J Stat Phys (2011) 144:793–812
DOI 10.1007/s10955-011-0277-9

Time-Scales to Equipartition in the Fermi–Pasta–Ulam Problem: Finite-Size Effects and Thermodynamic Limit

G. Benettin · A. Ponno

Received: 13 January 2011 / Accepted: 12 July 2011 / Published online: 5 August 2011
© Springer Science+Business Media, LLC 2011

Abstract We investigate numerically the common $\alpha+\beta$ and the pure β FPU models, as well as some higher order generalizations. We consider initial conditions in which only low-frequency normal modes are excited, and perform a *very* accurate systematic study of the equilibrium time as a function of the number N of particles, the specific energy ε, and the parameters α and β. While at any fixed N the equilibrium time is found to be a stretched exponential in $1/\varepsilon$, in the thermodynamic limit, i.e. for $N\to\infty$ at fixed ε, we observe a crossover to a power law. Concerning the (usually disregarded) dependence of $T_{\rm eq}$ on α and β, we find it is nontrivial, and propose and test a general law. A central role is played by the comparison of the FPU models with the Toda model.

Keywords Fermi–Pasta–Ulam · Equipartition time · Thermodynamic limit

1 Introduction

This paper is devoted to a numerical study of the long-term dynamics in the Fermi–Pasta–Ulam (FPU) problem [1]. The Hamiltonian (although some generalizations will also be taken into consideration) is the original FPU one:

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

with

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

G. Benettin (✉) · A. Ponno
Dipartimento di Matematica Pura e Applicata, Università di Padova, Via Trieste 63, 35131 Padova, Italy
e-mail: benettin@math.unipd.it

A. Ponno
e-mail: ponno@math.unipd.it

The motivation is also the original one, namely understanding how, and on which time scale, the model, if prepared in a state far from statistical equilibrium, approaches equilibrium.[1] The numerical investigation also follows FPU: energy is initially given to a small subset of normal modes of low frequency, and statistical equilibrium is identified with energy equipartition among all modes.

The normal modes are defined as usual as

$$Q_k = \sqrt{\frac{2}{N+1}} \sum_{i=1}^{N} q_i \sin \frac{\pi k i}{N+1}, \qquad P_k = \sqrt{\frac{2}{N+1}} \sum_{i=1}^{N} p_i \sin \frac{\pi k i}{N+1}; \tag{3}$$

the (harmonic) energy E_k and frequency ω_k of the k-th mode are then respectively

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

energy equipartition means that the time average of E_k up to time T, namely

$$\overline{E}_k(T) = \frac{1}{T} \int_0^T E_k\big(P(t), Q(t)\big)\, \mathrm{d}t, \tag{4}$$

for large T converges (up to minor nonlinear contributions) to the energy per degree of freedom $\varepsilon = E/N$, E denoting the total energy.

The literature on the subject, over more than 50 years, is so huge, that it is not even conceivable to summarize it here. The state of the art, updated to a few years ago, can be found in two collections of papers [2, 3], but it is hard—and this is a good indication of the difficulty of the problem—to extract, from all of these papers or reviews, a unitary view, free of contradictions. Here we assume, as a possible frame for a critical discussion, the point of view summarized, for example, in [4]. In the very essence: at small ε, (at least) two well separated time scales enter the problem.

- In a relatively short time, the system reaches a state different from the initial one—several modes, in addition to the initially excited ones, enter the game—but still very far from energy equipartition. This is the problematic state observed at low energy in the original FPU paper.
- Such a state, however, is only apparently stationary, in fact it is not: similarly to a metastable state of statistical mechanics, on a much longer time scale it does evolve towards statistical equilibrium.

Concerning the possible physical meaning of the FPU state, and in general the relevance of the FPU problem for the foundations of statistical mechanics, different authors expressed quite different points of view. But on the above outlined elementary scenario—the two time scales—as far as we know there is a rather general agreement.

Figure 1 (*left*) illustrates the scenario. The figure refers to an $\alpha + \beta$ model, with $N = 1{,}023$, $\alpha = -1$ and $\beta = 2$ (see the next section for comments on the choice of the parameters), at the rather small value of $\varepsilon = 10^{-4}$, and shows the averaged energy spectrum, i.e. $\overline{E}_k(T)$ vs. k/N (with a minor modification in the definition of $\overline{E}(T)$, see Sect. 2.2), at different times T in geometric progression. The energy was initially equidistributed among the

[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.*

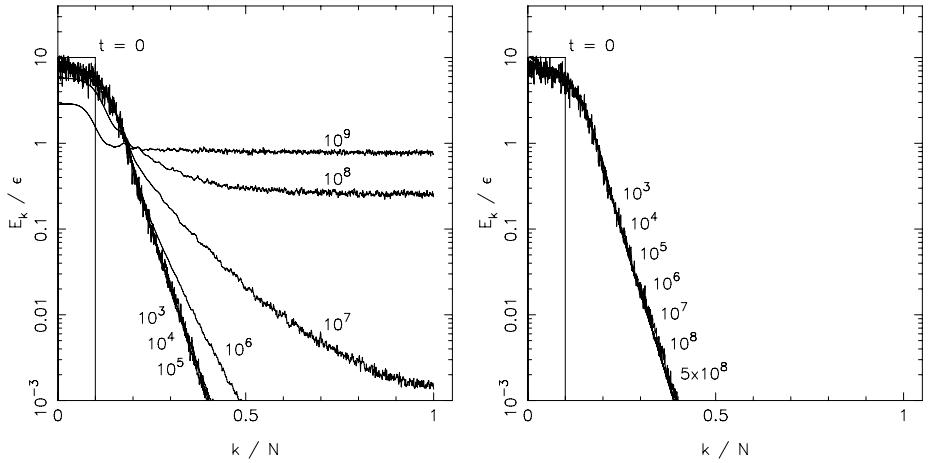

Fig. 1 *Left*: the averaged energy spectrum at different time T, for the $\alpha + \beta$ model; $N = 1{,}023$, $\alpha = -1$, $\beta = 2$, $\varepsilon = 10^{-4}$. *Right*: same quantity for the tangent Toda model (same N and ε). Initial excitation of modes with $0 < k/N \leq 0.1$

lower 10% of modes (the rectangular profile in the figure). 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. Afterwords the dynamics slowly evolves towards energy equipartition, the high-frequency modes being progressively involved into the energy-sharing game. If the phases of the initially excited modes are chosen randomly (this is crucial, see [5]), then the behavior, for large N at fixed small ε, is independent of N, i.e. 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 two-time scales scenario was first outlined, as far as we know, in two pioneering papers, nearly twenty years ago [6, 7]. More recently, it was reconsidered and systematically studied in several papers, both theoretical and numerical, among them [5, 8–11]. The situation however, in our opinion, is far from being clear. As far as we know:

- $\alpha + \beta$ model: the short time scale is very accurately described numerically and rather well understood theoretically, the main theoretical tool being the KdV equation. The use of KdV to explain FPU goes back to the celebrated paper [12]; recent papers on the subject include [5, 11]. As shown, in particular, in the systematic study [5], the formation of the metastable profile can be predicted exactly—dependence on N, on ε, on time, on the initial state—using as normal form for the system a convenient truncation of the KdV equation. *This means that the formation of the metastable state is integrable in character*. Besides KdV, the (integrable) Toda model has been also used as a comparison for the short time dynamics of FPU, see in particular [13–15]. The model, we recall, has potential

$$V_T(r) = V_0\big(e^{\lambda r} - 1 - \lambda r\big), \tag{5}$$

and for $V_0 = \lambda^{-2}$, $\lambda = 2\alpha$, it coincides with (2) up to the order r^3. As stressed in [15], in the short-time scale the $\alpha + \beta$ FPU is indistinguishable from Toda, the only difference being that for Toda a second time-scale does not exist; this is shown, for example, by the *right panel* of Fig. 1, where the Toda potential (5) was used in place of the FPU potential (2).

In front of this detailed knowledge of the short time scale of the $\alpha + \beta$ FPU, very few is known, in our opinion, concerning its long time scale, at least in the limit of large N at fixed (small) ε in which we are here interested.

- Pure β-model: here the situation is reversed; the short time scale, to our knowledge, has never been accurately studied, while interesting results do exist concerning the long time scale. In [8] it was shown, first of all, that even at low ε, on sufficiently long times, equipartition always occurs. The equipartition time $T_{\rm eq}$, suitably defined through the so-called spectral entropy and the related effective number of equipartizing degrees of freedom [16, 17], was roughly estimated to grow, for small ε, as ε^{-3}. More accurate computations in [10], performed as in [8], corrected the power law into a stretched exponential, actually

$$T_{\rm eq} \sim e^{(\varepsilon^*/\varepsilon)^{1/4}}, \tag{6}$$

and also stressed the independence of $T_{\rm eq}$ on N, if large; in both papers the initial data (this will be shown below to be crucial) were displaying energy equipartition in the interval $1/64 \le k/N \le 5/64$, the minimal frequency of the excited modes being bounded away from zero.

