of the paper, defining and investigating the large time scale around equilibrium is not trivial, but something can be done.

2 Generic initial data: the short time scale

2.1 A result by the Milano team

With deep intuition, a few years ago our colleagues in Milano studied [24] the FPU phenomenon for generic initial data—i.e. data extracted randomly according to the microcanonical measure—by

[2] The limit is delicate, and approaching it at low ε requires larger and larger N, see [19]; for example, for $\varepsilon = 10^{-4}$ the correct asymptotics appears for, say, $N = 1023$ or larger, while $N = 255$ is sufficient at $\varepsilon = 10^{-3}$. The two limits $N \to \infty$ and $\varepsilon \to 0$ do not commute at all.

[3] Of course the divergence stops for β so close to β_T, that the terms of order five, different in FPU and in Toda, become dominant; see [19].

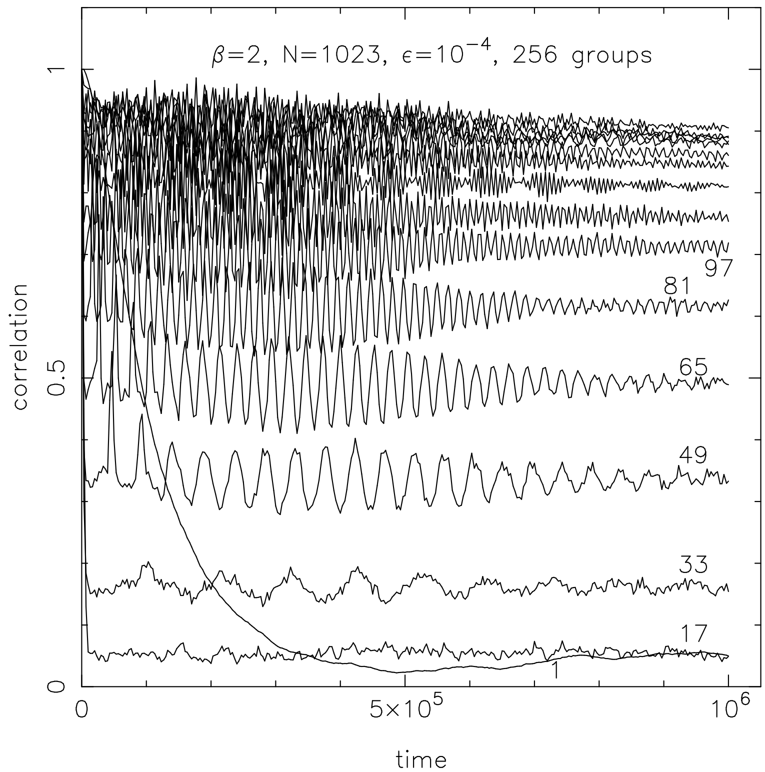


Figure 5: *The group correlation function* g_{e_j} *vs.* t*, for selected groups of modes:* $N = 1023$*,* $n = 256$*,* $j = 1, 17, 33, 49, ..., 256$*; FPU model with* $\alpha = 1$*,* $\beta = 2$*,* $\varepsilon = 10^{-4}$*.*

looking not at the time averages of the energies of the modes, which by construction are already in equipartition and statistically do not evolve any more, but at their time autocorrelations. The purpose is to distinguish, roughly, between modes that—within a given observation time t^*, interpreted as the short time scale—significantly exchange energy with other modes and decorrelate, and modes that instead, in the same time interval, keep almost unchanged their initial energy. Such a distinction corresponds, in the view of the authors (that we share), to the distinction observed by FPU between modes that, starting from their nonequilibrium initial condition, rapidly share energy with other modes, thus forming the intermediate state (figure 1), and modes that do not. We decided to revisit the computations of [24] to improve their accuracy (mainly through a larger statistics that cleans curves) as well as to investigate the relation between FPU and Toda for generic initial conditions, too.

For a generic observable f, let

$$G_f(t) = \langle f(t)f(0)\rangle - \langle f\rangle^2 \ ,$$

where $f(t)$ is an abbreviation for $f(p(t), q(t))$ and $\langle\,.\,\rangle$ denotes (with some approximation, see below) microcanonical averaging on initial data; the normalized correlation function of the observable is then defined, as usual, as

$$g_f(t) = G_f(t)/G_f(0) \ .$$

The authors divide the N normal modes into a number n of groups (to be kept fixed when N increases, in view of the thermodynamic limit), and choose, as the observables to look at, the energies e_j of the groups, $j = 1, \ldots, n$, defined as is obvious as the sum of the energies of the modes composing the group.

Figure 5 shows the behavior of g_{e_j}, limited to a few selected j ($j = 1, 17, 33, \ldots$, bottom to top), for a system of $N = 1023$ particles, at $\varepsilon = 10^{-4}$, modes being divided into $n = 256$ groups of 4 modes (3 in the last group). The figure shows that groups of modes with low j (bottom part of the figure) rather rapidly decorrelate, reaching a *plateau*; groups with higher j reach, after some oscillations, a higher plateau; groups with j close to n remain instead, within the time–scale $t = 10^6$ of the figure, strongly correlated.

Just a comment on the measure $\langle\,.\,\rangle$. The procedure used in [24], that we literally followed, is to neglect the nonlinear terms in the Hamiltonian, and to extract randomly, with Gaussian distribution, the coordinates Q_k, P_k of the normal modes; the coordinates are then rescaled, so as the model attains the prefixed specific energy ε. The approximation of the Gibbs measure by a Gaussian measure, ignoring nonlinear terms, is open to criticism in the thermodynamic limit (no matter how small is ε), but it is hard to do better; results confirm the validity of the approximation (even in the case, studied in [24], of $\beta = 0$ and, correspondingly, of non compact constant energy surfaces). Our statistics is (at least) $24,000$ random extractions ($10,000$ to $2,000$, depending on N, in [24]).

