The Fermi–Pasta–Ulam problem
and its underlying integrable dynamics:
an approach through Lyapunov Exponents

G. Benettin[*], S. Pasquali[†], A. Ponno[‡]

Abstract

FPU models, in dimension one, are perturbations either of the linear model or of the Toda model; perturbations of the linear model include the usual β-model, perturbations of Toda include the usual $\alpha+\beta$ model. In this paper we explore and compare two families, or hierarchies, of FPU models, closer and closer to either the linear or the Toda model, by computing numerically, for each model, the maximal Lyapunov exponent χ. More precisely, we consider statistically typical trajectories and study the asymptotics of χ for large N (the number of particles) and small ε (the specific energy E/N), and find, for all models, asymptotic power laws $\chi \simeq C\varepsilon^a$, C and a depending on the model. The asymptotics turns out to be, in general, rather slow, and producing accurate results requires a great computational effort. We also revisit and extend the analytic computation of χ introduced by Casetti, Livi and Pettini, originally formulated for the β-model. With great evidence the theory extends successfully to all models of the linear hierarchy, but not to models close to Toda.

[*]Università di Padova, Dipartimento di Matematica "Tullio Levi-Civita", Via Trieste 63, 35121 Padova (Italy); benettin@math.unipd.it

[†]Università di Roma Tre, Dipartimento di Matematica e Fisica; spasquali@mat.uniroma3.it

[‡]Università di Padova, Dipartimento di Matematica "Tullio Levi-Civita", Via Trieste 63, 35121 Padova (Italy); ponno@math.unipd.it

1 Introduction

1.1 The purpose and the considered models

This paper is devoted to the Fermi-Pasta-Ulam (FPU) problem [1], precisely to a numerical study of the maximal Lyapunov exponent χ in different one-dimensional FPU models. Numerical results will be also critically compared with the existing theory. The Hamiltonian is

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

V being some nearest-neighbours potential with single minimum in zero.

Two integrable models of this form are known, namely the linear model with purely quadratic potential $V_L(r) = \frac{1}{2}r^2$, which trivially separates into independent normal modes, and the Toda model [2] [3] [4] with potential

$$V_T(r) = \mu^{-2}(e^{\mu r} - 1 - \mu r) = V_L(r) + \mathcal{O}(\mu r^3) ; \tag{2}$$

no other FPU-like integrable systems do exist, see [5].

As in most papers on the subject, we are mainly interested in the dynamics at small energy per particle $\varepsilon = E/N$, paying however as much attention as possible to the asymptoticity in N at fixed ε (that is, ideally, to the thermodynamic limit).

The FPU models we are going to consider include the two most studied ones, namely:

i. The so-called "$\alpha+\beta$" model, with potential

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

The model is generally thought of as a perturbation of the linear model, the cubic and quartic term providing a small coupling between normal modes. The nonlinearity (the mode coupling) is given by the two quantities

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

at small ε the former dominates and correspondingly the distance to the linear model is roughly (forgetting constants)

$$\text{dist}\,(\alpha + \beta, \text{Lin}) \sim |\alpha|\sqrt{\varepsilon} .$$

The $\alpha + \beta$ model, however, is tangent up to the order three to the Toda model with $\mu = 2\alpha$, the distance being now

$$\text{dist}\,(\alpha + \beta, \text{Toda}) \sim |\beta - \beta_T|\varepsilon , \qquad \beta_T = \frac{2}{3}\alpha^2 .$$

The relevance of the underlying Toda dynamics in the $\alpha + \beta$ FPU model was observed and emphasized already in 1982 by Ferguson, Flaschka and McLaughlin in ref. [6] (a paper that we consider among the fundamental ones, after the original FPU paper).

ii.   The symmetric "pure β" model, with potential

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

Its closest integrable approximation is the linear model and

$$\mathrm{dist}\,(\text{pure}\ \beta, \mathrm{Lin}) \sim \beta \varepsilon .$$

Models with a higher order contact with Toda have been taken into consideration, as far as we know, in [7] and in a more recent paper by two of us [8]. Such models are:

iii.   The $\alpha + \beta$ model, with however $\beta = \beta_T = \frac{2}{3}\alpha^2$. We shall call it the "$\beta_T$" model. Quite trivially,

$$\mathrm{dist}\,(\beta_T, \mathrm{Toda}) \sim |\gamma_T| \varepsilon^{3/2} , \qquad \gamma_T = \frac{1}{3}\alpha^3 .$$

iv.   The "γ_T" model, with

$$V_{\gamma_T}(r) = \frac{r^2}{2} + \alpha \frac{r^3}{3} + \beta_T \frac{r^4}{4} + \gamma_T \frac{r^5}{5} + \delta \frac{r^6}{6} , \qquad 0 < \delta \neq \delta_T = \frac{2}{15}\alpha^4$$

(the term of degree 6, with positive coefficient, ensures stability). Roughly,

$$\mathrm{dist}\,(\gamma_T, \mathrm{Toda}) \sim |\delta - \delta_T| \varepsilon^2 .$$

A corresponding hierarchy of models closer and closer to the linear model can obviously be produced; for completeness, and to test the existing theory, we investigated two of them, namely

v.   The "$\gamma + \delta$" model, with potential

$$V_{\gamma + \delta}(r) = \frac{r^2}{2} + \gamma \frac{r^5}{5} + \delta \frac{r^6}{6} , \qquad \alpha \neq 0, \quad \delta > 0 ; \tag{5}$$

it is obviously

$$\mathrm{dist}\,(\gamma + \delta, \mathrm{Lin}) \sim \gamma \varepsilon^{3/2} .$$

vi.   The "pure δ" model, with potential

$$V_\delta(r) = \frac{r^2}{2} + \delta \frac{r^6}{6} , \qquad \delta > 0 , \tag{6}$$

and correspondingly

$$\mathrm{dist}\,(\text{pure}\ \delta, \mathrm{Lin}) \sim \delta \varepsilon^2 .$$

What is lacking in such a landscape of nearly integrable models is a model distant both from the linear and from the Toda model by cubic terms in V. The one we suggest here is

vii.   A generalization of $\alpha + \beta$, with site-dependent potential

$$V_i(r) = \frac{r^2}{2} + \alpha_i \frac{r^3}{3} + \beta \frac{r^4}{4}$$

(see later for the choice of α_i). We shall call it the "variable α" model. The distances are

$$\mathrm{dist}\,(\text{var.}\ \alpha, \mathrm{Lin}) \sim \mathrm{dist}\,(\text{var.}\ \alpha, \mathrm{Toda}) \sim \alpha \sqrt{\varepsilon} ,$$

α denoting here, roughly, the size of the α_i's and of their oscillation.
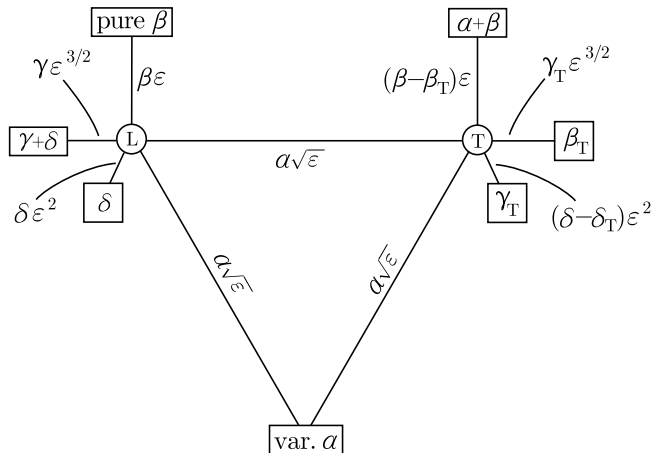
Figure 1: *The different FPU models studied in this paper and their rough distances to the linear (L) and to the Toda (T) models.*

Models i-vii form a kind of constellation around the two integrable models, which is symbolically represented in figure 1. The aim of this paper is to explore such a constellation by means of the most flexible tool, sensitive to the lack of integrability and independent of the choice of special coordinates such as the normal modes (as is especially important around Toda), namely the computation of the maximal Lyapunov exponent. The paper, in a sense, is a continuation of [8] and [9] (which in turn, in their aim, are continuations of [6]).

Remark. For the $\alpha + \beta$, β_T and γ_T models, the value of α, if different from zero, is irrelevant and simply fixes the energy scale: models with the same ratios β/α^2, γ/α^3, δ/α^4... have indeed identical dynamics up to trivial rescalings. We always used $\alpha = -1$; correspondingly, $\beta_T = 2/3$, $\gamma_T = -1/3$, $\delta_T = 2/15$. (Negative α means the "springs" between masses get stiff when compressed, as in typical molecular potentials.) Similar considerations hold for the leading nonlinearity constant of any of the above models.

1.2 Main connections with the literature

A. General references on FPU. The literature on FPU, over more than 60 years, is so huge, that it is not even conceivable to summarize it here. The state of the art, updated to 2005–06, can be found in two collections of papers [10] [11], but it is hard—and this is a good indication of the difficulty of the problem—to extract, from all of these papers or partial reviews, a unitary view, free of contradictions. In such an uncommon situation, we can only indicate which are the main references that have been important for us, to form our personal view of the FPU problem, and should be considered at the basis of the present work. (We are aware that the list is incomplete and important papers are possibly missing.)

After the original FPU paper, we wish to mention [12], where the presence of longer time recurrencies was established, and some early papers like [13] [14] [15] [16], raising and discussing, with different perspectives, the very crucial question about the persistence of the FPU paradox in the thermodynamic limit.