In this paper we shall consider both the $\alpha + \beta$ and the pure β models, and provide numerical evidence that:

(i) In the $\alpha + \beta$ model, in the thermodynamic limit, the dependence of the equilibrium time $T_{\rm eq}$ on ε is a power law,

$$T_{\rm eq} \simeq A\varepsilon^{-a}; \tag{7}$$

for given α, different values of β modify the constant A, but not the exponent a, which is numerically found to be approximately 2.25; one is tempted to say $a = 9/4$. However, if β is equal to (or very close to) the Toda value $\beta_T = (2/3)\alpha^2$, obtained from (5) by a Taylor expansion, then a grows to 3. In fact, by moving β at fixed ε, the equilibrium time is found to have a maximum around $\beta = \beta_T$. This encourages to regard FPU not as a linear model perturbed by the terms αr^3 and βr^4, the quartic term being introduced only to guarantee stability, but as a perturbed Toda model, with perturbation $V - V_T$ dominated by the quartic term $(\beta - \beta_T)r^4$, if nonvanishing. To exploit this idea, modified FPU models which are tangent to Toda at higher order ($\alpha + \beta + \gamma + \delta$ FPU models, with obvious meaning), are also taken into account. For such models too, we find power laws of the form (7), but with different exponents, apparently depending only of the "degree of tangency" of the potential V to V_T.

Besides the asymptotic power law (7), our numerical results lead us to conjecture and successfully test a general law of the form

$$T_{\rm eq} \sim \alpha^{-1/2}\varepsilon^{-1/4}\left[\left(\beta - c\alpha^2\right)\varepsilon\right]^{-2}, \quad c \simeq 2/3,$$

collecting the dependence of $T_{\rm eq}$ on α, β and ε (such an expressions holds out of rather small neighborhoods of $\alpha = 0$ and $\beta/\alpha^2 = c$, where crossover effects eliminate the divergences; see Sect. 4.3 for details).

Another interesting feature of the $\alpha+\beta$ model, already observed in [5, 9] for the short time dynamics and here confirmed for the long time dynamics too, is that the precise initial energy distribution among normal modes, provided only low modes are excited, is irrelevant.

(ii) In the pure β model, the initial energy distribution among the normal modes is instead relevant, both for the short and the long time dynamics. Indeed, at variance with the $\alpha+\beta$ model, the metastable state always keeps memory of the initial state, being practically a minor modification of it. Concerning the equilibrium time, it can be either a stretched exponential as in (6) or a power law as in (7), depending on the initial datum. Precisely, if the frequency of the initially excited modes, as in [8] and [10], is bounded away from $\omega=0$, then in the limit of large N (and for small ε, of course) we do obtain the exponential law (6); if instead frequencies arbitrarily close to zero are excited, then asymptotically for large N we get the power law (7), remarkably with the same exponent $a\simeq 2.25$ as in the $\alpha+\beta$ model.

The results reported in this paper required several months of computer time, in a cluster of 24 fast CPU's (about one year, including preliminary and less interesting computations).

2 Preliminaries

2.1 The Models

Let us consider the FPU Hamiltonian (1), with V as in (2) and $\alpha\neq 0$. As is well known, the values of α and β are not separately relevant, the only relevant parameter being the dimensionless ratio β/α^2: models with the same β/α^2 are identical up to a rescaling, otherwise they are not. More precisely a rescaling of the form $q=\varrho q'$, $p=\varrho p'$, turns α into $\alpha'=\varrho\alpha$ and $\beta'=\varrho^2\beta$, while at the same time ε turns into $\varepsilon'=\varrho^{-2}\varepsilon$; this is often reported by saying that the relevant quantities determining the dynamics are $|\alpha|\sqrt{\varepsilon}$ and $\beta\varepsilon$. Throughout the paper, for the $\alpha+\beta$ model, we made the simple choice[2] $\alpha=-1$ (only in Sect. 4.3, for a better understanding, we found convenient to vary α). Once α is fixed, β becomes relevant, and a special value of β turns out to be the Toda value β_T, obtained by expanding in Taylor series the Toda potential (5); choosing $V_0=1/(4\alpha^2)$ and $\lambda=2\alpha$, so as V_T coincides with V up to third order, one finds

$$V_T(r)=\frac{1}{2}r^2+\frac{\alpha}{3}r^3+\frac{\beta_T}{4}r^4+\frac{\gamma_T}{5}r^5+\frac{\delta_T}{6}r^6+\cdots,$$

with

$$\beta_T=\frac{2}{3}\alpha^2,\qquad \gamma_T=\frac{1}{3}\alpha^3,\qquad \delta_T=\frac{2}{15}\alpha^4.$$

If, as outlined in Sect. 1, one considers the $\alpha+\beta$ model as a perturbed Toda model, then for $\beta\neq\beta_T$ the leading perturbation is $(\beta-\beta_T)r^4$, while for $\beta=\beta_T$ the perturbation reduces to $-\gamma_T r^5$, much smaller at small energy. The game can go on, by naturally generalizing the potential (2) to

$$V(r)=\frac{1}{2}r^2+\frac{\alpha}{3}r^3+\frac{\beta}{4}r^4+\frac{\gamma}{5}r^5+\frac{\delta}{6}r^6 \tag{8}$$

[2]The negative sign is typical of molecular potentials, with elastic constant that gets harder by compressing the springs (the rescaling constant ϱ needs not to be positive).

(with $\delta > 0$ so as to ensure stability when $\gamma \neq 0$); for $\beta = \beta_T$ the leading perturbation is $(\gamma - \gamma_T)r^5$, while for $\gamma = \gamma_T$ it further reduces to $(\delta - \delta_T)r^6$. Following the idea that the FPU model should be regarded as a perturbed Toda model, we investigated all these possibilities.

Let us now come to the pure β model. For such a model the only substantial parameter, determining the dynamics up to a rescaling, is the product $\beta\varepsilon$; our simple choice was $\beta = 1$. Nothing else can be said: indeed an integrable chain like the Toda model, but with $\alpha = 0$, to be confronted with the pure β model, is not known, and so considering the pure β model as a perturbation of a linear model, with perturbation $\beta\varepsilon$, seems the best one can do up to now.

2.2 The Equipartition Time

In place of considering the time averages of the normal modes energies starting from $t = 0$, as in (4), we found it convenient to average over a running window of amplitude proportional to T, replacing (4) by

$$\overline{E}_k(T) = \frac{1}{\frac{1}{3}T} \int_{\frac{2}{3}T}^{T} E_k\big(P(t), Q(t)\big)\, \mathrm{d}t \tag{9}$$

(the precise choice of the width of the window, however, is largely irrelevant). This has the advantage of a quicker loss of the memory of the very special initial state, providing at the same time an averaging interval which becomes infinite for $T \to \infty$.

On the basis of the $\overline{E}_k$'s, one should then introduce a parameter providing a distance to equipartition. A frequently used parameter is the "effective relative number of degrees of freedom"

$$n_{\mathrm{eff}}(t) = \frac{\exp \eta(t)}{N},$$

where η is the "spectral entropy"

$$\eta(t) = -\sum_{k=1}^{N} w_k(t) \log w_k(t), \qquad w_k(t) = \frac{\overline{E}_k(t)}{\sum_{j=1}^{N} \overline{E}_j(t)};$$

see [16, 17]. Equipartition corresponds to $n_{\mathrm{eff}} = 1$. We tried to use this parameter, which however, in our experience, has the disadvantage to be quite sensitive to the re-distribution of energy among the already large w_k's (the flattening of the lower part of the spectrum in Fig. 1), but almost insensitive to the slow growth of the exponential tail of the high modes, till their w_k become large enough, although this is the crucial phenomenon eventually leading the system to equilibrium. So, we decided to focus the attention onto the high modes, practically (with some arbitrariness, of course) onto modes with $k > N/2$, looking at their overall energy normalized to the due amount in equipartition; this leads to consider the parameter

$$\mathcal{E}'(t) = \frac{\sum_{N/2<k\le N} \overline{E}_k(t)}{\frac{1}{2}\sum_{1\le k\le N} \overline{E}_k(t)}.$$

As a minor correction, to keep into account the nonflatness of the tail (this is relevant only in the transient, see Fig. 1) we multiplied $\mathcal{E}'$ by the effective relative number of degrees of