Let $g^*_{e_j}$ denote the value of the *plateau* attained by g_{e_j}, $j = 1, \ldots, n$; $g^*_{e_j}$ is pragmatically defined as the average of $g_{e_j}(t)$ in the time interval $(2/3) \times 10^6 \leq t \leq 10^6$ (such an averaging, not present in [24], considerably cleans the results). The bold curve in figure 6, upper left panel, shows $g^*_{e_j}$ vs. j/n in the same conditions as in figure 5, namely $N = 1023$, $\beta = 2$, $\varepsilon = 10^{-4}$. The other curves in the figure show the result at different N, namely, top to bottom, $N = 511$, 1023, 2047, 4095, 8191. It is rather evident that, as guessed in [24] although with poorer data, a sigmoidal limit curve for large N does exist. Such a curve synthesizes the difference between low and high modes in the FPU dynamics for generic initial data: (groups of) low modes do exchange energy in a short time, (groups of) high modes do not or do less. This is coherent with the information deduced from the spectrum appearing in figure 1; the great novelty, with respect to figure 1 and more generally with respect to traditional studies on FPU, is that generic initial states are now considered.

Let us now come to the Toda model. All of the previous computations have been repeated with the tangent Toda potential; the result is in figure 6, upper right panel. The curves of Toda and FPU are clearly very similar, although a little difference can be observed in the high frequencies: the correlations are a little lower in FPU than in Toda. For a better comparison, we investigated the β dependence of the above curves, restricting the attention to the cheaper case $N = 511$. The result is in the lower panels of figure 6; the right panel is just a zoom of the left one. Toda is in bold. The quantity Δ appearing in the figure is the relative difference between β and β_T, namely

$$\Delta = \frac{\beta - \beta_T}{\beta_T}\ . \tag{8}$$

For larger $|\Delta|$, the difference between FPU and Toda increases, and is revealed by the high modes, rather than by the low modes.

A few comments are in order:

– Even for generic initial data, the intermediate state of FPU seems to be the same as in Toda. This means that the large energy exchanges exhibited by low modes, which produce the decay of their correlations, are completely regular, and can be hardly interpreted as a partial thermalization. The natural conjecture (as anticipated in the Introduction) is that within the considered time scale, in both models, only the Toda angles advance and produce