A double turning point occurred in 1982: on the one hand, in the already emphasized reference [6], evidence was provided that the FPU results (in particular, the partial energy sharing among low
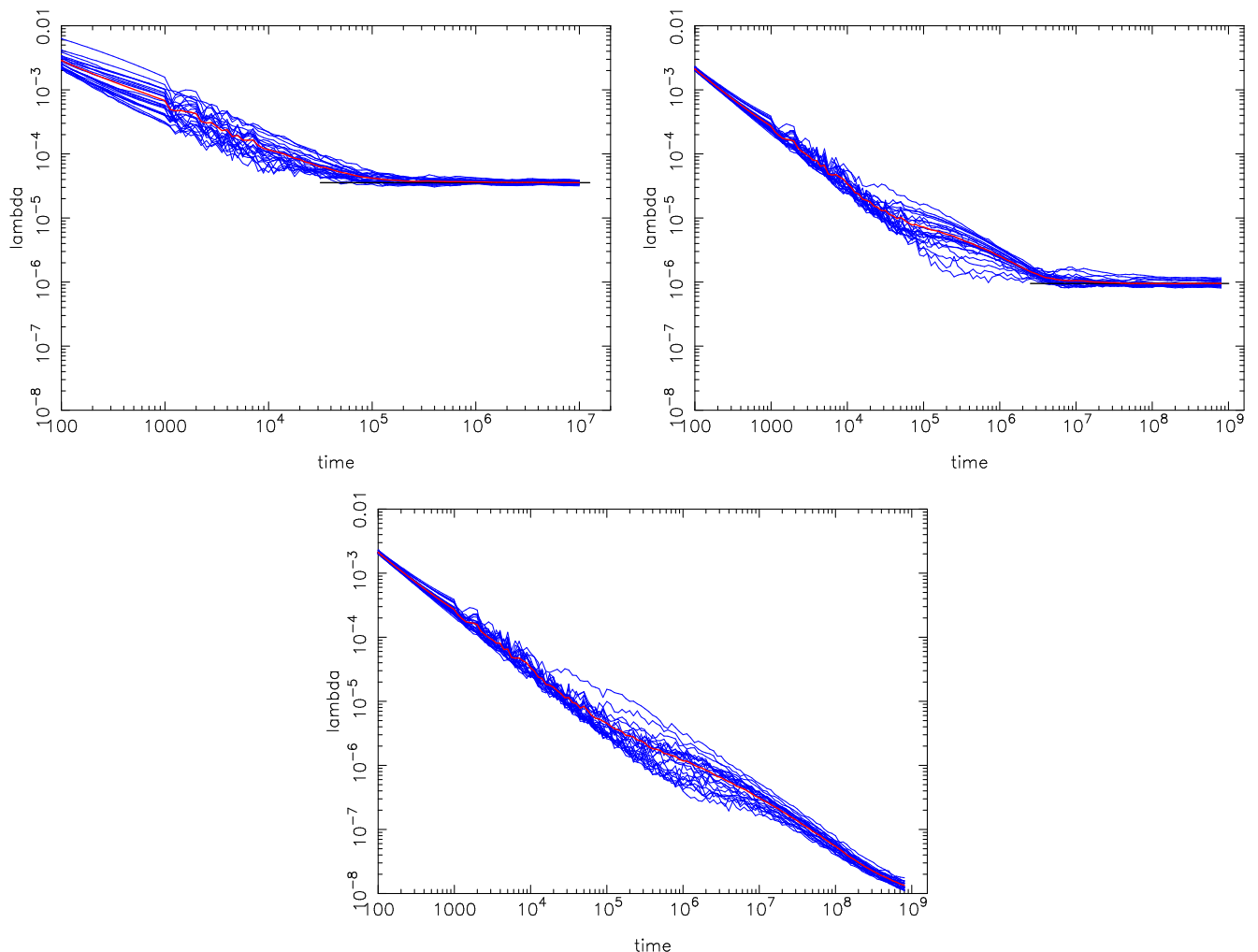
Figure 2: *Upper panels:* $\lambda(x,t)$ *for the* $\alpha + \beta$ *model with* $N = 1024$, $\alpha = -1$ *and* $\beta = 2$, *at* $\varepsilon = 10^{-3}$ *(left) and* 10^{-4} *(right); 24 random choices of the initial datum* x *(blue) and their average* $\overline{\lambda}(t)$ *(red). The black horizontal segment is useful to appreciate the reaching of the limit ??). Lower panel: same quantities for the Toda model, with* $N = 1024$, $\varepsilon = 10^{-4}$.

modes in the $\alpha + \beta$ model) are fully explained by the underlying integrable Toda dynamics; on the other hand, in two similarly crucial papers [17] [18], the so-called two-time scales (or metastability) scenario was introduced: in essence, on a first relatively short time scale the phenomena observed by FPU (essential lack of ergodicity and mixing, quasi-periodicity) occur, while on a much larger time scale the normal statistical behavior is recovered. The role of metastability was discussed and emphasized in later papers like [19] [20] [21].

We expressed our view on the one-dimensional FPU problem in [22], [8] and [9]. In short:

– For models of the Toda hierarchy, i.e. $\alpha + \beta$, β_T and γ_T, the ideas introduced in [6] (the under-lying integrable Toda dynamics) and in [17] [18] (the two time–scales) naturally combine in a unitary quite elementary picture: in the shorter time scale of the metastability scenario the Toda actions stay almost constant, while the angles turn on tori producing a partial averaging (producing in particular the FPU state, if only a few low frequency modes are excited); on

larger times the actions slowly diffuse throughout the phase space, with small speed for small ε, producing eventually complete microcanonical averaging. The puzzling special behavior of low modes in FPU, during the short time scale, is due to a lack of correspondence between linear modes and Toda actions, when the index k of the mode is, roughly, smaller than $\varepsilon^{1/4}N$.

– The thermodynamic limit, in all FPU models, is highly nontrivial: the limits $\varepsilon \to 0$ and $N \to \infty$ do not commute in an essential way, the asymptotics in N becoming slower and slower at small ε. Correspondingly, great attention must be paid in numerical computations.

– In the thermodynamic limit, for small ε, all time scales, including the large equilibrium ones, are in general (i.e., with the exception of special initial conditions [20] [8]) power laws ε^{-a}, the exponent a depending on the degree of contact with the nearest integrable model, linear or Toda. The computed values of the exponents [8] are compatible with, and conjectured to be, $a = 9/4$ for the $\alpha + \beta$ and for the pure β models, $a = 3$ for the β_T model and $a = 4$ for the γ_T model.

B. Specific references on the maximal Lyapunov exponent. Let us now draw the attention to some keynote papers in which Lyapunov Exponents are used to investigate the FPU problem. Here too, the literature is rather abundant, and we limit ourselves to mention the few papers which are more directly connected with the present work.

After the pioneering paper [23], the first explicit reference to Lyapunov exponents in connection with FPU seems to be [24]. Refs. [25] and [26] are devoted to the pure β model (as well as to the so-called φ^4 model), with N between 32 and 128; a crossover between a weak and a strong stochasticity regime is there made evident, and in the weak stochasticity regime, at small ε, the power law $\chi \sim \varepsilon^2$ is found. Ref. [27] contains instead a computation of χ in the pure α model (i.e. $\alpha + \beta$ with $\beta = 0$), with N between 8 and 64. In [7] the β_T model is studied again, with N up to 1024; in the weak stochasticity regime the same power law ε^2 as in the pure β model is suggested.

A very remarkable sequence of papers, including in particular [28] and [29] (see also sections 2.4.7 and 2.8.2 of the review paper [30]), develop a statistic–geometric theory, based on a subtle analysis of the fluctuations of the curvature in the phase space in a natural metrics, which leads to an analytic estimate of χ as a function of ε, in the thermodynamic limit. The theoretical curve fits impressively well the numerical data of the pure β model, both in the small ε and in the large ε regimes. Asymptotically for small ε the power law

$$\chi = C \, \varepsilon^2$$

(with computed C) is found.

Further comments on the theory, including a natural extension of it to all models in the hierarchy around the linear model (but not around Toda, where the theory apparently does not apply) are demanded to Section 3.

1.3 A sketch of results

Our results can be summarized by saying that:

– For all models, at large N and small ε, the Lyapunov exponent obeys a power law $\chi = C \, \varepsilon^a$, C and a depending on N; for large N, the exponent a always converges (although possibly very slowly) to a limit value, while the constant C, for the most important models namely $\alpha + \beta$ and pure β, possibly mantains a residual logarithmic dependence on N.

- For models of the linear hierarchy, the numerically found values of a (see the text for details) are close to "simple" values, namely to $a = 2$, in agreement with the above quoted literature, for the pure β model, to $a = 3$ and 4 respectively for the $\gamma + \delta$ and the pure δ model, and to $a = 1$ for those variable α models which, see Section 2.4, are close to the linear model.

- Models of the Toda hierarchy exhibit instead less understandable exponents, actually $a \simeq 1.5$, 1.9 and 2.2 respectively for the $\alpha + \beta$, the β_T and the γ_T model; $a \simeq 1.2$ for some other variable α models, which are close to Toda.

- The theory developed in [28] [29] in connection with the pure β model successfully extends to all models of the linear hierarchy, producing, for large N and small ε, the above mentioned values of a. The theory instead, as far as we understand it, does not extend to the models of the Toda hierarchy and to Toda itself (for which of course $\chi = 0$). The obstruction seems precisely to be the difficulty to detect the integrability of Toda, and consequently the peculiarities of the nearby models.

2 Numerical results

2.1 Computing χ

Trajectories have been computed by a symplectic integration algorithm, namely a fourth-order leap-frog based on [31]. For general questions concerning symplectic integration algorithms see [32], [33], [34]; for their use in connection with FPU models see [35], [36]. The Lyapunov exponent χ has been computed by the traditional algorithm [37] [38]. For any $x = (p, q)$ and any tangent vector $\xi = (\delta p, \delta q)$, let Φ^t and $D\Phi^t_x$ denote the flow and, respectively, its tangent application in x, and let

$$\lambda(x, t) = \frac{1}{t} \log \frac{\|D\Phi^t_x \xi\|}{\|\xi\|} \tag{7}$$

be the usual finite-time Lyapunov exponent. The limit for $t \to \infty$ of $\lambda(x, t)$ gives, for (almost) any choice of ξ, the maximal Lyapunov exponent in x. Practically, the dependence on ξ at the r.h.s. of (7) is soon lost, much before the limit, and for this reason ξ does not appear among the arguments of λ. The choice of the norm in the tangent space is similarly irrelevant (we used the Euclidian norm in the $\delta p, \delta q$ coordinates).