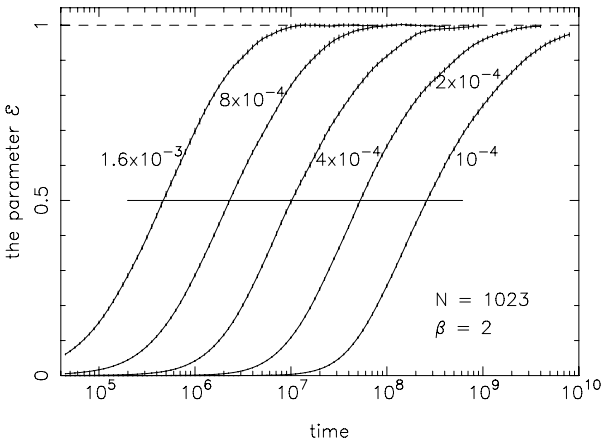

Fig. 2 The parameter $\mathcal{E}(t)$ vs. t, for the $\alpha + \beta$ model; $\alpha = -1$, $\beta = 2$, $N = 1{,}023$ and (left to right) $\varepsilon = 1.6 \times 10^{-3}$, 8×10^{-4}, 4×10^{-4}, 2×10^{-4}, 10^{-4}. Curves are the averages over 24 different random choices of the initial phases; error bars, just visible with respect to the line thickness, correspond to three std. deviations

freedom, restricted to the tail: practically, the parameter we used is not $\mathcal{E}'$ but

$$\mathcal{E}(t) = \frac{\exp \eta'(t)}{N/2} \mathcal{E}'(t),$$

where

$$\eta'(t) = -\sum_{N/2<k\le N} w'_k(t) \log w'_k(t), \quad w'_k(t) = \frac{\overline{E}_k(t)}{\sum_{N/2<j\le N} \overline{E}_j(t)}.$$

At equipartition $\mathcal{E} = 1$. Figure 2 shows the typical behavior of $\mathcal{E}(t)$ as a function of t, in semi-log scale. The figure refers to the $\alpha + \beta$ model, with $\alpha = -1$, $\beta = 2$, $N = 1{,}023$, and $\varepsilon = 4 \times 10^{-4}$, 2×10^{-4}, 10^{-4}. The initial condition was equipartition among the lowest 10% of modes, as in Fig. 1. The curves in the figure do not really represent $\mathcal{E}(t)$, but its average $\langle \mathcal{E}(t) \rangle$ on 24 different random choices of the initial phases of the excited modes; the (rather small, nearly invisible) error bars traced across the lines correspond to three standard deviations. We will come back on Fig. 2 in the last section, where further details concerning the "sigmoidal" curves are reported and commented.

Let us finally come to our definition of the equilibrium time $T_{\rm eq}$. Pragmatically, we decided to look, for each choice of the phases, at the time $\mathcal{T}$ at which $\mathcal{E}(t)$ reaches the value 0.5; then we define the equilibrium time $T_{\rm eq}$ as the average of $\mathcal{T}$ on the phases, $T_{\rm eq} = \langle \mathcal{T} \rangle$. The statistics on 24 different choices of the phases does not provide only the average $T_{\rm eq}$, but also a dispersion, and will allow us to trace, in our figures, error bars (three standard deviations). Let us stress that, on the whole, the procedure we use to identify $T_{\rm eq}$, is a refinement of what is usually done in the literature; the improved procedure leads, in our experience, to more accurate results, but does not exploit new ideas.

2.3 The Numerical Method

For the numerical integration, we used a leap-frog algorithm of order four, obtained by combining three leap-frog steps of order two, according to the Yoshida method [18]. The typical time-step was $\tau = 0.1$; the corresponding relative error in energy conservation, when

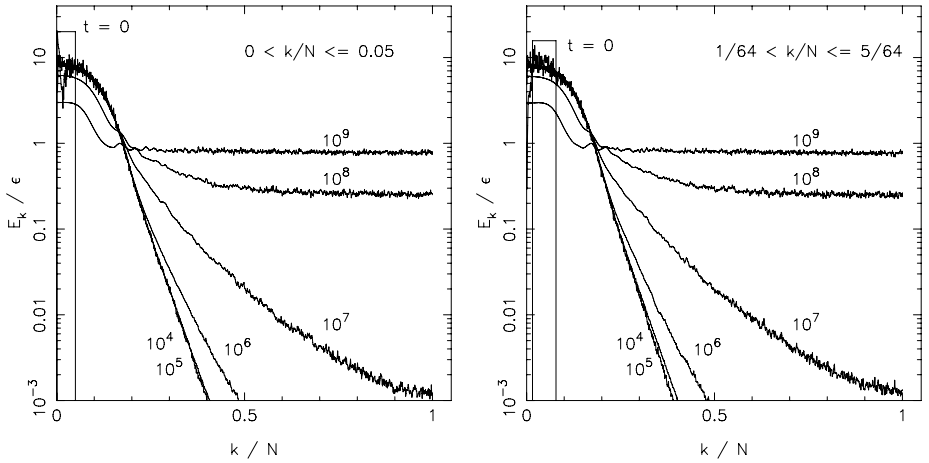

Fig. 3 As in Fig. 1 (*left*), but initial excitation of modes with $0 < k/N \leq 0.05$ (*left panel*) and $1/64 \leq k/N \leq 5/64$ (*right panel*). Time $t = 10^3$ was omitted

all modes are excited and do contribute to the total energy,[3] is around 10^{-5}. Occasional tests with smaller τ have been performed, with absolutely identical result.

Experience shows that leap-frog performs very well, for FPU problems. The essential reason is that, in spite of the errors, such an algorithm preserves very well two main features of the model, namely the structure of resonances among normal modes (in particular, the big acoustic resonance $\omega_k = k\omega_1 + \mathcal{O}(k^3/N^3)$) and, as is crucial, the strict "selection rules" in the nonlinear terms providing the interaction among normal modes. For a critical discussion see [19, 20] and references there quoted.

3 Results for the $\alpha + \beta$ Model

3.1 Irrelevance of the Initial Excitation

As already remarked in Sect. 1, the metastable state of the $\alpha + \beta$ model—as far as only low modes are excited, with random choice of the phases—is known to be independent of the details of the initial excitation [5, 9]. This is not obvious nor is easily interpreted, the short time behavior being essentially integrable; for further comments, see Sect. 5. Here we add the further nontrivial information that such an independence extends to the long time scale, too. This can be seen for example in Fig. 3, similar to Fig. 1 (*left*), but referring to initial equipartition among modes with $0 \leq k/N \leq 0.05$ (*left*) and $1/64 \leq k/N \leq 5/64$, as in [8, 10] (*right*). The spectrum at $t = 10^3$ has been omitted, since for narrower initial excitation the metastable state is reached in a longer time [5] (the figure would get a little

[3] By the way: the error in energy conservation is much smaller, as all numerical researchers on FPU experience, during the short time scale. This might appear to be in contrast with the symplecticity of the algorithm, which prevents energy drifts, but it is not: the point is that during the short time scale, only low modes appreciably contribute to the total energy, and the dimensionless quantity $\omega_k \tau$, which is the only relevant one, for such modes is small. But the apparent higher accuracy in this part of the dynamics is illusory, since the growth of the higher modes is precisely the phenomenon to be looked at.

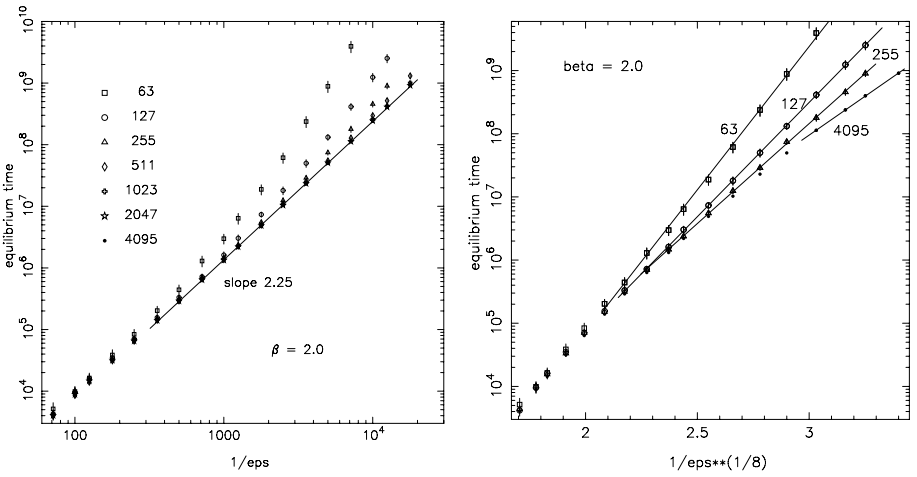

Fig. 4 *Left*: the equilibrium time $T_{\rm eq}$ as function of ε^{-1}, in log–log scale, for $N = 63$, 127, 255, 511, 1,023, 2,047, 4,095. *Right*: the same quantity as function of $\varepsilon^{-1/8}$, in semi-log scale, for $N = 63$, 127, 255, 4,095

confused), but the data at $t = 10^4$, 10^5 show that the short-time energy spectrum is the same, and moreover, as is very important, the long-term evolution towards energy equipartition is then practically identical: lines at different times do exactly superimpose. We exploited other choices of the initial data, with identical result.