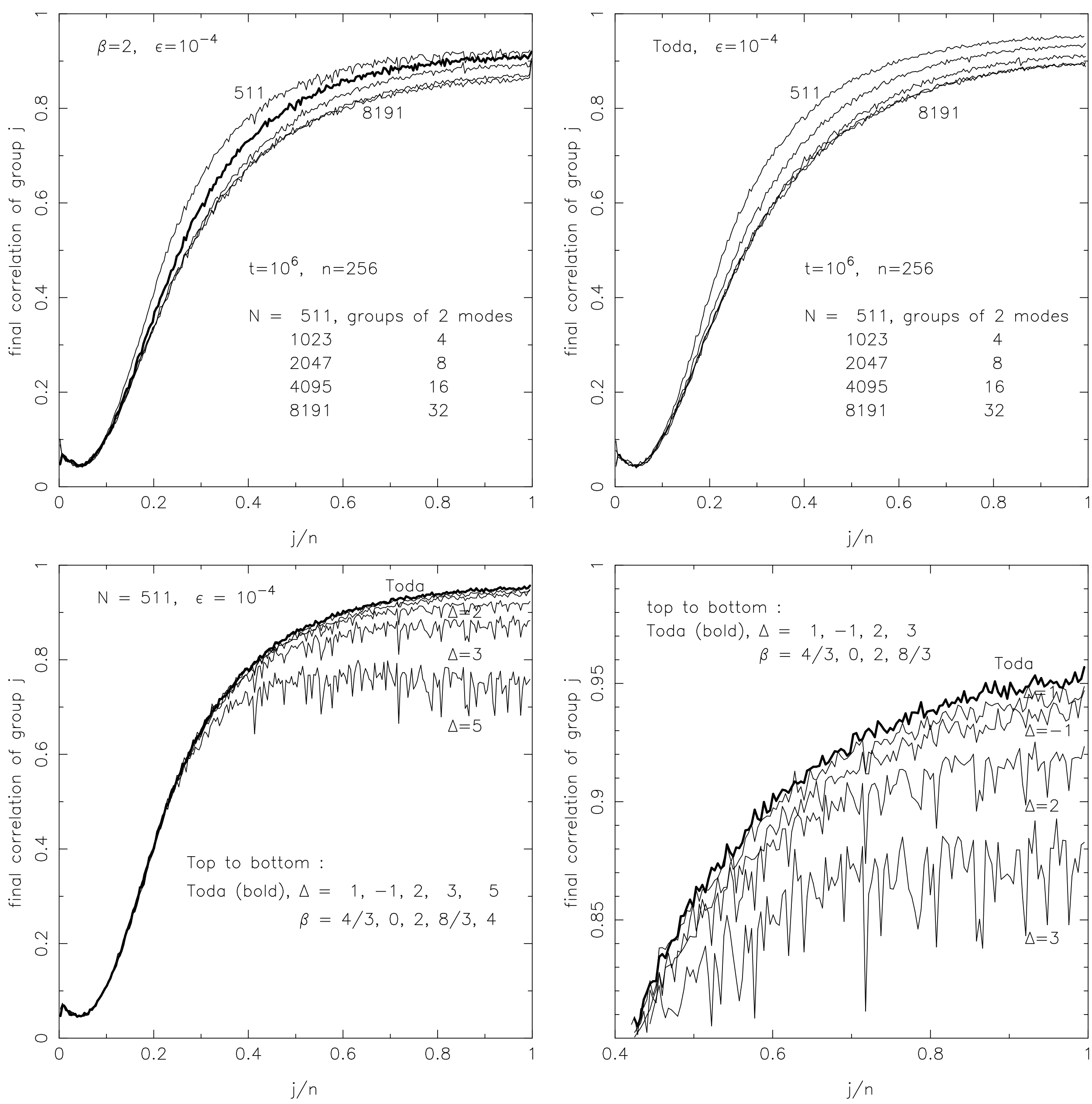


Figure 6: *Upper left panel: the plateau value* $g^*_{e_j}$ *vs.* j/n*, for a FPU model with* $\alpha = 1$, $\beta = 2$, $\varepsilon = 10^{-4}$*, and* $N = 511, 1023, 2047, 4095, 8191$*; the bold curve refers to* $N = 1023$ *and corresponds to figure 5. Upper right panel: the same, for the tangent Toda model. Lower left panel: a comparison between Toda (the upper curve) and FPU models with different values of* $\Delta = (\beta - \beta_T)/\beta_T$*, namely, top to bottom,* $\Delta = -1, 1, 2, 3, 5$ $(\beta = 0, 4/3, 2, 8/3, 4)$*. The lower right panel is a zoom.*

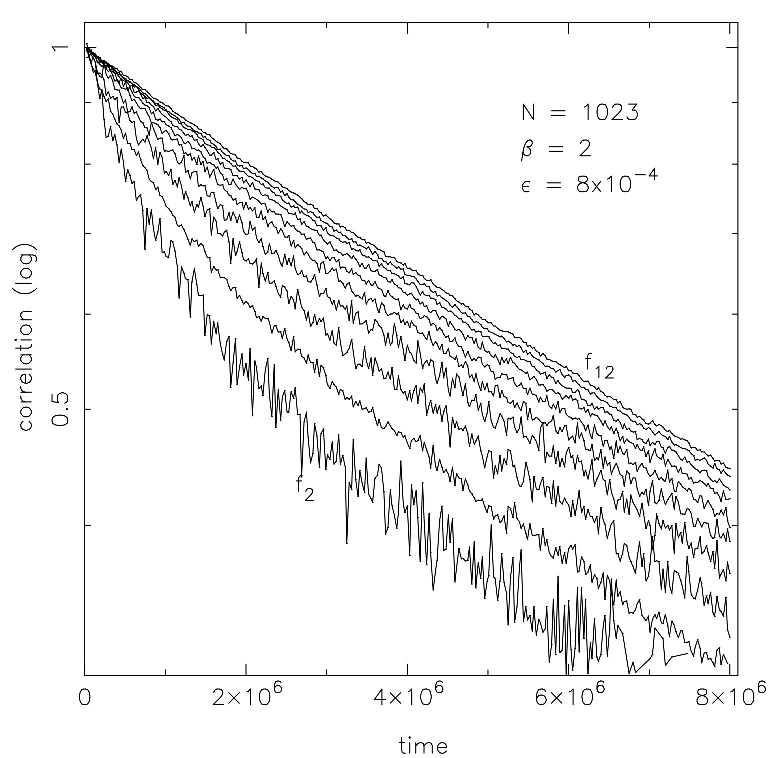


Figure 7: *The decay of the correlation functions $g_1, \ldots, g_{12}$ of the first few Toda constants of motion $f_2, \ldots, f_{12}$, in the FPU dynamics; semilog scale. Parameters: $N = 1023$, $\alpha = 1$, $\beta = 2$, $\varepsilon = 8 \times 10^{-4}$.*

a partial randomization, while the Toda actions stay constant. The different behavior of the low and the high modes is easily explained if one further conjectures that low modes depend more strongly on the angles, and decorrelate because of the angle randomization occurring in the integrable dynamics, while high modes do less, being closer to functions of the actions only. Unfortunately, explicit expressions of the Toda action–angle variables are not known, so a direct test of this conjecture is problematic. This thought however prompted us to study the Toda constants of motion in the FPU dynamics, as will be explained in the next section.

– High modes are more sensitive than low modes to the non–integrability of the FPU model. Their decorrelation, larger in FPU than in Toda, shows, before low modes, that the FPU model, on a larger time scale, is chaotic and ultimately obeys statistical mechanics; low modes are lazier. By the way: if we look again at figure 1, we can observe there too that low modes move immediately to produce the intermediate state, but then are slower than high modes to reach equipartition. In other words, when nonintegrability starts to play a significant role, then a hydrodynamic-like regime sets in, with a decay (with respect to the Toda level) of the energy autocorrelation of a given mode which is the faster, the shorter is the wavelength of the mode.

– The behavior of high modes, when β is varied, shows rather clearly that the distance from Toda, and not the nonlinearity, produces the non–integrability: for example, the decorrelation of high modes is (slightly) more pronounced at $\beta = 0$ ($\Delta = -1$) than at $\beta = \beta_T$ ($\Delta = 0$) or $\beta = 4/3$ ($\Delta = 1$), in spite of the smaller nonlinearity.

3 Generic initial data: the Toda constants of motion; the large time scale

The problem arises whether it is possible to investigate numerically the long time scale for generic initial conditions. At first sight this might appear prohibitive: indeed the necessity of a large statistics on the initial data, of the order of 10^4 or possibly more, makes it unrealistic, at least to our means, to reach times substantially larger than 10^6, certainly not 10^9 or 10^{10}.