For any model, any N and any ε, we found convenient to consider 24 different initial data x, extracted randomly on the chosen constant energy surface with a Gaussian measure on normal modes (a measure close to the microcanonical one). As an example, figure 2, upper panels, shows $\lambda(x, t)$ as a function of t, as usual in log-log scale, for the $\alpha + \beta$ model with $N = 1024$, $\beta = 2$, at $\varepsilon = 10^{-3}$ (left) and $\varepsilon = 10^{-4}$ (right); the blue curves are the 24 individual curves $\lambda(x, t)$, the red curve is their average $\overline{\lambda}(t)$ (the average on the 24 initial data). Quite clearly, the finite time behavior depends on x (due to the different paths in the phase space), but the limit is the same; correspondingly, the precise choice of the 24 random initial data is quite irrelevant, and averaging is only helpful to produce a cleaner curve. (For a result on Lyapunov exponents, in which the average on the initial datum at fixed time plays a relevant role, see [39]). Practically, different samples of the 24 initial data produce hardly distinguishable averages $\overline{\lambda}(t)$ much before the limit.
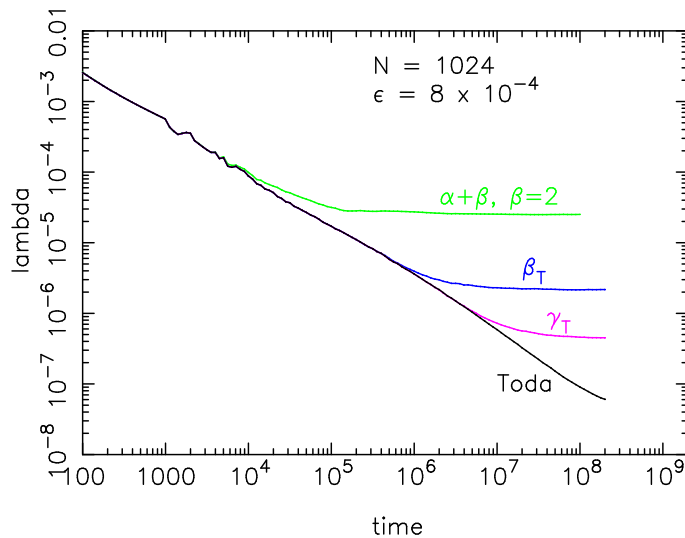
Figure 3: $\overline{\lambda}(t)$ *as function of t for the models of the Toda hierarchy. Models, top to bottom: $\alpha + \beta$ with $\beta = 2$, β_T, γ_T, Toda. Same $N = 1024$, $\varepsilon = 8 \times 10^{-4}$.*

For all models and any N and ε, the behavior of $\lambda(x,t)$ and $\overline{\lambda}(t)$ is as in figure 2. The *plateau* is obviously absent in Toda, for which (if the numerical integration is accurate, see Section 4.A) the maximal Lyapunov exponent vanishes and correspondingly both $\lambda(t,x)$ and the average $\overline{\lambda}(t)$ decrease indefinitely, see the lower panel of figure 2.

We shall denote by $\chi(N,\varepsilon)$ the limit value of $\overline{\lambda}(t)$. This is the quantity we are going to investigate.

2.2 χ for the Toda hierarchy

A. Generalities. In this subsection we address to models close to Toda, namely $\alpha + \beta$, β_T and γ_T. A preliminary indication that the proximity to Toda is relevant is provided by fig. 3, where $\overline{\lambda}(t)$ is plotted vs. t for these models as well as for Toda (same initial data). As is not surprising, the closer a model is to Toda, the longer its $\overline{\lambda}(t)$ follows $\overline{\lambda}(t)$ of Toda.

Remark. As a general well known fact, the finite time Lyapunov exponent $\lambda(x,t)$, in any nearly-integrable system, follows closely the corresponding curve of the reference integrable model, until it detaches and reaches a *plateau* (sometimes from above as in figure 3, sometimes from below). Using the average $\overline{\lambda}(t)$ in place of $\lambda(x,t)$ cleans the curves and makes the phenomenon more evident (look how the curves in figure 3 superimpose to the Toda curve, even in the details). For FPU, this phenomenon is commented for example in [27] and, with main attention to the case of small N, in [40]. We propose here a quite elementary interpretation (making reference, in the notation, to our $\overline{\lambda}(t)$). In an integrable system, the norm of the tangent vector, supposing it be one at $t = 0$, grows roughly linearly in time, say $n(t) \simeq 1 + ht$, h being some norm of the Hessian of the Hamiltonian in action-angle variables. For a nearly-integrable system it is, again roughly, $n(t) \simeq 1 + ht + c(e^{\chi t} - 1)$, χ small, and correspondingly

$$\overline{\lambda}(t) \simeq \frac{1}{t} \log[1 + ht + c(e^{\chi t} - 1)] . \tag{8}$$

Quite clearly, for $t < \chi^{-1}$ the behavior is almost identical to the corresponding integrable system,
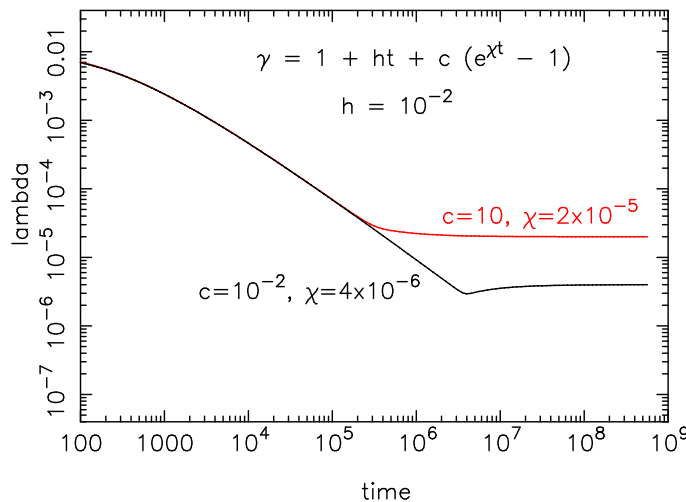
Figure 4: *A plot of (8), for $h = 10^{-2}$ and $c = 10$, $\chi = 2 \times 10^{-5}$ (upper curve), $c = 10^{-2}$, $\chi = 4 \times 10^{-6}$ (lower curve).*

and $\overline{\lambda}(t)$ decreases in time as $\log(1 + ht)/t$. At $t \simeq \chi^{-1}$ there is a crossover to a *plateau* and asymptotically

$$\overline{\lambda}(t) \simeq \chi + t^{-1} \log c \, ,$$

the *plateau* being reached from above for $c > 1$, from below for $c < 1$. $\overline{\lambda}$ as in (8) is plotted vs. t in figure 4; the upper curve has $c > 1$, the lower one has $c < 1$. The profiles are very typical of the computation of χ, in nearly-integrable models. Let us remark that our interpretation does not need assuming that trajectories move, after a certain "trapping time", from a regular to a chaotic region. We do not claim our view based on (8) is original, but we are not aware of a paper where it is suggested (for a related discussion, see however [41]). Let us further comment that, in this perspective, the detachment time is not an independent information on the dynamics, such a time essentially coinciding with χ^{-1}.

B. The $\alpha + \beta$ model. Figure 5, upper right panel, shows $\chi(N, \varepsilon)$ as function of ε for the $\alpha + \beta$ model with $\beta = 2$, $N = 1024$. The data for $\varepsilon < 2 \times 10^{-2}$ are quite well fitted by a line,[1] suggesting, as in [25] [26] [7] [28], a power law

$$\chi \sim \varepsilon^a \, . \tag{9}$$

The value of the exponent (least-squares best fit) is $a = 1.57$, but the value depends on N and, as shown in the upper right panel of the figure, for large N it seems to approach $a = 3/2$. The asymptotics looks unespectedly slow. The lower panel provides an overview of $\chi(N, \varepsilon)$ for several values of N between 64 and 16 384; values are less accurate for smaller N, as is shown by the larger now well visible error bars.

In fact, some criticism is here mandatory, for the very existence of a limit $N \to \infty$ for $\chi(N, \varepsilon)$ at fixed ε is questionable. Figure 6, left panel, shows $\chi(N, \varepsilon)$ as a function of N, log-log scale, for ε between 10^{-3} and 10^{-4}. The curves, at least at the values of N we reached, seem to mantain a