3.2 The Equilibrium Times

Let us now come to the main target of our investigation, namely the dependence of the averaged equilibrium time $T_{\rm eq}$, as defined above, on N and ε.

Figure 4 (*left*) shows the behavior of $T_{\rm eq}$ vs. $1/\varepsilon$, in log–log scale, for different values of N, namely $N = 63$, 127, 255, 511, 1,023, 2,047 and 4,095, at fixed $\beta = 2$. Error bars (three std. deviations) are included, although not well visible for large N because much smaller then the symbol size. The figure shows that individual curves, for small N, are not straight lines, i.e., are not power laws; for large N, however, they progressively flatten, for larger and larger intervals, on a line, and are practically indistinguishable from it, in the exploited range of ε, already for $N = 1{,}023$. So, *the line, that is a power law as in* (7), *provides the dependence of* $T_{\rm eq}$ *on* ε *in the thermodynamic limit*. The computed value of the exponent, i.e. the slope of the line in the figure, is

$$a \simeq 2.25 = 9/4.$$

Concerning individual curves, a careful inspection shows that the upper part of them is rather well interpolated by a stretched exponential, precisely

$$T_{\rm eq} \sim \exp(\varepsilon_N^*/\varepsilon)^b, \quad b = 1/8;$$

see the *right panel* of the figure. By growing N, we did not observe any convergence of the lines appearing in the *right panel* to a limit line: their slope ε_N^* progressively decreases, while the point at which the linear behavior (the stretched exponential) starts, moves to the

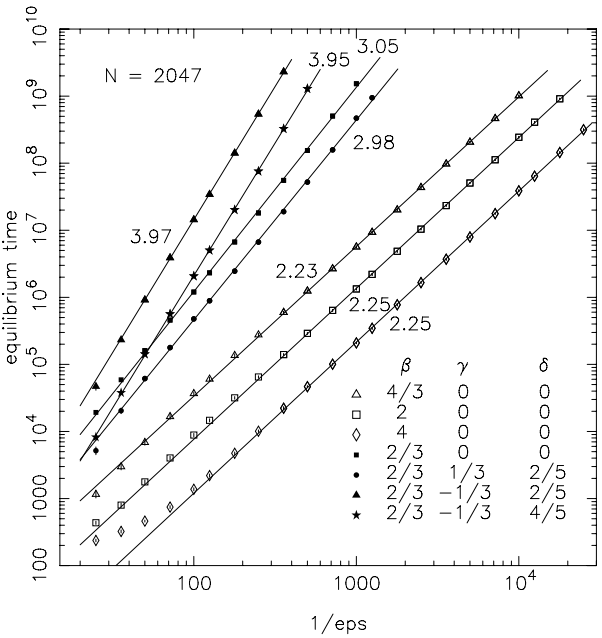

Fig. 5 The equilibrium time T_{eq} as function of ε^{-1}, in log–log scale, for the $\alpha + \beta$ model and its generalization (8). The *three rightmost lines*: $\gamma = \delta = 0$ and $\beta = 2\beta_T = 4/3$ (*triangles*), $\beta = 2$ (*squares*), $\beta = 4$ (*diamonds*). The *two intermediate lines*: $\beta = \beta_T$ and $\gamma = \delta = 0$ (*squares*); $\gamma = -\gamma_T = 1/3$, $\delta = 3\delta_T = 2/5$ (*dots*). The *two leftmost lines*: $\beta = \beta_T$, $\gamma = \gamma_T$ and $\delta = 3\delta_T = 2/5$ (*triangles*), $\delta = 6\delta_T = 4/5$ (*stars*). For all lines, $N = 2{,}047$ and $\alpha = -1$

right; correspondingly, in the *left panel* of the figure, the point where individual curves separate from the limit line moves, by growing N, to lower values of ε. Such an N dependence clearly shows that the occurrence of exponentially long times is, in the case presently analyzed, a finite-size effect. The crossover from stretched exponential to power-law is not a peculiarity of the $\alpha + \beta$ model: we shall see below that in the pure β model it is even more evident. Concerning the value b of the exponent in the stretched exponential, $b = 1/8$ provides the best interpolation, although close values such as $b = 1/9$ are also compatible; $b = 1/8$ is significantly one half of the value $b = 1/4$ found in [10] for the β model, that we shall also find below in our analysis of the β model.

3.3 The Role of β

The value of the exponent a turns out to be independent of β, as far as β is different from the Toda value $\beta_T = 2/3$. This was checked by computing T_{eq} as function of ε at sufficiently large $N = 2{,}047$ (so that the thermodynamic limit, in the exploited range of ε, is practically achieved), for different values of β, namely $\beta = 2\beta_T = 4/3$, 2 and 4. The result is reported in Fig. 5, see the *three rightmost curves*; the computed values of the slope turns out to be $a \simeq 2.25$ for all such values of β, within 1%.

The power-law changes, however, if $\beta = \beta_T$: in this case indeed we find $a = 3$, within about 1%; see in the figure the line interpolating the *black squares*. As is remarkable, the same power law is found for the generalized FPU potential (8), whenever $\beta = \beta_T$ and $\gamma \neq \gamma_T$. For example, the line interpolating the dots in Fig. 5 refers to $\gamma = -\gamma_T = 1/3$ and $\delta = 3\delta_T = 2/5$ (as already remarked, $\delta > 0$ is necessary to avoid instability). In this case too, the computed exponent is 3 within 1%. It is then natural to continue the investigation, by considering models tangent to Toda one order up, i.e. with $\gamma = \gamma_T$, too, and $\delta \neq \delta_T$. This was done for $\delta = 3\delta_T$ and $6\delta_T$, and the result are the *two leftmost lines* in Fig. 5. The exponent now is $a = 4$, within the same precision. For all curves, error bars in principle

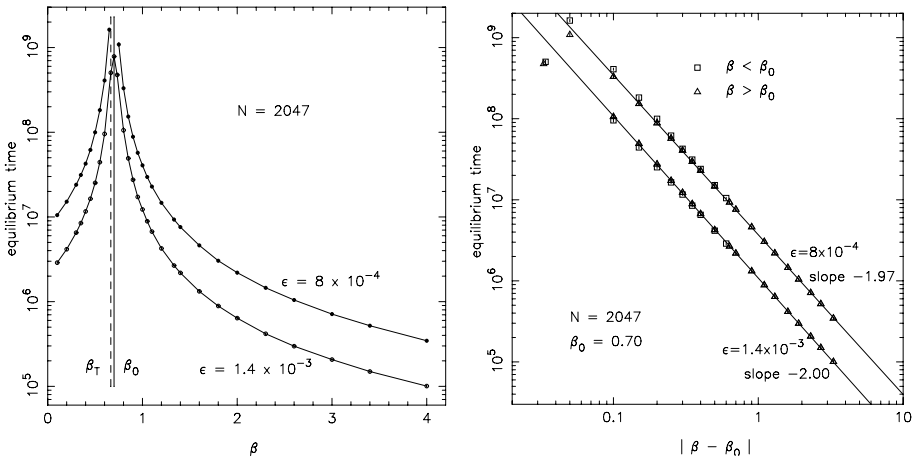

Fig. 6 *Left*: the averaged equilibrium time $T_{\rm eq}$ as function of β, for $N = 2{,}047$ and $\varepsilon = 1.4 \times 10^{-3}$, semi-log scale. *Right*: same quantity as function of $|\beta - \beta_0|$, $\beta_0 = 0.7$; *squares*: $\beta < \beta_0$, *triangles*: $\beta > \beta_0$

are reported in the figure, but are smaller than the symbols, and so almost invisible. It is worthwhile to stress that $\beta \neq \beta_T$ is the only case for which the exponent a is noninteger. We did not study the further case $\delta = \delta_T$; we guess that to a tangency of order $s \geq 4$ there corresponds an exponent $a = s - 1$.