A more careful inspection shows that the question is subtler. Indeed the phenomenon we wish to look at, which according to the first of the above comments characterizes the long time scale, is the slow drift of the actions in the phase space. Such a motion does not start after a certain time: it is obviously progressive and takes place from the beginning, but if one studies any function $f(I,\varphi)$ depending both on the actions I and on the angles φ of the Toda system, like the energies of the normal modes or of groups of them, then for quite a long time (the first time scale) the behavior of f is dominated by the motion of the angles, and only on much larger times (the second time scale) the evolution of the actions is sufficiently large to be practically appreciated.

The way out should be clear: the evolution of the actions can be observed from the very beginning, without the "noise" produced by the motion of the angles, if one looks at functions $f(I)$ depending only on the Toda actions and not on the angles. The actions themselves, as already remarked, cannot be easily computed (see however [13]). But a set of independent constants of motion is well known, namely the functions introduced in [11, 12] to prove the integrability of the Toda model. Such functions are the same (up to a normalization) in the two papers; they refer to a periodic model, but as is well known, motions of a model with fixed ends and N moving particles are a subset of motions of a periodic model with $N' = 2N+2$, namely motions for which the coordinates (q', p') of the periodic model satisfy, initially and consequently at any time, the symmetry

$$q'_{N+1+i} = -q'_{N+1-i} \ , \qquad p'_{N+1+i} = -p'_{N+1-i} \ , \qquad i = 0, \dots, N \ , \tag{9}$$

with of course $q'_i = q_i$, $p'_i = p_i$ for $i = 1, \dots, N$. The constants $f_1, \dots, f_N$ of the fixed–ends model with potential (3) are defined as follows: let

$$a_i = \frac{1}{\lambda} e^{\frac{\lambda}{2}(q'_{i+1} - q'_i)} \ , \qquad i = 1, \dots, N' \ ,$$

and consider the matrix

$$L = \begin{pmatrix} p'_1 & a_1 & & & & & a_{N'} \\ a_1 & p'_2 & a_2 & & & & \\ & a_2 & p'_3 & a_3 & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & & & a_{N'-2} & p'_{N'-1} & a_{N'-1} \\ a_{N'} & & & & & a_{N'-1} & p'_{N'} \end{pmatrix} ;$$

then it is[4]

$$f_s = \mathrm{Tr}(L^{2s}) \ , \qquad s = 1, \dots, N \ .$$

[4] Odd powers of L do provide the additional constants of motion of the larger periodic model, but identically vanish for motions satisfying (9).

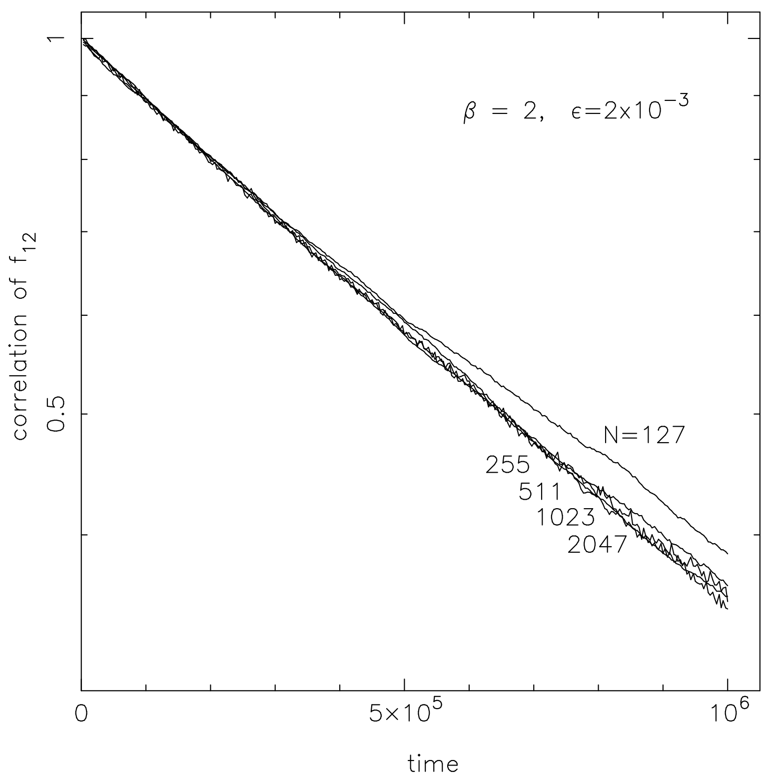


Figure 8: *The correlation function $g_{12}(t)$ vs. t for the FPU model with $\alpha = 1$, $\beta = 2$, $\varepsilon = 2 \times 10^{-3}$, and different values of N, namely $N = 127$ (upper curve) and $N = 255, 511, 1023, 2047$ (lower almost indistinguishable curves).*

The first constant f_1 is easily checked to be (up to a multiplicative constant) the energy of the Toda model, almost conserved in FPU too and thus not much interesting. We focused our attention on the first few of the remaining ones, namely on $f_2, \ldots, f_{12}$, and computed their correlation functions $g_i \equiv g_{f_i}$, $i = 2, \ldots, 12$, in the FPU dynamics. At variance with the g_{e_j}, such correlation functions are only sensitive to the nonintegrability of FPU. *Any decay of such correlation functions depends on the drift of the Toda actions in the FPU dynamics* and can reveal it.

Figure 7 shows the decay of $g_2, \ldots, g_{12}$ for a FPU model with $\beta = 2$, $N = 1023$, at $\varepsilon = 8 \times 10^{-4}$, in semi–log scale. Quite remarkably, the correlation functions accumulate, by increasing the index, on a limit curve $\tilde{g}(t)$; this is not far from a straight line i.e. an exponential $\tilde{g}(t) \sim e^{-t/\tau}$, and so defines a time scale τ. This is the large time scale we are looking for, namely the time scale in which the motion of the actions presumably makes the system ergodic, only the total energy being conserved. The "noise", so to speak, produced by the advancing of the angles in the integrable dynamics, has been filtered out, and we could reliably observe the slow action drift without waiting too much. *The decay of such correlations is a mixing property; the time scale τ of such a decay seems to us to be the time scale FPU were searching* (see the quotation in footnote 1).