---

[1]In principle, the data in the figure include an error bar, deduced in a standard way from the dispersion of the 24 individual data: let $\chi_1, \ldots, \chi_n$, $n = 24$, be the values of $\lambda(x, t)$ for the n considered trajectories at the largest computed t; then the error bar corresponds to $\pm 3\sigma/\sqrt{n-1}$, σ being the standard deviation of the χ_i's. Bars are practically invisible because of the same size of the symbols or smaller.
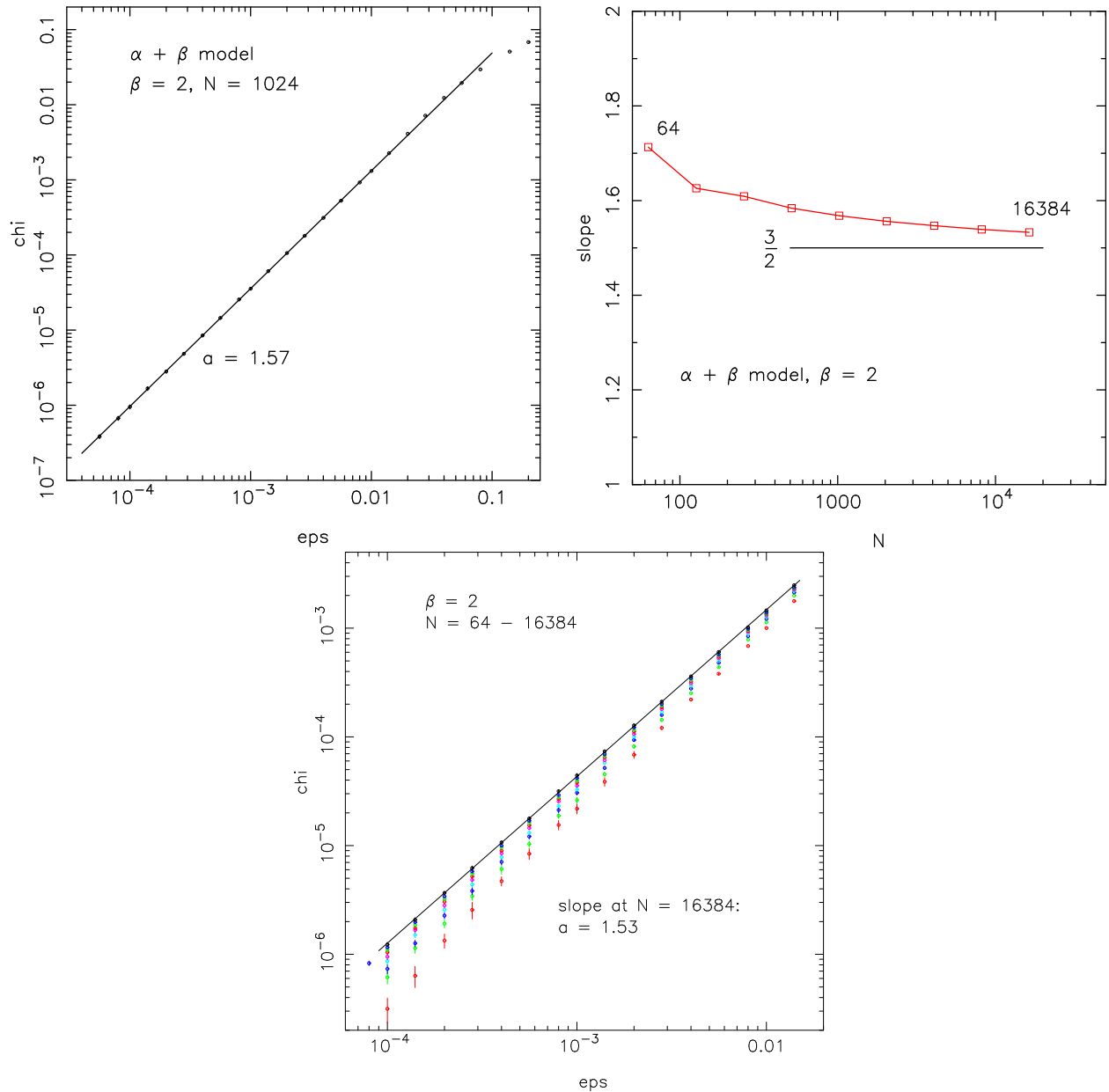
Figure 5: *Up left: $\chi(N, \varepsilon)$ as function of ε for the $\alpha + \beta$ model with $\beta = 2$ and $N = 1024$. Up right: the exponent a as function of N, at fixed $\beta = 2$ (least squares for $\varepsilon < 2 \times 10^{-2}$). Down: $\chi(N, \varepsilon)$ as function of ε, for $N = 64, 128, \ldots, 16\,384$. The line interpolates data for $N = 16\,384$.*
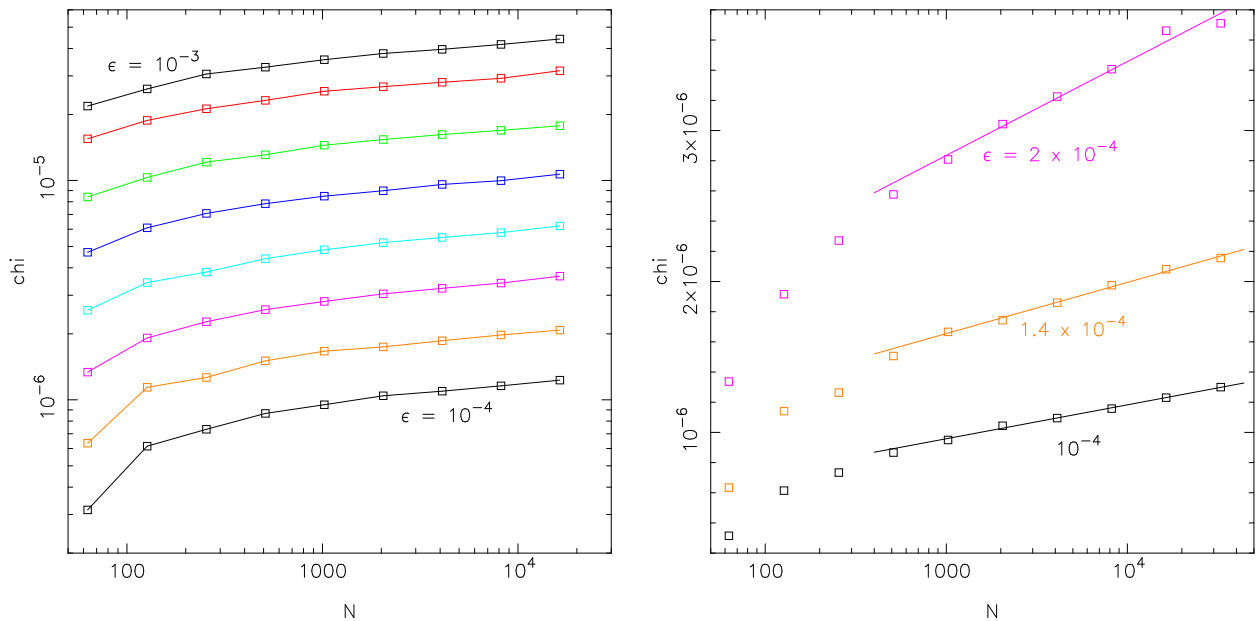
Figure 6: $\chi(N, \varepsilon)$ as function of N, for some values of ε, in the $\alpha + \beta$ model with $\beta = 2$. Left, log–log scale; right, selected values in semilog scale.

residual dependence on N (in spite of the fact that the slopes $a(N)$, as shown in fig. 5, apparently attain a limit). The right panel of figure 6 reports a few of such curves in semilog scale, in the attempt to put in evidence a possible logarithmic dependence on N. The interpolation with lines, for $N \geq 10^3$, looks acceptable. A power law N^b with small b, namely b between 0.07 and 0.09 depending on the curve, fits the data equally well, a larger range of N being necessary to decide. Let us stress the miss of the thermodynamic limit by a residual logarithmic dependence on N would be welcome, and we conjecture this is the case. We are not aware of a similar situation, at least not in connection with FPU. Great computational effort has been necessary to produce data as accurate as in figs. 5 and 6.

C. Other models of the Toda hierarchy. Computations have been repeated for the two mentioned models of the Toda hierarchy, namely β_T and γ_T, with N up to[2] 4 096. Qualitatively the behavior is similar to the $\alpha + \beta$ model, $\overline{\lambda}$ obeying power laws like (9). Figure 7 summarizes the results, showing together the different power laws we found for the different models. In detail:

○ The β_T model: we found $a \simeq 1.90$, a value close to $a = 2$ approximately found, and conjectured as true, in [7]. We obviously agree that $a = 2$ is an attracting conjecture, but such a value does not seem compatible with the numerical data. This is shown in figure 8, where $\chi(N, \varepsilon)$ is reported as a function of ε for $N = 256, 512, \ldots, 4\,096$ (only the data for small ε between 10^{-4} and 10^{-3} are reported). For each N, data look definitely well interpolated by lines, with slopes oscillating (without order) between 1.88 and 1.90. By comparison, a dashed line with slope 2, through the last point at $N = 4\,096$, is also drawn. The slopes look well distinguishable.

○ The γ_T model: we found $a \simeq 2.2$. Here too, the closest "attracting" value is $a = 2$, but

---

[2]Obtaining accurate results for such models is rather painful: the convergence of $\overline{\lambda}(t)$ to its limit is slow and requires very long computations. For this reason, N was limited to 4 096.
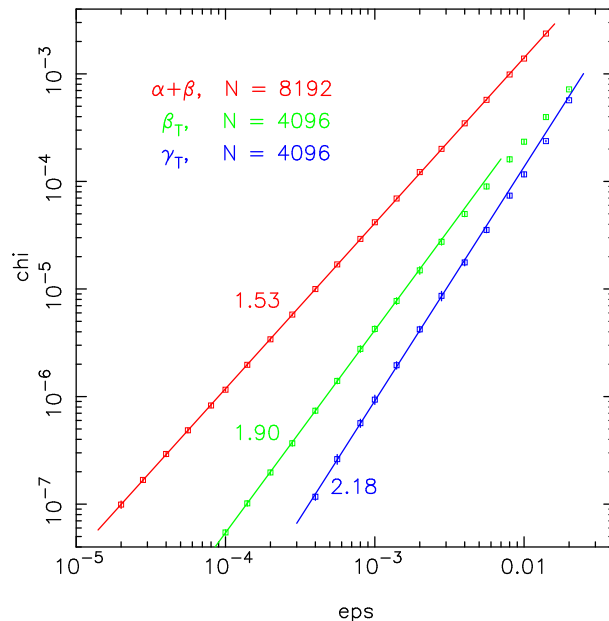
Figure 7: *The different power laws for models of the Toda hierarchy ($N = 8\,192$ for $\alpha + \beta$, $n = 4\,096$ for β_T and γ_T).*

> such a value looks not compatible with the numerical data. (In any case, it would be hard to conjecture $a = 2$ for both the β_T and the γ_T models.)

The overall situation looks rather puzzling. We shall further discuss this point among the Concluding remarks.

2.3 χ for the hierarchy around the linear model

A. The pure β model We studied the power law (9) for the pure β model, although established in the literature, so as to increase the accuracy of the existing results. As already remarked in the Introduction, both the available numerical data [25] [26] and the theory [28] [29] suggest $a = 2$. We confirm such a result, asymptotically for large N, with however quite slow asymptotics in N, as for the $\alpha + \beta$ model.

Figure 9 shows, in the left panel, $\chi(N, \varepsilon)$ as a function of ε, for $N = 128$, 256, ..., 8192. The line interpolates data for $N = 8\,192$. Its slope is $a = 2.05$, but as shown in the right panel of the figure, the slope depends on N, and for $N \to \infty$ it likely converges to 2. Nevertheless, a residual dependence of χ on N, approximately logarithmic as for the $\alpha + \beta$ model, apparently remains: see figure 10, similar to figure 6.

B. The next models.

As remarked in the Introduction, we investigated the two next models of the hierarchy around the linear model. The main reason, besides the pleasure of the symmetry of figure 1, is to make a test of the theory developed in [28], [29].