We found it interesting to investigate more carefully the (commonly neglected) role of β in the $\alpha + \beta$ model, by looking at the dependence of $T_{\rm eq}$ on β at fixed ε. Figure 6, *left panel*, shows $T_{\rm eq}$ as function of β, for $\varepsilon = 1.4 \times 10^{-3}$ and $N = 2{,}047$, in semi-log scale. The figure shows a very pronounced peak (recall the vertical scale is logarithmic) close to β_T; quite clearly, even models with β close to zero—including the pure α model—appear here as perturbations of Toda. The center of the peak is not exactly at β_T, but at $\beta_0 \simeq 0.7$, and this is likely due to the fact that, for β close to β_T, the higher order terms get also important. The *right panel* of the figure shows $T_{\rm eq}$ as function of $|\beta - \beta_0|$, this time in log–log scale. The result is with good approximation a line of slope -2, representing a law of the form

$$T_{\rm eq} \sim (\beta - \beta_0)^{-2}. \tag{10}$$

The *line* in the figure is a least-squares fit of all points with $\beta > \beta_0$ but the leftmost one, and remarkably, the same *line* reasonably fits also most points with $\beta < \beta_0$. The exceptional points are the closest to β_0 and thus to the Toda value β_T; it is not surprising that for $\beta \simeq \beta_T$, the higher order terms in the potential become important and the law (10) fails. The small difference between β_0 and β_T looks like a phenomenological way to keep into account the effect of such terms, and one can conjecture it would disappear by repeating computations with $V(r) = V_T(r) + (\beta - \beta_T) r^4$ in place of the FPU potential (2). Keeping into account the power-law (7) with $a = 9/4$, the resulting law for $T_{\rm eq}$, in the thermodynamic limit, is

$$T_{\rm eq} = C\, \varepsilon^{-1/4} \big[(\beta - \beta_0)\varepsilon\big]^{-2}, \quad \beta_0 \simeq \beta_T, \tag{11}$$

with some positive constant C.

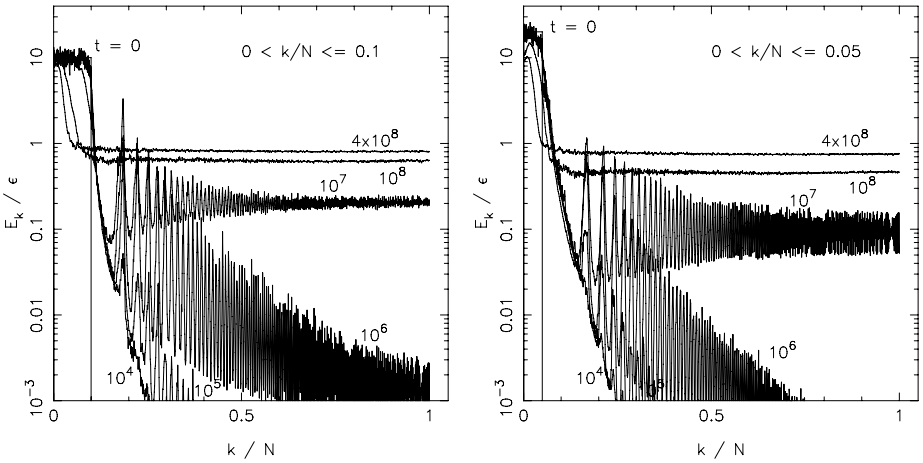

Fig. 7 The averaged energy profile at different times T, for the pure β model at $\varepsilon = 8 \times 10^{-4}$ and $N = 1{,}023$; initially excited modes: $0 < k/N \leq 0.1$ (*left*) and $0 < k/N \leq 0.05$ (*right*)

4 Results for the Pure β Model

4.1 Relevance of the Initial Excitation

The substantial irrelevance of the initial excitations (provided only low modes are excited), that we observed in the $\alpha + \beta$ model (Figs. 1 (*left*) and 3), no longer holds for the pure β model. Figure 7 shows the energy profile of the pure β model at selected times, at $\varepsilon = 8 \times 10^{-4}$ and $N = 1{,}023$, for initial excitation of the modes with $0 < k/N \leq 0.1$ (*left*) and $0 < k/N \leq 0.05$ (*right*). Although the presence of two time scales is still visible in both cases, it is evident that the metastable profile is not the same, and some difference is also present in the route to equipartition. But the difference becomes especially relevant, if the frequency of the initially excited modes is bounded away from $\omega = 0$. Figure 8 refers to N and ε as above, but now

$$\frac{1}{64} \leq \frac{k}{N} \leq \frac{5}{64}, \tag{12}$$

as in Refs. [8, 10] and (for the $\alpha + \beta$ model) as in Fig. 3 (*right*). Quite clearly, see the detail in the *right panel* of the figure, in the short time scale energy does not flow backward so as to fill the gap between $k/N = \frac{1}{64}$ and zero, while the road to equipartition is different and slower. As we shall see, such a difference will be quite important for the equipartition time too, in particular in the limit of large N. Let us remark that the lack or presence of an effective backward flow in the low mode region can affect the dynamics of high modes, by enhancing or inhibiting specific energy transfers from low to high frequencies, thus being relevant for the rate of approach to equilibrium.

4.2 The Equilibrium Times

Figure 9 shows the behavior of the averaged equilibrium time T_{eq} as function of the specific energy ε, for different values of N between 127 and 32,767. The initial datum, as is now important, is energy equipartition among modes with $0 < k/N \leq 0.1$. The figure should be

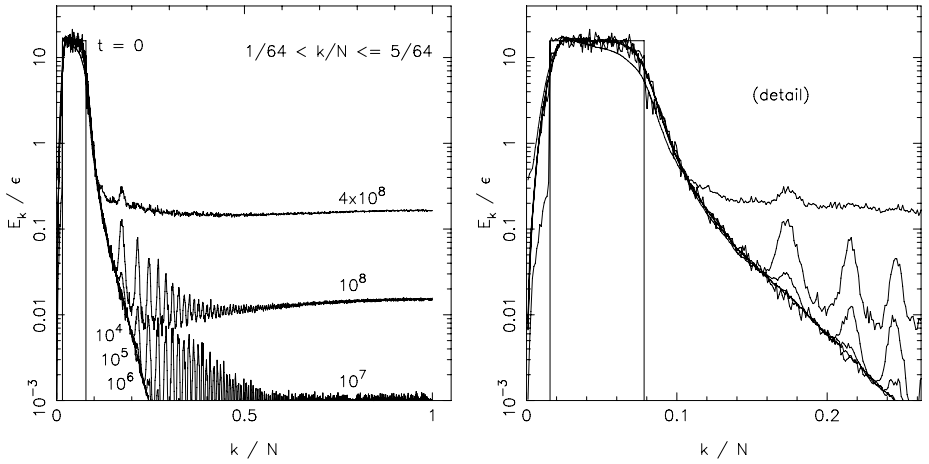

Fig. 8 *Left*: as in Fig. 7, but initial excitation of modes with $\frac{1}{64} \leq k/N \leq \frac{5}{64}$. *Right*: a zoom on the small k/N region

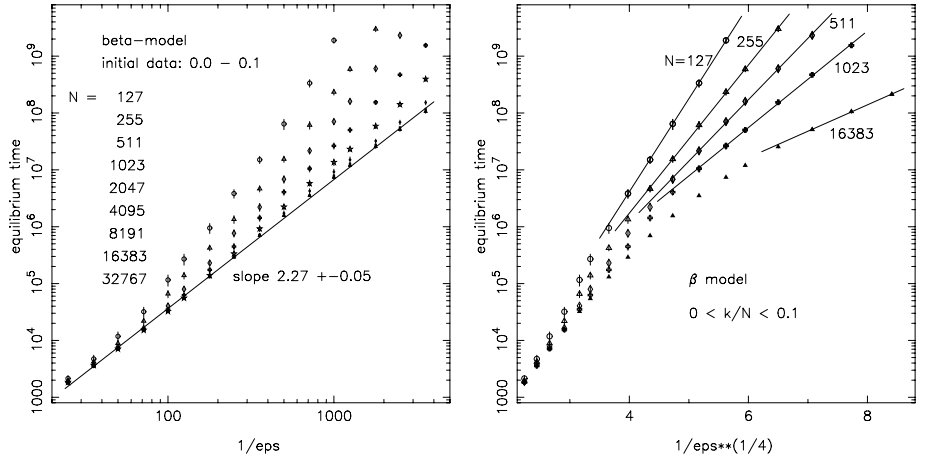

Fig. 9 *Left*: the equilibrium time $T_{\rm eq}$ as function of ε^{-1}, in log–log scale, for the pure β model; $N = 127$, 255, 511, 1,023, 2,047, 4,095, 8,191, 16,383, 32,767. *Right*: the same quantity as function of $\varepsilon^{-1/4}$, in semi-log scale, for $N = 127$, 255, 511, 1,023, 16,383. Initially excited modes: $0 < k/N \leq 0.1$

compared with Fig. 4, which instead refers to the $\alpha + \beta$ model. The behavior is clearly very similar: for large N (*left panel*) all curves flatten on a line, that is, for the pure beta model too, in the thermodynamic limit, $T_{\rm eq}$ obeys a power law. Quite significantly, the exponent looks the same as for the $\alpha + \beta$ model; the best numerical result is indeed $a = 2.26 \pm 0.05$. For small N, similarly to the $\alpha + \beta$ model, individual curves follow a stretched exponential (*right panel*). The best choice of the exponent is now $1/4$ (as in [10]) rather than $1/8$.