In principle, one should perform a systematic study of the decay time τ, so as to determine its dependence on N, ε and β. Such a systematic study, however, is hard and would require a quite considerable numerical effort as well as a careful analysis of errors, so we defer it to a future paper. Here we limit ourselves to a preliminary study, sufficient however to give a hint on such dependencies and to roughly compare the behavior of τ, as a function of N, ε, β, with the quantity

$T(N, \varepsilon, \beta)$ mentioned in Section 1.4.

Concerning the dependence on N, figure 8 shows the behavior of $g_{12}(t) \simeq \tilde{g}(t)$ vs. t, for $\beta = 2$ and[5] $\varepsilon = 2 \times 10^{-3}$, N ranging from 127 to 1023. Quite clearly, at least at this value of the specific energy, the behavior of g_{12} is substantially independent of N for, say, $N = 255$ or bigger.

Concerning the dependence on β, this has been investigated for $N = 511$ and $\varepsilon = 2\times 10^{-3}$. The question is whether the quadratic law (7), observed in connection with FPU–like initial conditions, is valid for generic initial data, too. The answer seems to be positive. Figure 9 reports our result: the left panel shows g_{12} for different values of β, more precisely for $\Delta = \pm 1/2, 1, 3/2, 2, 3, 4$, with Δ as in (8); already here it is clear that the highest values of the inverse slope τ are obtained for Δ close to 0, and that $\Delta = \pm 1/2$ give the same τ. The right panel shows the same quantities g_{12}, plotted however vs. the rescaled time

$$t_\beta = \Big(\frac{\beta - \beta_T}{2 - \beta_T}\Big)^2 t \ ;$$

the collapse of different curves into a single one is not perfect, but in our opinion is satisfactory; in particular, the three lower curves, referring to $\Delta = \pm 1/2, 1$, almost exactly superimpose.

Finally we come to the dependence of τ on ε. Here we have a little novelty: the power law (6), obtained in connection with FPU–like initial data, is not fully confirmed. Figure 10, upper panel, shows g_{12} as a function of the unrescaled time t for $N = 1023$, $\beta = 2$ and several values of ε ranging from $\varepsilon = 4 \times 10^{-3}$ to 8×10^{-4}, in geometric progression. The lower left panel shows the same quantities, plotted however vs. the rescaled time

$$t_\varepsilon = \Big(\frac{\varepsilon}{2 \times 10^{-3}}\Big)^a t \ ,$$

with $a = 9/4$; the result is clearly unsatisfactory. A better rescaling (the best among power laws) is the one with slightly larger exponent $a = 5/2$, see the lower right panel. One should take into account that the time scales $T(N, \varepsilon, \beta)$ and $\tau(N, \varepsilon, \beta)$ we are here comparing differ in one important point: T is sensitive only to the modes in the tail, namely such that $k/N > 1/2$, while τ is sensitive to all degrees of freedom; but as already remarked, low modes are a little lazier than higher modes in the vicinity of equilibrium. This is of course only one of the possible explanations of the difference in the power laws, and the question certainly needs further investigation.

Remarks on the numerical integration. For the numerical integration we used a symplectic algorithm of order 4 obtained by suitably composing three *leap frog* (or Störmer–Verlet) elementary steps, according to the Yoshida scheme [25]. The time–step we used was either $h = 0.1$ or, when computing correlation functions g_j, $h = 0.05$.

It is very difficult to estimate the accuracy of long–time computations. What makes us confident that computations are sensible are a few facts: first, the overall coherence of results, and their independence of the time–step (as we occasionally checked); second, the great difference in the long–time behavior between FPU and Toda: looking for example at figure 1, it is hard to believe that the growth of the tail leading eventually to equipartition is due to numerical errors, since a similar growth does not occur at all in Toda. Third, the accuracy of the conservation of energy, with relative error of the order 10^{-5}, and the conservation of the Toda constants of motion in the Toda dynamics, with similar accuracy; such a good conservation also seems to exclude that the observed evolution of the Toda constants of motion in the FPU dynamics, that is the long time

[5] Here and in the following, smaller values of ε would be desirable, but times then get much larger and the numerical work goes behind our possibilities.