Concerning the $\gamma + \delta$ model, the choice of the constants is $\gamma = 1$ and $\delta = 0.8$. The result is a nice power law with $a \simeq 3.0$, essentially independent of N at least between $N = 1024$ and
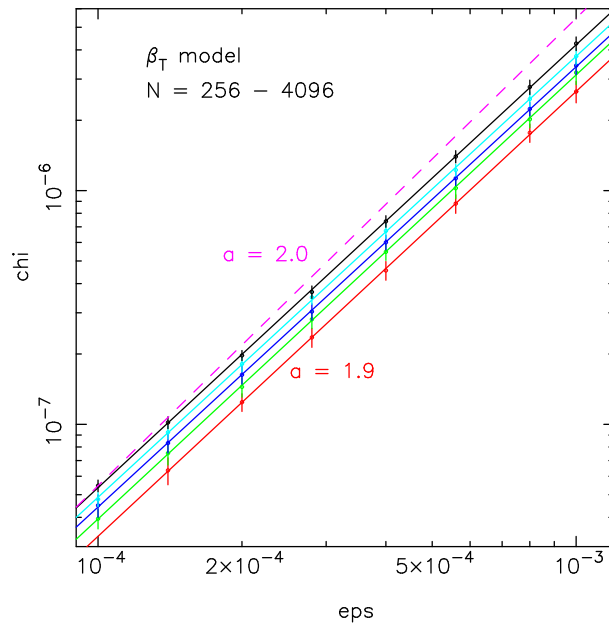
Figure 8: *Symbols and solid interpolating lines:* $\chi(N, \varepsilon)$ *as function of* ε, *for (bottom to top)* $N = 256, 512, \ldots, 4\,096$. *Dashed line: a line with slope* $a = 2$, *through the last point at* $N = 4\,096$.
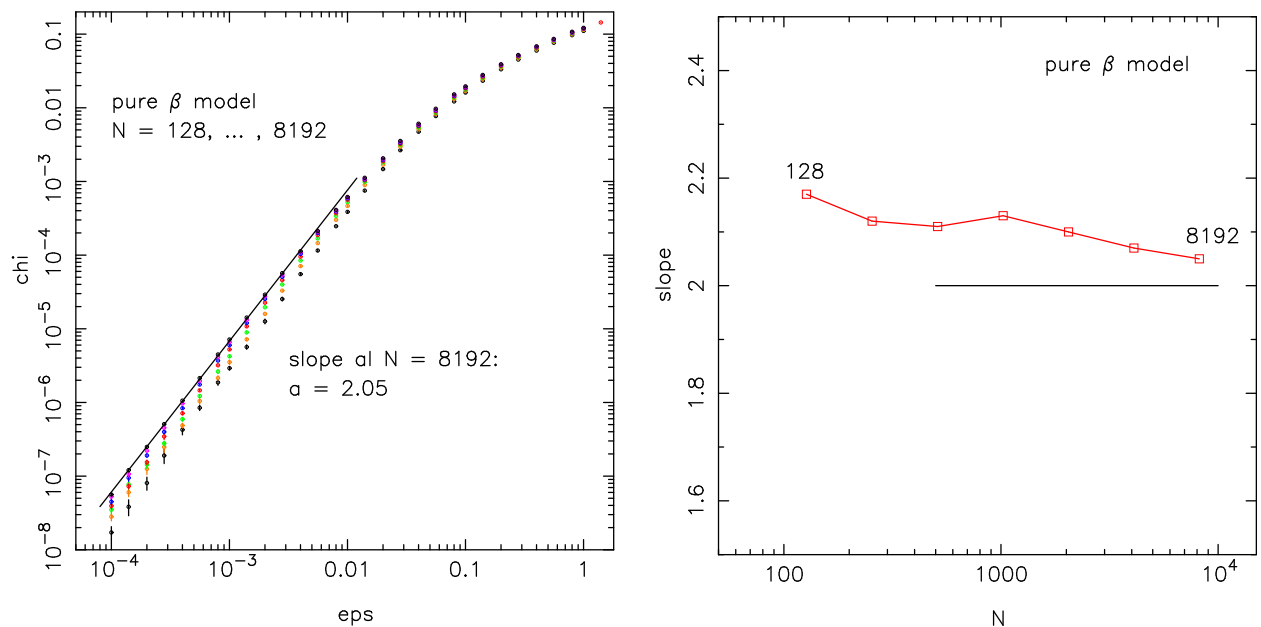

Figure 9: *Left:* $\chi(N, \varepsilon)$ *as function of* ε, *for the pure* β *model, for* $N = 128, 256, \ldots, 8\,192$; *the line with slope* $a = 2.05$ *interpolates data for* $N = 8\,192$. *Right: the computed slope as function of* N, *suggesting a possible convergence to* $a = 2$.
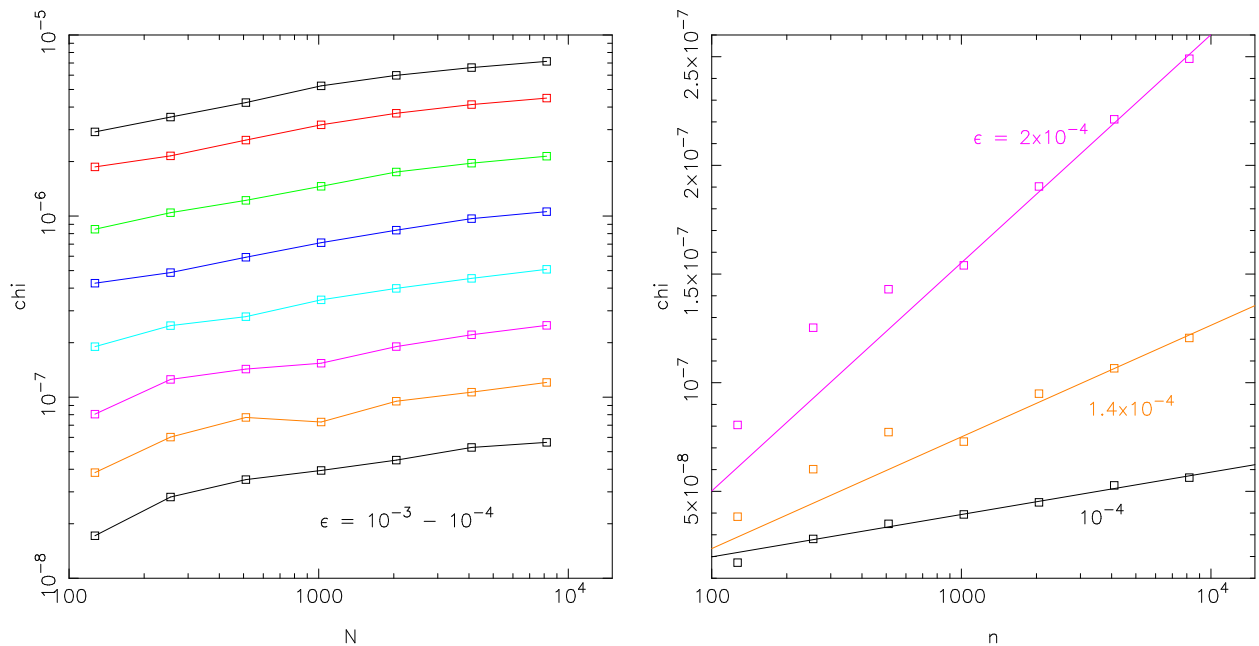
Figure 10: $\chi(N, \varepsilon)$ as function of N, for some values of ε, in the pure β model. Left, log–log scale; right, selected values in semilog scale.

$N = 8\,192$; see figure 11, upper curve. Concerning instead the pure δ model, the natural choice is $\delta = 1$; for such a model we found a close to 4, precisely $a \simeq 4.1$, see the lower curve of the same figure (numerics was a little painful here). The very simple rule for the exponent, also including the pure β model, seems to be

$$\chi \sim \varepsilon^s \qquad \text{for} \qquad V(r) = \frac{1}{2}r^2 + \mathcal{O}(r^{2+s}) \,, \tag{10}$$

for $s \geq 2$. As we shall comment in Section 3, such a rule is in agreement with the theoretical prediction.

2.4 χ for the variable α model

We studied several variable α models, among them:

(a) $\alpha_i = \pm 1$ (b) $\alpha_i = 1/2 \pm 1$
(c) $\alpha_i = 1 \pm 1/2$ (d) $\alpha_i = 1 \pm 1/3$,

randomly with equal probability. (We also considered models with the α_i modulated by a few Fourier components; results are similar, but a little less clear.) Different models do not behave identically: χ always follows asymptotically a power law $\chi \sim \varepsilon^a$, but the exponent a does depend on the model. More precisely:

– For models (a) and (b) of the above list, we found $a = 1$. Such a value fits the rule (10) of the linear hierarchy, actually extending it to $s = 1$, and suggests models (a) and (b) are part of it, Toda being apparently too far and not influencing the low ε dynamics. See the two upper lines of figure 12 (hardly distinguishable, for the data almost superimpose).
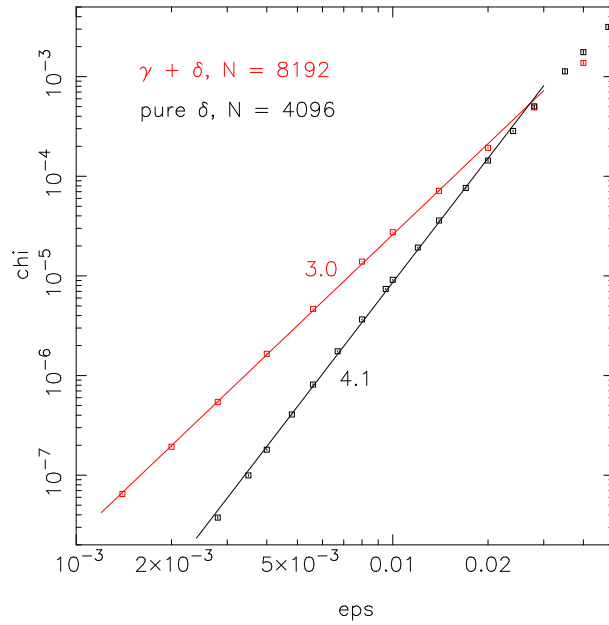
Figure 11: $\chi(N, \varepsilon)$ as function of ε for the $\gamma + \delta$ model at $N = 1024$ (upper curve), and the pure δ model at $N = 8192$ (lower curve); the computed slopes are, respectively, $a = 3.0$ and $a = 4.1$.
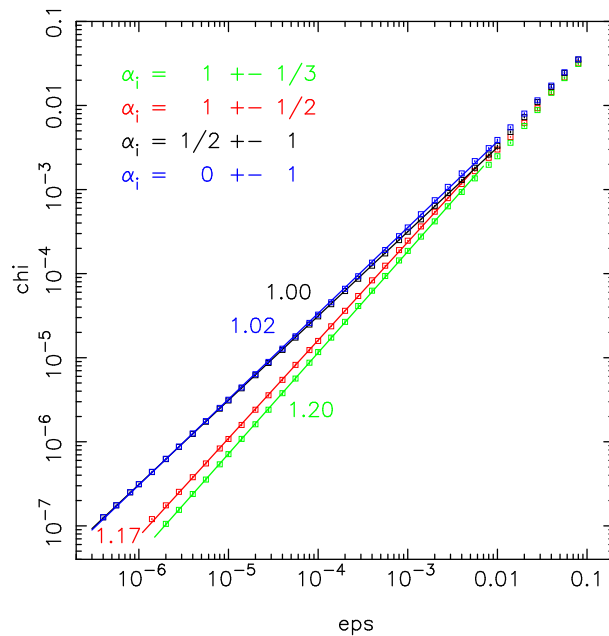

Figure 12: $\chi(N, \varepsilon)$ as function of ε for the variable α models (a)-(d), $N = 8192$. Lower curves: models (a) and (b); upper curves (almost superimposed): models (c) and (d). Computed slopes, in the order, $a = 1.00$, 1.02, 1.17, 1.19.