Figure 9 shows with greater evidence than Fig. 4 that the study of the thermodynamic limit is not trivial. Let us use here the self-evident notation $T_{\rm eq}(N, \varepsilon)$; quite clearly, to work out the behavior of the model in the thermodynamic limit, it is not enough to select a large $\overline{N}$, and to study $T_{\rm eq}(\overline{N}, \varepsilon)$ as function of ε: unavoidably, by lowering ε, this introduces finite-

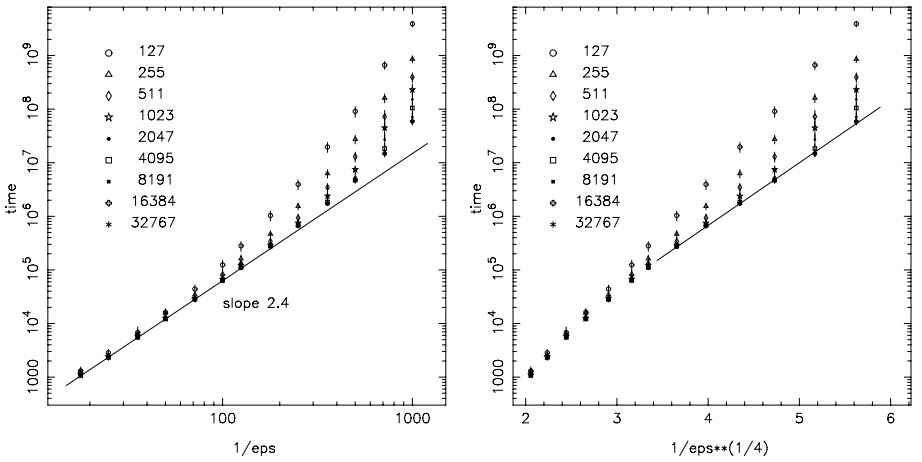

Fig. 10 Same as Fig. 9, but initial excitation of modes with $\frac{1}{64} < k/N \leq \frac{5}{64}$

size effects. It is instead mandatory, for each ε, to preliminarily grow N till $T_{\rm eq}(N, \varepsilon)$ practically reaches the limit $T_{\rm eq}(\infty, \varepsilon)$; this requires N to be larger and larger, for small ε. This is very expensive numerically, if one wants to reach small values of ε, so as to distinguish a power law from an exponential, but cannot be avoided: the two limits $N \to \infty$ and $\varepsilon \to 0$ definitely do not commute.

At variance with the $\alpha + \beta$ model, the thermodynamic limit of the pure β model depends on the initial data (even if low modes only are excited), more precisely it is sensitive to the fact that the frequency of the excited modes is bounded away from zero. Figure 10 refers to initial data with k/N as in (12) and shows, in the usual way, $T_{\rm eq}$ as a function of ε for selected values of N, ranging as before from 127 to 32,767. The difference with respect to Fig. 9, *left panel*, is evident: the thermodynamic limit, in the explored range of ε, has been reached, but this is no longer a power law. The *right panel* of the figure, where $T_{\rm eq}$ is plotted vs. $\varepsilon^{-1/4}$ in log–log scale, shows that for sufficiently small ε, even in the thermodynamic limit, $T_{\rm eq}(\varepsilon)$ is a stretched exponential. This is precisely the law obtained in [10] (by observing $N = 511$), that we here confirm.

Figure 10 (*left*) shows that for not too small specific energy ($\varepsilon \geq 10^{-2}$) the equilibrium time follows a power of $1/\varepsilon$, with exponent 2.4. In a similar range of energies, a power law with exponent 3 was instead obtained in [8]. Such a difference is due to the different indicators chosen to measure the equilibrium time: indeed, as explained in Sect. 2.2, our parameter $\mathcal{E}$ is only sensitive to the growth of the tail ($k > N/2$), while the usual spectral entropy considered in [8] depends on all modes (and is mainly sensitive, in our experience, to the redistribution of energy among low modes, which is not the same phenomenon). We confirm the validity of the exponent 3, in that range of ε, if the usual indicator is chosen.

4.3 The Crossover from the $\alpha + \beta$ Model to the Pure β Model

Let us reconsider the expression (11) of $T_{\rm eq}$ for the $\alpha + \beta$ model. Such an expression was obtained for $\alpha = -1$, but taking into account that the only relevant parameters in the problem are $|\alpha|\varepsilon^{1/2}$ and $\beta\varepsilon$, the expression naturally generalizes to

$$T_{\rm eq}^{(\alpha\beta)}(\alpha, \beta, \varepsilon) = C|\alpha|^{-1/2}\varepsilon^{-1/4}\left[\left(\beta - c\alpha^2\right)\varepsilon\right]^{-2}, \quad c \simeq \frac{2}{3}. \tag{13}$$

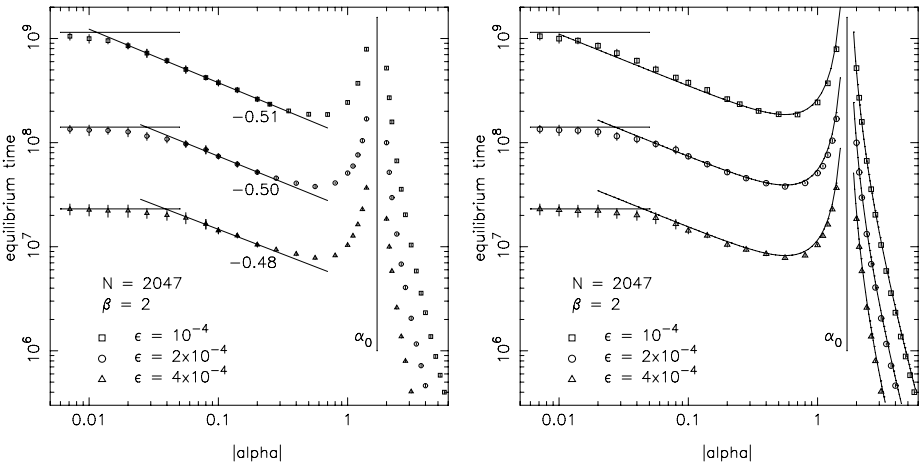

Fig. 11 A test of expression (13), also showing the crossover from the $\alpha+\beta$ to the pure β model. Parameters: $\beta = 2$, $\varepsilon = 10^{-4}$ (*squares*), 2×10^{-4} (*circles*), 4×10^{-4} (*triangles*). Both panels show $T_{\rm eq}^{(\alpha\beta)}$ vs. $|\alpha|$, in log–log scale. *Left*: the *lines with slope* (approximately) -0.5 show the factor $|\alpha|^{-1/2}$ in front of (13); the *horizontal lines* provide the value of $T_{\rm eq}^{(\beta)}$ of the pure β model, so as to show the crossover. *Right*: same data; the *continuous line* represents (13), with C chosen so as to fit exactly one point (same C for the three curves)

It is interesting to test such an expression by moving α to zero, at fixed β and ε; indeed, in this limit the content of the square bracket goes to a constant, so it is possible to isolate and test separately the presence of the factor $\alpha^{-1/2}$ in front of it. Expression (13) has however a problem: the divergence for $\alpha \to 0$ cannot be true, since the limit of the model for $\alpha \to 0$ is not an integrable model, but the pure β model (similarly, for $\beta \to \beta_0$ we do not have a real divergence but only a peak). Practically, for $\alpha \to 0$ we should expect a crossover from expression (13) to the expression found for the pure β model, namely

$$T_{\rm eq}^{(\beta)}(\beta, \varepsilon) = D\varepsilon^{-1/4}(\beta\varepsilon)^{-2},$$

with a suitable constant D.