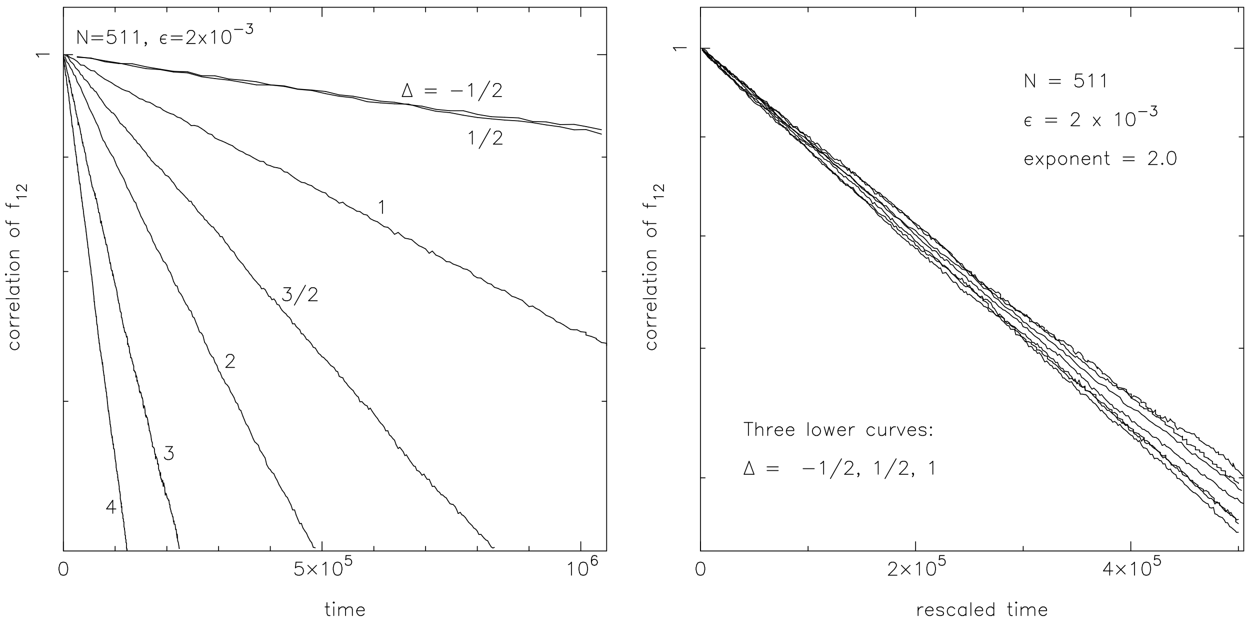


Figure 9: *Left panel: the correlation function* $g_{12}(t)$ *vs.* t *for the FPU model with* $N = 511$, $\alpha = 1$, $\varepsilon = 2 \times 10^{-3}$, *and different* β, *namely* $\Delta = \pm 1/2, 1, 3/2, 2, 3, 4$, Δ *as in (8). Right panel: same quantities plotted vs. the rescaled time* t_β. *The three almost coinciding lower curves in the right panel refer to* $\Delta = \pm 1/2, 1$.

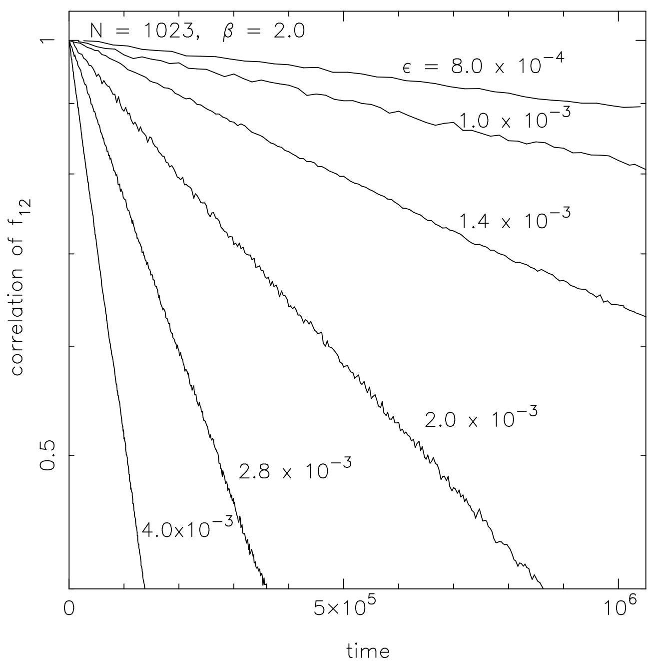
evolution of FPU, is the result of numerical errors, as one could suspect (and colleagues frequently ask at conferences). For a recent discussion on numerical errors in the integration of FPU–like systems, see [26].

For our computations we could use a cluster of over 100 CPU's, for a reasonable fraction of time. A few data will give an idea of the numerical effort we had to make. With $h = 0.1$ and $N = 1023$, a single FPU trajectory, up to $t = 10^6$, takes approximately 11 CPU–seconds; for $t = 10^{10}$ as in the left panel of figure 1, and 24 independent initial data, this requires about 15 CPU–days simultaneously on 24 CPU's, with an overall amount of almost one CPU-year.[6]

Integrating the Toda dynamics, due to the necessity to compute at each step an exponential in place of a polynomial of degree 3, is almost three times as expensive (for this reason the integration time, see figure 2, has been reduced). For the Toda model with $N = 8191$, up to $t = 10^6$ as in figure 6, the required CPU–time, using $h = 0.1$, is approximately 1.1 hours per initial condition; for 24,000 initial condition this corresponds to about 20 CPU–days, using simultaneously 100 CPU's (this is the most expensive computation of the paper, done of course for a long wall–time in background).

Concerning figures 7–10, taking into account the little extra time to compute the f_j's and the use of $h = 0.05$, each trajectory up to $t = 10^6$, for $N = 1023$, requires approximately 5 minutes; for 24,000 initial data this gives an approximate overall amount of 2,000 CPU–hours. The corresponding overall cost of figure 9 is about 17,000 CPU–hours, that is two weeks simultaneously on 50 CPU's; figure 10 is almost twice as expensive.

[6] It is not easy to compare the speed of MANIAC I used by FPU with the speed of modern CPUs. A prudent estimate gives a ratio of the order 10^5; this means that each panel of figure 1 would require, on MANIAC I, about 10^5 CPU years.