– Models (c) and (d) instead have a larger exponent $a \simeq 1.2$: a somehow misterious value, as all values of a for the Toda hierarchy are; see the two lower lines of figure 12. The difference with respect to (a) and (b) suggests the models feel the presence of Toda, and are part of its hierarchy, although, in lack of any interpretation of the exponents of the Toda hierarchy, it is hard to draw any conclusion.

Our feeling is that the variable α models feel the proximity of the linear model, when the variance of α_i, in some sense to be better understood, dominates on the average; by the way, in such a case the average seems to be irrelevant, as is shown by the exact superposition of data for models (a) and (b). If instead the average gets important, the presence of Toda gets relevant. In lack of any theoretical basis, we did not further investigate this point.

3   Theoretical items

As remarked in the Introduction, there exists an approximate theory, based on statistical and geometric considerations, adapted to weakly nonlinear systems with many degrees of freedom like FPU, which leads to an analytic estimate of the maximal Lyapunov exponent [28][29][30]. The theory was tested on the pure β FPU model (as well as in the so-called φ^4 model and in a chain of weakly coupled rotators) and the agreement with numerical data turned out to be excellent, both for small and for large specific energy.

The underlying idea is that, for systems like FPU, the exponential separation of nearby trajectories is not produced (as in Anosov systems) by a negative curvature of the manifold where the motion occurs, i.e. of the constant energy surface, rather it is a phenomenon of parametric instability due to the fluctuations of the curvature along trajectories, in a convenient natural metrics. Schematically the theory, as developed in the above quoted references, proceeds as follows:

i. Within certain assumptions (the most important one is that, in the chosen coordinates, the oscillations or rotations of the tangent vector ξ do not play an important role), the problem is reported to a Hill equation for a single oscillator, namely

$$\ddot{\psi} + \Omega(t)\psi = 0 \ , \tag{11}$$

where however $\Omega(t)$ is not an assigned function of time but a stochastic process; $\psi(t)$ represents any of the components of ξ. The process is assumed to be Gaussian and δ–correlated on a convenient time scale τ. Its average Ω_0 and variance σ (after a nontrivial geometrical analysis) are found to be

$$\Omega_0 = \frac{1}{N}\langle\Delta\mathcal{V}\rangle \ , \qquad \sigma^2 = \frac{1}{N}\left[\langle(\Delta\mathcal{V})^2\rangle - \langle\Delta V\rangle^2\right] \ ,$$

where $\Delta\mathcal{V}$ is the Laplacian of the potential energy

$$\mathcal{V}(q) = \sum_{i=1}^{N} V(r_i) \ , \qquad r_i = q_i - q_{i-1} \ ,$$

and $\langle\,.\,\rangle$ denotes microcanonical averaging. Concerning the correlation time τ, dimensional arguments lead to two candidates [28]:

$$\tau_1 = \sqrt{\frac{2}{\Omega_0}} \ , \qquad \tau_2 = \sqrt{\frac{\Omega_0}{2\sigma^2}} \ .$$

The choice of the authors, leading to good results, is

$$\tau^{-1} = (\tau_1^{-1} + \tau_2^{-1}) . \tag{12}$$

The idea underlying the composition rule (12) is that, if two very different correlation times enter the dynamics, χ is sensitive to the shorter one. In the asymptotics $\varepsilon \to 0$ one has $\sigma \ll \Omega_0$, and thus $\tau = \tau_1$. The numerical factors looks a little arbitrary; on the other hand, any dimensional estimate includes by itself an arbitrary factor.

ii. Once the process is defined, a result by Van Kampen [44] applies, leading to

$$\chi(\Omega_0, \sigma, \tau) = \frac{1}{2}\left(\Lambda - \frac{4\Omega_0}{3\Lambda}\right) ,$$

where

$$\Lambda = \left[2\tau\sigma^2 + \sqrt{(4\Omega_0/3)^3 + (2\tau\sigma^2)^2}\right]^{1/3} .$$

Asymptotically for small ε one has $\sigma \ll \Omega_0$ and correspondingly

$$\Omega_0 \to 2 , \qquad \tau \to 1 , \qquad \chi \simeq \frac{1}{8}\sigma^2 . \tag{13}$$

So, everything is reported to estimating the variance of

$$\Delta\mathcal{V} = 2\sum_{i=1}^{N} V''(r_i) - V''(r_1) - V''(r_N) ;$$

the last two terms are due to the choice of fixed ends, and are negligible for large N. Quite clearly, for the potentials we are dealing with,

$$\Delta\mathcal{V} = 2N + 4\sum_{i=1}^{N}\alpha_i r_i + 6\beta\sum_{i=1}^{N} r_i^2 + 8\gamma\sum_{i=1}^{N} r_i^3 + 10\delta\sum_{i=1}^{N} r_i^4 , \tag{14}$$

where the possibility of site depending α_i has been taken into account.

Let us forget for a moment the constants and look only at the dependence of σ, and thus of χ, on ε. The computation gets straightforward: since at small ε the harmonic energy dominates, it is $r_i \sim \sqrt{\varepsilon}$, and consequently

$$\chi \sim \sigma^2 \sim \varepsilon^s \qquad \text{for} \qquad V(r) = \tfrac{1}{2}r^2 + \mathcal{O}(r^{2+s}) , \qquad s \geq 1 .$$

This is precisely rule (10) governing the linear hierarchy. *For $s = 1$, however, it is necessary that the constants α_i do depend on the site*: if α is constant, as in the $\alpha + \beta$ model, then, because of the fixed ends, the linear term in (14) exactly vanishes (the same would happen with periodic boundary conditions, actually whenever $\sum_i r_i$ is not allowed to fluctuate). For the $\alpha + \beta$ model, the theory predicts $\chi \sim \varepsilon^2$, precisely as for the pure β model.

In fact, the theory predicts the exponent $a = 2$ not only for the $\alpha + \beta$ model, but for all models of the Toda hierarchy, *including Toda itself*. This is not surprising: indeed the theory captures the nonlinearity, rather than the lack of integrability, and cannot be sensitive to the absolutely peculiar properties of Toda and of the models close to it.

Computing the exact asymptotics of σ in ε, including the constants, can be done for all models of the linear hierarchy. The computation we did is sketched in the Appendix; it takes into account the possibility that not only α, but also β, γ and δ depend on the site. Denoting by $\overline{(\cdot)}$ and by $\sigma^2_{(\cdot)}$, respectively, the arithmetic mean and the variance of the variables $(\cdot)$, the asymptotic result for χ turns out to be

$$
\chi(\varepsilon) \simeq
\begin{cases}
2\sigma^2_\alpha\, \varepsilon & \text{variable } \alpha \\[2mm]
\frac{9}{2}\left(\overline{\beta^2} + \sigma^2_\beta\right)\varepsilon^2 & \text{pure } \beta \\[2mm]
48\left(\overline{\gamma^2} + \frac{3}{2}\sigma^2_\gamma\right)\varepsilon^3 & \gamma + \delta \\[2mm]
750\left(\overline{\delta^2} + \frac{3}{5}\sigma^2_\delta\right)\varepsilon^4 & \text{pure } \delta
\end{cases}
. \tag{15}
$$

For the pure β model, this asymptotic result reproduces well the asymptotic line reported in [28][29][30]. Such a line fits well our numerical data for $N = 1024$ or 2048, while for larger N the numerical data lie slightly above it; see figure 13, upper-left panel. For the other models of the linear hierarchy, the situation is as follows:

– The variable α models (a) and (b): the theoretical line lies somehow above the data, approximately by a factor 6.

– The $\gamma + \delta$ model: the theoretical line lies above the data, approximately by a factor 2.

– The pure δ model: the agreement looks quite good.

See the other panels of figure 13. We are not surprised that a factor of order one is needed, due to the necessarily rough definition of τ, and we consider the theory developed in the above quoted references to be remarkably correct and successful, for all models of the linear hierarchy.

4   Further remarks

A. Numerical errors?   Numerical integration of differential equations is never safe. The procedure we used, mentioned at the beginning of Section 2, is however among the most standard ones, both for trajectories and for the computation of Lyapunov exponents.

Concerning trajectories, let us say we successfully used fourth-order leap-frog in several papers, also investigating theoretically the behavior of this algorithm in connection with FPU-like models [36]. Among the crucial tests, the algorithm is observed to preserve well not only energy, but also the Toda constants of motion, when used to integrate Toda (it is obviously essential that the diffusion transversal to Toda tori is produced by the dynamics, not by the algorithm).