Figure 11 shows the dependence of $T_{\rm eq}^{(\alpha\beta)}(\alpha, \beta, \varepsilon)$ on $|\alpha|$ (we are using negative α) for $N = 2{,}047$, $\beta = 2$ and three different values of ε, namely $\varepsilon = 10^{-4}, 2 \times 10^{-4}, 4 \times 10^{-4}$, in log–log scale. In the *left panel* one can observe (i) the peak when $|\alpha| = \sqrt{\beta/c} \simeq 1.7$; (ii) an interval where the factor $\alpha^{1/2}$ in front of (13) appears rather clearly (the computed slope, top to bottom, is 0.51, 0.50, 0.48; *solid symbols* indicate the interval used to fit the slope) and (iii) the crossover to the constant value $T_{\rm eq}^{(\beta)}(\beta, \varepsilon)$, represented in the figure by *horizontal lines*.

The *right panel* shows the same numerical points, but in place of the two interpolating lines, useful to observe the crossover, we plotted the curve (13), with C (the same constant for the three curves) chosen so as to fit exactly one particular point in the center of the figure ($\varepsilon = 2 \times 10^{-4}$, $|\alpha| = 0.1$). Although minor discrepancies are present, it looks rather evident that expression (13) is substantially correct. We admit we have been a little surprised: expression (13) was guessed just as a possible match between two asymptotic laws for $T_{\rm eq}$, namely the $\varepsilon^{-9/4}$ behavior and the $(\beta - \beta_0)^{-2}$ behavior around $\beta = \beta_0$; it is not obvious it holds so generally. Understanding (13) theoretically looks a noneasy challenge.

5 Concluding Remarks and Open Problems

5.1 Integrability, but Loss of Memory of Initial Conditions?

The behavior of the $\alpha + \beta$ model might appear paradoxical: in the short time scale, as we remarked (and as is well known in the literature), the model cannot be distinguished from an integrable one, nevertheless both the metastable state and the subsequent path to equipartition look independent of the particular initial data (provided only low modes are excited, with random phases). To see there is, at least, no contradiction, let us make the example of a linear chain, observed however in the "wrong" coordinates, namely in the particle coordinates rather than in the normal modes coordinates. The model is obviously integrable. Nevertheless, if we excite randomly a given small subset of particles, typically all normal modes will be excited. The phases of modes are not at all random (antitransforming, only a few particles will be out of the rest position). But due to the ergodicity of the linear motion of the phases on the N-torus, the phases of the modes get randomized (this is a second randomization, of dynamical nature, independent of randomness in the initial excitation). After this happens, if we antitransform, we will find that all particles got excited, and the state will appear to be independent of the particular initial datum, in spite of the integrability of the model; the trivial but crucial point is that an excitation localized in the particles, is not localized in the modes, and conversely. A reasonable conjecture (for which we are indebted to D. Ryabov) is that for the $\alpha + \beta$ model, *as far as low frequency modes are initially excited,* there is a similar relation between the linear modes (corresponding to the particles, in the above example) and the Toda modes (corresponding, in the example, to the modes): no matter which set of linear modes are excited, all Toda modes *up to a cut-off depending on the specific energy* get excited; some time later, when the phases of the excited Toda modes get randomized by the linear dynamics on their torus, the energy spectrum of the linear modes will become practically independent of the particular initial excitation, as we did observe. Quantitatively, to have a similar universality, the fraction of modes which are initially excited needs to be smaller than the width of the metastable profile reported in Fig. 1, which is proportional to $\varepsilon^{1/4}$ [9]. (We did not investigate the case in which the latter fraction is comparable to or even larger than $\varepsilon^{1/4}$, where a completely different phenomenology might occur.)

5.2 What About the β Model?

All of the above considerations concern the $\alpha + \beta$ model. For the β model, instead, the situation is by far less clear. By analogy with the $\alpha + \beta$ model, one should expect that if a metastable state does exist, its profile must have a width proportional to $\sqrt{\varepsilon}$ (the nonlinear coupling constants of the $\alpha + \beta$ and β models are indeed $\alpha\sqrt{\varepsilon}$ and $\beta\varepsilon$, respectively). However, a discrete integrable model approximating (tangent to) the β model, as the Toda model is for the $\alpha + \beta$ model, is not known, and this might be very relevant. As a matter of fact, separating integrable short time phenomena from chaotic long time ones, in the β model is problematic. In particular for the β model, with initial excitation touching $k = 0$, we could never identify a metastable profile somehow independent of the initial excitation: even for very narrow initial excitation and not too small ε (for example: k/N down to 5×10^{-3}, ε up to 10^{-2}), the short time evolution keeps strong memory of the width of the initial excitation, and the time of approach to equilibrium, at variance with the $\alpha + \beta$ model, also depends on it. Understanding the difference between the β and the $\alpha + \beta$ model looking at their best integrable approximations (possibly including nonlinear wave equations) is, in our opinion, an interesting open problem.

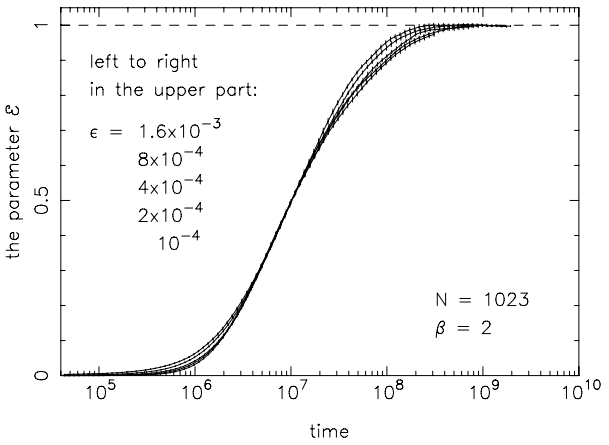

Fig. 12 The sigmoids appearing in Fig. 2 are here horizontally translated, in such a way the coincide at $\mathcal{E} = 0.5$ (the one at $\varepsilon = 4 \times 10^{-4}$ is kept fixed). Decreasing values of ε from left to right, in the *upper part* where $\mathcal{E}$ approaches one

5.3 A Third Time-Scale?

Coming back to the $\alpha + \beta$ model, a nontrivial fact, in our opinion, is that the long time evolution towards energy equipartition (which involves the tail, i.e. modes far beyond the acoustic resonance) also looks independent of the initial data. As a matter of fact, before such evolution takes place, in a third time scale, intermediate between the formation of the metastable state ($T \sim \varepsilon^{-3/8}$, according to [5]) and the subsequent growth of the tail (our $T_{\rm eq} \sim \varepsilon^{-9/4}$), some chaotic randomization occurs. This can be observed by computing the Lyapunov exponents of the model: preliminary computations, not reported here, apparently indicate that the Lyapunov time $T_L = \lambda_{\rm max}^{-1}$, where $\lambda_{\rm max}$ is the maximal Lyapunov exponent, is intermediate; tentatively, $T_L \sim \varepsilon^{-3/2}$. The presence of some chaoticity is important, in our opinion, to possibly interpret the intermediate state as an effective metastable state, relevant for a short-time statistical mechanics.

5.4 A Further Time-Scale? Criticism on the Notion of Equilibrium Time

Throughout the paper we used the expression "equipartition time", and even "equilibrium time". However, what we really observe are states which got far from the initial state—the energy of the tail has grown to one half of the due amount—but are still far from equilibrium: *very* far, if one looks at the microcanonical probability of similar states, which represent incredibly large fluctuations with respect to equilibrium.

As already remarked, the chosen threshold for $\mathcal{E}$ to define the equipartition time (actually $\mathcal{E} = 0.5$) is fairly irrelevant. Indeed if, similarly to [8], we rigidly shift the sigmoids of Fig. 2 in such a way they coincide at $\mathcal{E} = 0.5$, they do superimpose in a rather wide interval, see Fig. 12. From the figure it is not clear if, asymptotically for small ε, the curves converge to a limit curve in a wider interval, in particular in the *upper part*, when $\mathcal{E}$ approaches one. We guess yes, but we admit there is no clear evidence, at least not within the level of precision adopted all over the paper. Should this be the case, it would mean that the time scale we find, more precisely its ε dependence, provides a complete picture of the approach to equilibrium, including the last part of the path, when the system is close to equilibrium.

On the other hand, the time scale for equilibrium, in principle, should be determined by producing small fluctuations with respect to equilibrium, looking then at their correlation time $T_{\rm corr}$. Some numerical studies in this direction do exist in the literature, see in particular [21]. In such a reference the presence of long term correlations is evident, but due to

computer limitations, a quantitative study of the dependence of $T_{\rm corr}$ on ε—whether it is a power law or a stretched exponential—could not be performed. There is a chance that $T_{\rm corr}$ represents a fourth, possibly longer, time-scale in the problem. Such a possibility was outlined, with some emphasis, by our colleagues of Milan in private discussions, as a possible way out to reconcile the relatively short equilibrium times we find here, with the point of view expressed, for example, in [4, 22]. This is obviously an open question.