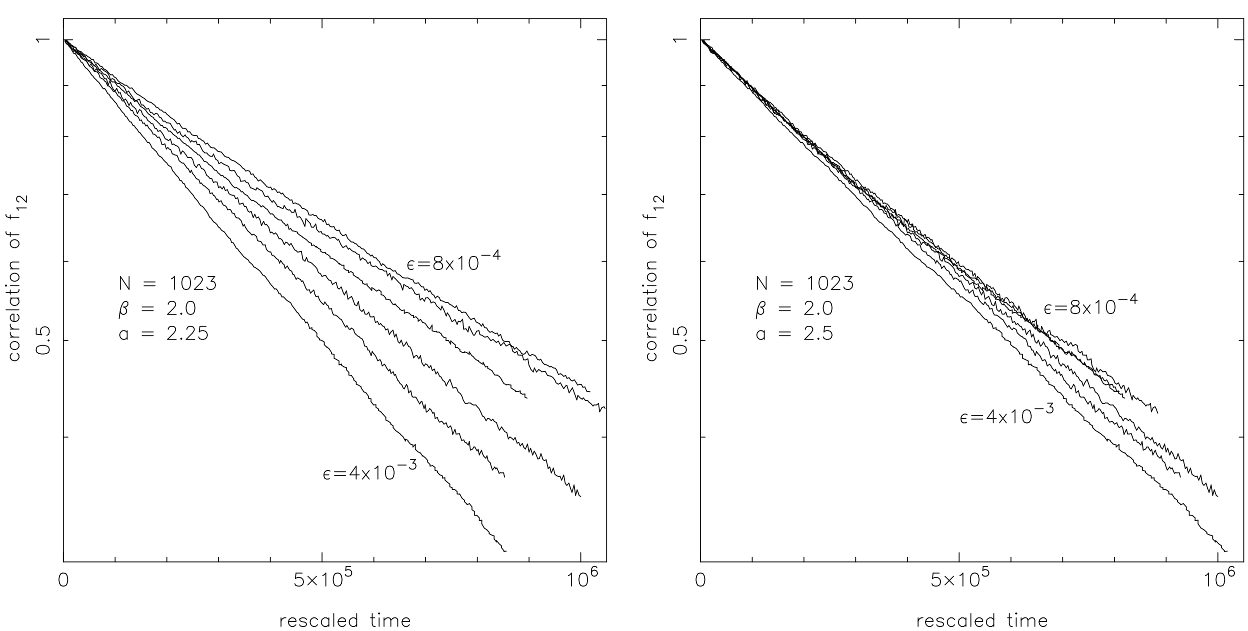


Figure 10: *Upper panel: the correlation function $g_{12}(t)$ vs. t for the FPU model with $N = 1023$, $\alpha = 1$, $\beta = 2$, and different values of ε ranging from 8×10^{-4} to 4×10^{-2}, in geometric progression. Lower panels: the same quantities plotted vs. the rescaled time t_ε, with exponent $a = 9/4$ (left) and $a = 5/2$ (right).*

4 Conclusions

We think we provided good evidence that the $\alpha+\beta$ FPU model should be regarded as a perturbed Toda model, rather than as a perturbed linear model, for any kind of initial conditions. As far as only long wave motions are present, the KdV wave equation also provides a reference integrable dynamics (a reliable normal form, according to [22]), but if all degrees of freedom are excited, in particular for generic states with respect to the microcanonical measure, then the reference integrable system is Toda. In this view, the presence of two time–scales is confirmed, and interpreted in the easiest way: for any function $F(I,\varphi)$ of the Toda action–angle variables—harmonic normal modes, groups of them, any other—and for a time T corresponding to the first time scale, it is

$$\overline{F}(T) \simeq \langle F(I_0,\,.\,)\rangle_\varphi \ , \tag{10}$$

where of course I_0 is the initial action and $\langle\,.\,\rangle_\varphi$ denotes averaging on the Toda angles φ; such a partial averaging provides the FPU intermediate state. For Toda of course (10) is an asymptotic equality, which holds without the occurrence of a second time scale.

On much larger times one has instead

$$\overline{F}(T) \longrightarrow \langle F\rangle \ ,$$

i.e. ergodicity. An exception is the case of the Toda constants of motions f_s we studied in Section 3, which do not depend on the angles and so the above partial averaging (10) is meaningless.

Unfortunately, as already remarked, the transformation from (q,p), or (Q,P), to (I,φ), is not known, and so a direct test of this picture cannot be produced. There exist however a procedure to compute numerically at least the actions [13]; following it, for example, it would be possible to know, for each initial datum, the number of significantly excited actions, and this, in our opinion, could be rather helpful to understand the behavior of the system for FPU–like initial data, in particular the width of the intermediate state. Such a study (which is based on the search of the zeros of a polynomial of order N) gets problematic for large N, and certainly requires a considerable effort. We plan to investigate this point in a near future.

The presence of the integrable Toda model makes the $\alpha+\beta$ model exceptional. The $\alpha+\beta$ and the pure β model are often regarded, in the literature, as interchangeable. As far as only long wave modes are excited and the continuum approximation holds, they probably are, at least to some extent: in the same sense the $\alpha+\beta$ FPU model is approximated by the KdV, the pure β model is approximated by the mKdV, which is also integrable. Already in [19], however, some difference between $\alpha+\beta$ and pure β are made evident, and we expect the differences get stronger if generic initial data are considered (our colleagues in Milano showed to us some preliminary results going in this direction). We also think that models with an on–site potential, like the φ^4 one, constitute a class apart; for a particular model within this class, a rigorous proof of the existence of exponentially long relaxation times has been given in [27].

Even more different the situation is expected to be in dimension larger than one, for the absence of discrete integrable models or integrable wave equations, that can be used as good approximations of convenient extensions of FPU in dimension two or three. One of us has some experience of work in dimension two, see [28, 29]; after the revisitation of dimension one we made in this paper, and before it in [22, 19], it is time to go back to dimension two, and we hope to be able to do it rather soon.

Acknowledgements: We feel indebted to all our colleagues of Milano, in particular Luigi Galgani, Antonio Giorgilli, Dario Bambusi, Andrea Carati, Simone Paleari and Tiziano Penati, for (years of) intense discussions; their criticism has been particularly helpful. The results contained in the paper have been preliminarily shown to Giovanni Gallavotti and Giorgio Parisi (Rome), and more recently to the research group around Jean-Pierre Eckmann (Genéve); the comments we received have been quite useful to improve the exposition of our results. This research has been supported by the University of Padova, also covering the post–doc position of H.C.