Concerning specifically the computation of Lyapunov exponents, the main risk, as far as we understand, is overestimation: consider the case in which a small or vanishing exponent is due to the fact that, in the metrics at hand, a relatively large expansion (easy example: local hyperbolicity, tangent vector pointing in an expanding direction) is followed, after some time T, by a compensating contraction (local hyperbolicity, evolved tangent vector pointing now in a contracting direction). Then, if T is large, there is a chance such a correlation is lost, compensation is partially missed and expansion prevails.
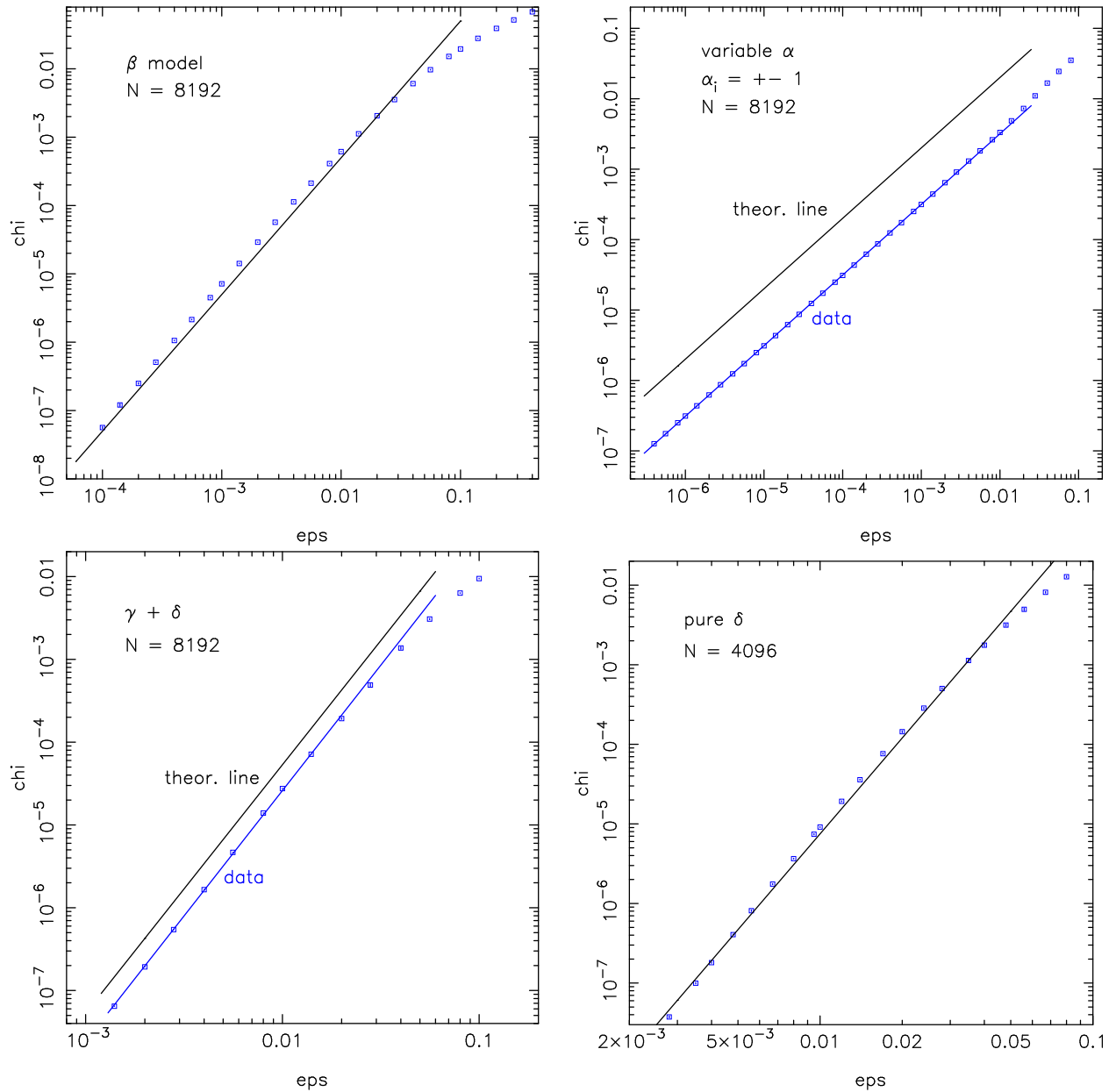
Figure 13: *A comparison between the theoretical lines and the numerical data. Upper left: the pure β model; upper right: the variable α model (a); lower left: the γ + δ model; lower right: the pure δ model. N it the largest computed, namely 8192 for the first three models, 4096 for the last one.*

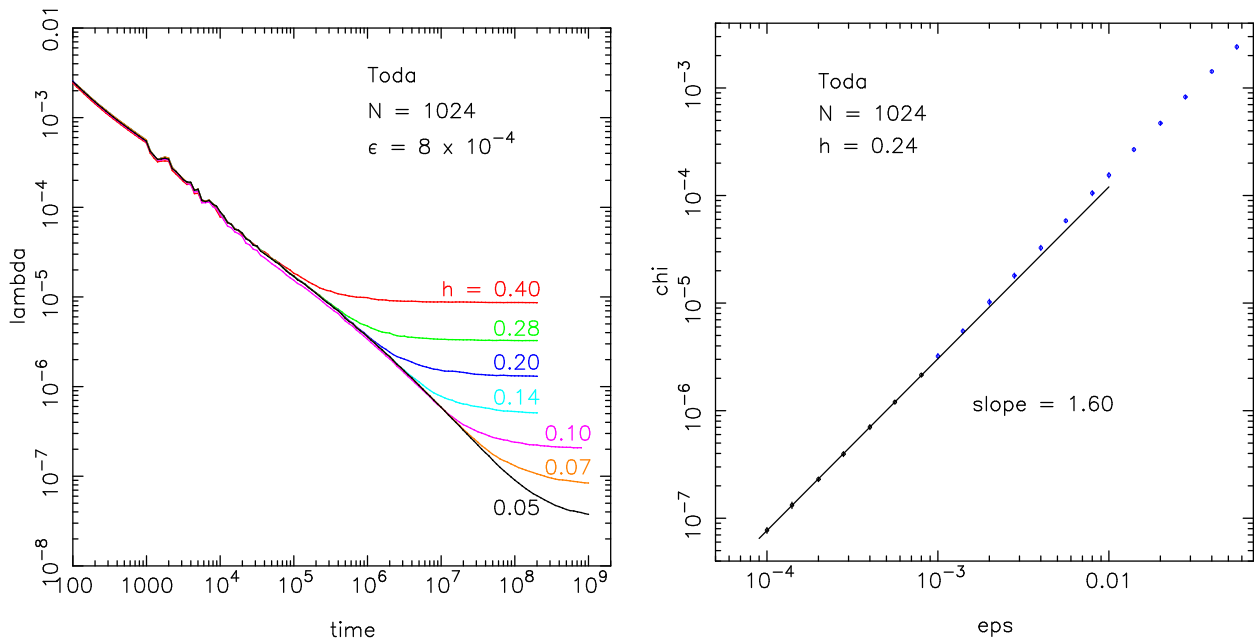Figure 14: *Left: $\overline{\lambda}(t)$ vs. t for the Toda model with $N = 1024$, at $\varepsilon = 8 \times 10^{-4}$, for several choices of the integration step h between 0.05 and 0.4. Right: the dependence of the limit value χ on ε, for fixed (large) $h = 0.24$.*

A good test is then computing χ for Toda, for which compensation is exact, and looking how close to zero is the computed value; intentionally, the integration step h can be taken large, so as to emphasize the effect on χ of a rough algorithm. Results are in figure 14. The left panel shows $\overline{\lambda}(t)$ vs. t for the Toda model with $N = 1024$, at $\varepsilon = 8 \times 10^{-4}$, h ranging from 0.05 to 0.40; for each h, the *plateau* reached by $\overline{\lambda}(t)$ provides an estimate of the possible error in χ in models close to Toda. The figure clearly resembles figure 3, as is not surpring, since the essential effect of the integration algorithm is that of introducing a perturbation [32] [33] [34]. In fact, the situation is different, since in figure 14 we are moving a parameter while in figure 3 the order of tangency changes discontinuously: nevertheless, the "perturbed" system might be considered, in our opinion, a further member of the Toda hierarchy. The computed value of χ (the error) also depends on ε, as is shown in the right panel of the figure, where χ is reported vs. ε for fixed $h = 0.24$ (a quite large value, so as to have large enough computable errors). As is remarkable, here too we find a well defined power law ε^a, with $a \simeq 1.60$. The value of a, almost identical to a of the $\alpha + \beta$ model at the same N, suggests that the effective perturbation on Toda introduced by the algorithm is quartic. We did not further investigate this point, although it could be done along the lines of [36].

Practically, in our numerical computations we always took care that the error due to the algorithm was much smaller than the measured value of χ. In most cases, $h = 0.1$ (the typical value we used with confidence, in the past, with fourth order leap-frog) was fairly enough; only for the γ_T and the pure δ models, at small ε, it was necessary to use $h = 0.05$.

B. Anomalous exponents for the Toda hierarchy? The power laws we produced look rather accurate. For the linear hierarchy the exponents look "simple"and close to the theoretical values, while for the Toda hierarchy they are more problematic. For the $\alpha + \beta$ model we observed a slow asymptotics in N leading, likely, to $a = 1.5$, while for the other models of the hierarchy simple

values were not observed and cannot be suggested. As far as we know, all power laws met in FPU, numerically or theoretically, in the weak stochasticity regime, have exponents either integer, or half-integer, or with denominator 4. The exponents we found for the β_T and the γ_T models, as well as for the variable α models (c) and (d), look then anomalous and are hard to understand. Some theory, similar and as powerful as the one reported in [28][29][30], but adapted to the neighborhood of the Toda model, seems to be necessary, but developing it looks a rather difficult task.

The Appendix: analytical details

In this Appendix we report some details concerning the analytical estimate of the Lyapunov exponents for the linear hierarchy. As stressed above, the theory is essentially that developed in [28, 29] for the pure β-model: we just extended it to the other models of the hierarchy, limiting ourselves to compute the asymptotic behavior of χ as $N \to \infty$ and $\varepsilon \ll 1$. The symbol '$\simeq$', in the sequel, means asymptotic equality in this limit. As pointed out in Section 3, according to the theory the maximal Lyapunov exponent is given by

$$\chi(\varepsilon) \simeq \frac{\sigma^2}{8} = \frac{\langle (\Delta \mathcal{V})^2 \rangle - \langle \Delta \mathcal{V} \rangle^2}{8N} \ , \tag{A.1}$$

where $\langle \cdot \rangle$ denotes the microcanonical average on the constant energy surface $H(p, q) = N\varepsilon$. The right hand side of (A.1) has to be computed to leading order in ε. We admit the possibility that the constants, at any order of nonlinearity, depend on the site, that is

$$\mathcal{V} = \sum_{i=1}^{N} V_i(r_i) = \sum_{i=1}^{N} \left(\frac{r_i^2}{2} + \alpha_i \frac{r_i^3}{3} + \beta_i \frac{r_i^4}{4} + \gamma_i \frac{r_i^5}{5} + \delta_i \frac{r_i^6}{6} \right) \ , \tag{A.2}$$

only assuming that at any site, the pair potential $V_i(r)$ displays a single minimum at $r = 0$ and diverges to $+\infty$ for $r \to \pm\infty$ ($\delta_i > 0$, or $\delta_i = \gamma_i = 0$ and $\beta_i > 0$). The corresponding expression of the Laplacian is

$$\Delta \mathcal{V} = 2N + 4 \sum_{i=1}^{N} \left(\alpha_i r_i + 6 \sum_{i=1}^{N} \beta_i r_i^2 + 8 \sum_{i=1}^{N} \gamma_i r_i^3 + 10 \sum_{i=1}^{N} \delta_i r_i^4 \right) \ . \tag{A.3}$$

As pointed out in [28, 29], the microcanonical quadratic fluctuation of $\Delta \mathcal{V}$, appearing in the numerator on the right hand side of (A.1), can be computed in terms of suitable *canonical* averages, according to the theory of Lebowitz, Percus and Verlet (LPV) [45]. For $\varepsilon \ll 1$ the LPV formula reads, to leading order in ε,

$$\frac{\langle (\Delta \mathcal{V})^2 \rangle - \langle \Delta \mathcal{V} \rangle^2}{N} \simeq \frac{\langle (\Delta \mathcal{V})^2 \rangle_c - \langle \Delta \mathcal{V} \rangle_c^2}{N} - \varepsilon^2 \left(\frac{\langle \Delta \mathcal{V} \rangle_c'}{N} \right)^2 \ , \tag{A.4}$$

where $\langle \cdot \rangle_c$ denotes the canonical average at temperature ε, $\langle \Delta \mathcal{V} \rangle_c' = d\langle \Delta \mathcal{V} \rangle_c/d\varepsilon$, and it is understood that all quantities on the right hand side are computed to leading order in ε. Using the simple scaling argument $r_i \sim \sqrt{\varepsilon}$, valid when $\varepsilon \ll 1$, from expression (A.3) one easily realizes that, according to the LPV formula (A.4), the microcanonical and the canonical quadratic fluctuation

of $\Delta \mathcal{V}$ differ, to leading order, only for leading nonlinearity of even order, that is for the β-models ($\alpha_i = 0$), and the δ-models ($\alpha_i = \beta_i = \gamma_i = 0$).