5.5 Any Hope to Understand Theoretically?

Understanding theoretically the long time scale of the FPU problem, as is well known, is a quite difficult task. As was shown (after hard work) in [23], KAM theory does apply, but unfortunately, the strong N-dependence of the threshold in ε for the existence of quasiperiodic motions, makes the approach useless in the thermodynamic limit. To our knowledge, very few perturbation results do exist in the thermodynamic limit, and none of them concerns the usual FPU models. For example, in [24] exponential estimates for the equilibrium times have been produced, for an infinite chain of rotators at nonvanishing specific energy; the initial data, however, need to be such that only a finite number of different frequencies is effectively present. The rotators model has been also studied in [25], with a relevant novelty: a suitable "Gibbs norm" replaces the common supremum norm of perturbation theory; the use of such a norm allowed to perform a few perturbation steps uniformly in N, for generic initial data. The same idea has been exploited very recently in [26], in connection with a φ^4-model in which both the nonlinearity and the nearest-neighbor coupling are small (so as the model is a perturbation of a model with just one frequency); for such a system the authors could prove the existence of an adiabatic invariant, which remains almost constant up to an exponentially long time-scale. The natural question is whether similar results are expected to hold for FPU models, too. Our numerical results, namely the power law $\varepsilon^{-9/4}$ for the equilibrium times in the thermodynamic limit, provide in our opinion an indication (valid at least for the nongeneric FPU-like initial data we are using) that some essential obstruction at finite order might occur in the perturbation construction, which prevents reaching the optimal perturbation order and thus getting exponentially long times. The right face of the medal is that working out the appropriate law for the equilibrium time in the thermodynamic limit could require only a finite (small) number of perturbation steps. All of this of course is very conjectural.

5.6 Is Dimension One Peculiar?

In two papers by one of us [27, 28], some two dimensional versions of the FPU model have been studied; as a result, power laws ε^{-a} have been found for the equilibrium times, and the too quick conclusion was drawn—essentially based on the comparison with [10], where exponentials are made evident—that dimension two is qualitatively different from dimension one. As a matter of fact, the behavior of $T_{\rm eq}(N, \varepsilon)$ in dimension two, as found in [27, 28], is very similar to the one reported in Figs. 4 and 9: exponentials for finite N, which however flatten on a line, i.e. on a power law, for $N \to \infty$. The numerically computed exponent is different, and actually dependent on the border conditions: namely $a \simeq 1$ for fixed or open boundary conditions, $a \simeq 7/4$ for periodic boundary conditions. Our comment now is that some difference does exist—equilibrium times in dimension two are indeed shorter than in dimension one—but the difference does not look very essential. By the way: understanding better the difference between dimensions one and two was the occasional motivation for us, to start the study which eventually led to this paper.

Acknowledgements We feel deeply indebted to the Milano Group—in particular A. Carati, L. Galgani, A. Giorgilli, S. Paleari—for many helpful vivid discussions. We also profited of useful conversations with Heleni Christodoulidi (Padova), Denis Ryabov (Rostov-on-Don) and Bob Rink (Amsterdam).

References

1. Fermi, E., Pasta, J., Ulam, S.: Studies of non linear problems, Los-Alamos internal report, Document LA-1940 (1955). In: Enrico Fermi Collected Papers, vol. II, pp. 977–988. The University of Chicago Press, and Accademia Nazionale dei Lincei, Chicago and Roma (1965). (Also reproduced in Ref. [3])
2. Chaos focus issue: The "Fermi–Pasta–Ulam" problem—the first 50 years. Chaos **15** (2005)
3. Gallavotti, G. (ed.): The Fermi–Pasta–Ulam Problem: A Status Report. Lect. Notes Phys., vol. 728. Springer, Berlin-Heidelberg (2008)
4. Benettin, G., Carati, A., Galgani, L., Giorgilli, A.: The Fermi–Pasta–Ulam problem and the metastability perspective. In: Gallavotti, G. (ed.) The Fermi–Pasta–Ulam Problem: A Status Report. Lect. Notes Phys., vol. 728, pp. 151–189. Springer, Berlin-Heidelberg (2008)
5. Benettin, G., Livi, R., Ponno, A.: The Fermi–Pasta–Ulam problem: scaling laws vs. initial conditions. J. Stat. Phys. **135**, 873–893 (2009)
6. Fucito, E., Marchesoni, F., Marinari, E., Parisi, G., Peliti, L., Ruffo, S., Vulpiani, A.: Approach to equilibrium in a chain of nonlinear oscillators. J. Phys. (Paris) **43**, 707–713 (1982)
7. Livi, R., Pettini, M., Ruffo, S., Sparpaglione, M., Vulpiani, A.: Relaxation to different stationary states in the Fermi–Pasta–Ulam model. Phys. Rev. A **28**, 3544–3552 (1983)
8. De Luca, J., Lichtenberg, A.J., Ruffo, S.: Finite times to equipartition in the thermodynamic limit. Phys. Rev. E **60**, 3781–3786 (1999)
9. Berchialla, L., Galgani, L., Giorgilli, A.: Localization of energy in FPU chains. Discrete Contin. Dyn. Syst. **11**, 855–866 (2004)
10. Berchialla, L., Giorgilli, A., Paleari, S.: Exponentially long times to equipartition in the thermodynamic limit. Phys. Lett. A **321**, 167–172 (2004)
11. Bambusi, D., Ponno, A.: Resonance, metastability and blow-up in FPU. In: Gallavotti, G. (ed.) The Fermi–Pasta–Ulam Problem: A Status Report. Lect. Notes Phys., vol. 728, pp. 191–205. Springer, Berlin-Heidelberg (2008)
12. Zabusky, N.J., Kruskal, M.D.: Interaction of solitons in a collisionless plasma and the recurrence of initial states. Phys. Rev. Lett. **15**, 240–245 (1965)
13. Ferguson, E.E., Flashka, H., McLaughlin, D.W.: Nonlinear Toda modes for the Toda chain. J. Comput. Phys. **45**, 157–209 (1982)
14. Casetti, L., Cerruti–Sola, M., Pettini, M., Cohen, E.D.G.: The Fermi–Pasta–Ulam problem revisited: stochasticity thresholds in nonlinear Hamiltonian systems. Phys. Rev. E **55**, 6566–6574 (1997)
15. Ponno, A., Christodoulidi, H., Flach, S., Skokos, H.: The two-stage dynamics in the Fermi–Pasta–Ulam problem: from regular to diffusive behavior. Preprint (2011)
16. Livi, R., Pettini, M., Ruffo, S., Sparpaglione, M., Vulpiani, A.: Equipartition threshold in nonlinear large Hamiltonian systems: the Fermi–Pasta–Ulam model. Phys. Rev. A **31**, 1039–1045 (1985)
17. Goedde, C.G., Lichtenberg, A.J., Lieberman, M.A.: Physica D **59**, 200 (1992)
18. Yoshida, H.: Construction of higher order symplectic integrators. Phys. Lett. A **150**, 262–268 (1990)
19. Paleari, S., Penati, T.: Numerical methods and results in the FPU problem. In: Gallavotti, G. (Ed.) The Fermi–Pasta–Ulam Problem. Lect. Notes Phys., vol. 728, pp. 239–282. Springer, Berlin-Heidelberg (2008)
20. Benettin, G., Ponno, A.: On the numerical integration of FPU-like systems. Physica D **240**, 568–573 (2011). doi:10.1016/jphysd.2010.11.008
21. Carati, A., Galgani, L., Giorgilli, A., Paleari, S.: FPU phenomenon for generic initial data. Phys. Rev. E **76**, 022104/1–4 (2007)
22. Carati, A., Galgani, L., Giorgilli, A.: The Fermi–Pasta–Ulam problem as a challenge for the foundations of physics. Chaos 15 (2005)
23. Rink, B.: Proof of Nishida's conjecture on anharmonic lattices. Commun. Math. Phys. **261**, 613–627 (2006)
24. Bambusi, D., Giorgilli, A.: Exponential stability of states close to resonance in infinite-dimensional Hamiltonian systems. J. Stat. Phys. **71**, 569–606 (1992)
25. Carati, A.: An averaging theorem for Hamiltonian dynamical systems in the thermodynamic limit. J. Stat. Phys. **128**, 1057–1077 (2007)
26. Carati, A., Maiocchi, A.: Exponentially long stability times for a nonlinear lattice in the thermodynamic limit (2011). Preprint arXiv:1011.5846v1 [math-ph]

27. Benettin, G.: Time-scale for energy equipartition in a two-dimensional FPU model. Chaos **15**, 15105/1–8 (2005)
28. Benettin, G., Gradenigo, G.: A study of the Fermi–Pasta–Ulam problem in dimension two. Chaos **18**, 013112/1–13 (2008)