References

[1] E. Fermi, J. Pasta, and S. Ulam: *Studies of Non Linear Problems,* Los-Alamos Internal Report, Document LA-1940 (1955), in: *Enrico Fermi Collected Papers*, Vol. II, The University of Chicago Press, Chicago, and Accademia Nazionale dei Lincei, Roma, 1965, pp. 977-988 (also reproduced in ref. [4]).

[2] J.L. Tuck, M.T. Menzell, *The superperiod of the nonlinear weighted string (FPU) problem,*, Adv. Math. **9**, 399–407 (1972); computations go back to 1961 (see Ulam's presentation of the FPU paper in [1]).

[3] Chaos focus issue: *The "Fermi–Pasta–Ulam" problem—the first 50 years.* Chaos **15**, 2005.

[4] G. Gallavotti (Ed.): *The Fermi-Pasta-Ulam Problem: A Status Report*, Lect. Notes Phys. **728**, Springer, Berlin-Heidelberg, 2008.

[5] E. Fucito, F. Marchesoni, E. Marinari, G. Parisi, L. Peliti, S. Ruffo and A. Vulpiani, *Approach to equilibrium in a chain of nonlinear oscillators,* J. de Physique **43**, 707–713 (1982).

[6] R. Livi, M. Pettini, S. Ruffo, M. Sparpaglione and A. Vulpiani, *Relaxation to different stationary states in the Fermi-Pasta-Ulam model,* Phys. Rev. A **28**, 3544–3552 (1983).

[7] L. Berchialla, L. Galgani and A. Giorgilli, *Localization of energy in FPU chains,* DCDS - A **11**, 855–866 (2004).

[8] L. Berchialla, A. Giorgilli and S. Paleari, *Exponentially long times to equipartition in the thermodynamic limit,* Phys. Lett. A **321**, 167–172 (2004).

[9] G. Benettin, A. Carati, L. Galgani and A. Giorgilli, *The Fermi–Pasta–Ulam problem and the metastability perspective,* in ref. [4], pp. 151–189.

[10] N.J. Zabusky and M.D. Kruskal, *Interaction of solitons in a collisionless plasma and the recurrence of initial states,* Phys. Rev. Lett. **15**, 240–245 (1965).

[11] M. Hénon, *Integrals of the Toda lattice,* Phys. Rev. B **9**, 1921–1923 (1974).

[12] H. Flaschka, *The Toda Lattice. II. Existence of integrals*, Phys. Rev. B **9**, 1924–1925 (1974).

[13] W.E. Ferguson, H. Flaschka and D.W. McLaughlin, *Nonlinear Toda Modes for the Toda Chain,* Journ. Comput. Phys. **45**, 157–209 (1982).

[14] S. Isola, R. Livi, S. Ruffo and A. Vulpiani, *Stability and chaos in Hamiltonian dynamics,* Phys. Rev. A **33**, 1163–1170 (1986).

[15] L. Casetti, M. Cerruti–Sola, M. Pettini and E.D.G. Cohen, *The Fermi–Pasta–Ulam problem revisited: Stochasticity thresholds in nonlinear Hamiltonian systems,* Phys. Rew. E **55**, 6566–6574 (1997).

[16] A. Giorgilli, S. Paleari and T. Penati, *Local chaotic behaviour in the Fermi-Pasta-Ulam system,* Discr. Cont. Dyn. Sys. B **5**, 991–1004 (2005).

[17] N.J. Zabusky, Zhanyu Sun and Gaozhu Peng, *Measures of chaos and equipartition in integrable and nonintegrable lattices,* Chaos **16**, 013130/1–12 (2006).

[18] A. Henrici and T. Kappeler, *Nekhoroshev theorem for the periodic Toda lattice,* Chaos **19**, 033120/1–13 (2009).

[19] G. Benettin and A. Ponno, *Time–scales to equipartition in the Fermi–Pasta–Ulam problem: finite-size effects and thermodynamic limit,* Journ. Stat. Phys. **144**, 793–812 (2011).

[20] A. Ponno, H. Christodoulidi, Ch. Skokos and S. Flach, *The two-stage dynamics in the Fermi-Pasta-Ulam problem: From regular to diffusive behavior,* Chaos **21**, 043127/1–14 (2011).

[21] T. Genta, A. Giorgilli, S. Paleari and T. Penati, *Packets of resonant modes in the Fermi–Pasta–Ulam system*, Phys. Lett. A **376**, 2038–2044 (2012).

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

[23] D. Bambusi and A. Ponno, *On Metastability in FPU,* Commun. Math. Phys. **264**, 539–561 (2006).

[24] A. Carati, L. Galgani, A. Giorgilli and S. Paleari, *FPU phenomenon for generic initial data,* Phys. Rev. E **76**, 022104/1–4 (2007).

[25] H. Yoshida, *Construction of higher order symplectic integrators,* Phys. Lett. A **150**, 262–268 (1990).

[26] G. Benettin and A. Ponno, *On the numerical integration of FPU–like systems,* Physica D **240**, 568-573 (2011).

[27] A. Carati and A. Maiocchi, *Exponentially long stability times for a nonlinear lattice in the thermodynamical limit,* Commun. Math. Phys. **314**, 129–161 (2012).

[28] G. Benettin, *Time–scale for energy equipartition in a two–dimensional FPU model,* Chaos **15**, 15105/1–8 (2005).

[29] G. Benettin and G. Gradenigo, *A study of the Fermi–Pasta–Ulam problem in dimension two,* Chaos **18**, 013112/1–13 (2008).