Up to this point we did not introduce any novelty, apart from extending formulas to all models of the linear hierarchy. We now proceed by computing the statistical averages for the different models; for the sake of consistency, we shall take into account the constraint $\sum_{i=1}^{N} r_i = 0$, which holds both for fixed ends and for periodic boundary conditions, whereas in [28, 29] the analytic computations are made in the simpler case of free ends. Such a constraint, as already mentioned in Section 3, plays a subtle role.

All canonical averages on the right hand side of (A.4) are clearly sums, with easily computed coefficients, of moments of the form

$$\langle r_j^m r_k^n \rangle_c = \frac{\int r_j^m r_k^n \; e^{-\frac{1}{\varepsilon}\mathcal{V}(r_1,\ldots,r_N)} \delta(\sum_i r_i) \; dr_1 \ldots dr_N}{\int e^{-\frac{1}{\varepsilon}\mathcal{V}(r_1,\ldots,r_N)} \delta(\sum_i r_i) \; dr_1 \ldots dr_N} \; ,$$

where j, k are any indices from 1 to N, while m, n are exponents between 0 and 4 and $\delta(x)$ denotes the Dirac function (expressing the constraint $\sum_{i=1}^{N} r_i = 0$), all integrals running from $-\infty$ to ∞. By representing the Dirac function in the integral form $\delta(x) = \frac{1}{2\pi} \int e^{iyx} dy$, using the expression (A.2) of the potential energy, one gets

$$\langle r_j^m r_k^n \rangle_c = (-\imath)^{m+n} \; \frac{\int \left[\prod_{\substack{i=1 \\ i \neq j,k}}^{N} f_i(y) \right] f_j^{(m)}(y) f_k^{(n)}(y) \; dy}{\int \prod_{i=1}^{N} f_i(y) \; dy} \; , \tag{A.5}$$

where $\imath$ denotes the imaginary unit, f_i is defined as

$$f_i(y) = \int e^{-\frac{1}{\varepsilon}V_i(r)+\imath yr} \; dr \; , \tag{A.6}$$

and $f_i^{(s)}$ denotes the s-derivative of f_i. Now, under the hypothesis made above that $V_i(r)$ has an unique minimum at $r = 0$, one easily shows, by the Laplace method (see for example [46]), that for small ε it is

$$f_i(y) \simeq \sqrt{2\pi\varepsilon} \; e^{-\varepsilon \frac{y^2}{2}} \; ,$$

independent of i (this is not surprising since in (A.6), for small ε, only the i-independent harmonic part of V_i contributes to the integral). In such a way, the average (A.5) assumes the simple asymptotic form

$$\langle r_j^m r_k^n \rangle_c = (-\imath)^{m+n} \; \frac{\int e^{-N\varepsilon \frac{y^2}{2}} P_{m+n}(y) \; dy}{\int e^{-N\varepsilon \frac{y^2}{2}} \; dy} \; , \tag{A.7}$$

where $P_{m+n}(y)$ is a suitable polynomial of degree $m + n$; moreover, it turns out that if $m + n$ is even (odd), then $P_{m+n}(y)$ contains only even (odd) terms, so in particular, for odd $m+n$, the right hand side of (A.7) vanishes.

In this way one reduces all the analytic work to computing trivial Gaussian moments. Taking a bit carefully into account the dependence on N and summing up all terms, yields the asymptotic estimates (15).

References

[1] E. Fermi, J. Pasta, and S. Ulam, *Studies of Non Linear Problems,* Los-Alamos Internal Report, Document LA-1940 (1955), first published in: *Enrico Fermi Collected Papers*, Vol. II, The University of Chicago Press, Chicago, and Accademia Nazionale dei Lincei, Roma, 1965, pp. 977-988. The paper was later reprinted in Lect. Appl. Math. **15**, 143–156 (1974) and inside ref. [11].

[2] M. Toda, *Vibration of a Chain with Nonlinear Interaction*, Journ. Phys. Soc. Japan **22**, 431–436 (1967); *Wave Propagation in Anharmonic Lattices*, Journ. Phys. Soc. Japan **23**, 501–506 (1967); *Mechanics and Statistical Mechanics of Nonlinear Chains*, Journ. Phys. Soc. Japan Suppl. **26**, 109–111 (1969); *Waves in nonlinear lattice*, Progr. Teor. Phys. Suppl. **45**, 174–200 (1970).

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

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

[5] B.A. Dubrovin, *On universality of critical behavior in Hamiltonian PDEs*, in: Geometry, topology, and mathematical physics: S.P. Novikov's seminar, 2006-2007, V.M. Buchstaber and I.M. Krichever editors. AMS Translations series 2, vol. 24 - Providence, R.I. 2008. (See Sect. 4.3.)

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

[7] M. Cerruti-Sola, M. Pettini and E. G. D. Cohen, *Phase space geometry and stochasticity thresholds in Hamiltonian dynamics*, Phys. Rev. E **62**, 6078–6081 (2000).

[8] 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).

[9] G. Benettin, H. Christodoulidi and A. Ponno, *The Fermi–Pasta–Ulam problem and its underlying integrable dynamics*, Journ. Stat. Phys. **152**, 195–212 (2013).

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

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

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

[13] F.M. Izrailev and B.V. Chirikov, *Statistical properties of a nonlinear string*, Sov. Phys. Dokl. **11**, 30–34 (1966).

[14] F.M. Izrailev, A.I. Khisamutdinov and B.V. Chirikov, *Numerical experiments with a chain of coupled anharmonic oscillators*, Report 252, Institute of Nuclear Physics, Novosibirsk, URSS (1968). English translation: LA 4440 TR, Los Alamos (1970).

[15] P. Bocchieri, A. Scotti, B. Bearzi and A. Loinger, *Anharmonic chain with Lennard–Jones interaction*, Phys. Rev. A **2**, 2013–2019 (1970).

[16] L. Galgani and A. Scotti, *Recent progress in classical nonlinear dynamics*, Rivista Nuovo Cim. **2**, 189–209 (1972).

[17] 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).

[18] 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).

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

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

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

[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] M. Casartelli, E. Diana, L. Galgani and A. Scotti, *Numerical computations on a stochastic parameter related to the Kolmogorov entropy*, Phys. Rev. A **13**, 1921–1925 (1976).

[24] G. Benettin, G. Lo Vecchio and A. Tenenbaum, *Stochastic transition in two-dimensional Lennard–Jones system*, Phys. Rev. A22, 1709–1719 (1980).

[25] M. Pettini and M. Landolfi, *Relaxation properties and ergodicity breaking in nonlinear Hamiltonian dynamics*, Phys. Rev. A **41**, 768–783 (1990).

[26] M. Pettini and M. Cerruti-Sola, *Strong stochasticity threshold in nonlinear large Hamiltonian systems: Effect on mixing times*, Phys. Rev. A **44**, 975–987 (1991).

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

[28] L. Casetti, R. Livi and M. Pettini, *Gaussian model for chaotic instability of Hamiltonian flows*, phys. Rev. Lett. **74**, 375–378 (1995).

[29] L. Casetti, C. Clementi and M. Pettini, *Riemannian theory of Hamiltonian chaos and Lyapunov exponents*, Phys. Rev. E **54**, 5969-5984 (1996).

[30] A.J. Lichtenberg, R. Livi, M. Pettini and S. Ruffo, *Dynamics of oscillator chains*, in [11].

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

[32] G. Benettin and A. Giorgilli, *On the Hamiltonian interpolation of near to the identity symplectic mappings, with application to symplectic integration algorithms*, J. Stat. Phys. **74**, 1117–1144 (1994).

[33] E. Hairer, *Backward analysis of numerical integrators and symplectic methods,* Annals of Numerical Mathematics 1, 107–132 (1994).

[34] E. Hairer, C. Lubich and G. Wanner, *Geometric numerical integration,* Springer–Verlag (Berlin–Heidelberg, 2002; second edition 2006).

[35] S. Paleari and T. Penati, *Numerical methods and results in the FPU problem,* in: G. Gallavotti Ed., *The Fermi–Pasta–Ulam problem,* Lect. Notes Phys. **728**, 239-282 (2008).

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

[37] G. Benettin, L. Galgani and J.-M. Strelcyn, *Kolmogorov Entropy and Numerical Experiments*, Phys. Rev. **A14**, 2338 (1976).

[38] G. Benettin, L. Galgani, A. Giorgilli and J.-M. Strelcyn, *Lyapunov Characteristic Exponents for Smooth Dynamical Systems and for Hamiltonian Systems; a Method for Computing all of Them. Part 1: Theory*, Meccanica **15**, 9 (1980); *Part 2: Numerical Applications*, Meccanica **15**, 21 (1980).

[39] Ya. G. Sinai, *A remark concerning the thermodynamical limit of the Lyapunov spectrum*, Int. Journ. Bifurcations Chaos **6**, 1137-1142 (1995).

[40] A. Giorgilli, S. Paleari and T. Penati, *Local chaotic behavior in the Fermi-Pasta-Ulam system*, DCDS-B **5**, 991-1004 (2005).

[41] G. Lukes-Gerakopoulos, N. Voglis and C. Efthymiopoulos, *The production of Tsallis entropy in the limit of weak chaos and a new indicator of chaoticity*, Physica A 387, 1907-1925 (2008).

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

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

[44] N.G. Van Kampen, *Stochastic differential equations,* Phys. Rep. **24**, 171–228 (1976).

[45] J.L. Lebowitz, J.K. Percus and L. Verlet, *Ensemble Dependence of Fluctuations with Application to Machine Computations*, Phys. Rev. **153**, 250–254 (1967).

[46] M.J. Ablowitz and A.S. Fokas, *Complex Variables*, 2nd edition, Cambridge University Press, Cambridge, 2003.