

## ARNOLD DIFFUSION IN LARGE SYSTEMS

B. V. Chirikov\*, V. V. Vechevslavov

*Budker Institute of Nuclear Physics  
630090, Novosibirsk, Russia*

Submitted 19 November 1996

A new regime of Arnold diffusion in which the diffusion rate has a power-law dependence on the perturbation strength is studied theoretically and in numerical experiments. The theory developed predicts this new regime to be universal in the perturbation intermediate asymptotics, the width of the latter increasing with the dimensionality of the perturbation frequency space, particularly, in large systems with many degrees of freedom. The results of numerical experiments agree satisfactorily with the theoretical estimates.

## 1. INTRODUCTION: UNIVERSAL NONLINEAR INSTABILITY

One of the most interesting phenomena in Hamiltonian dynamics is the so-called Arnold diffusion (AD), a distinctive universal instability of multidimensional nonlinear oscillations [1, 2]. This global instability was predicted by Arnold [3]; its chaotic nature was discovered in Refs. [4, 5, 1] and further studied in detail in Refs. [6–11, 14, 15, 17].

First, following Ref. [17], we briefly recall the diffusion mechanism, which is related to the interaction of nonlinear resonances. Consider a general Hamiltonian describing multidimensional oscillations:

$$H(I, \theta, t) = H_0(I) + \varepsilon \sum_{n,m} V_{nm}(I) \exp(in \cdot \theta + itm \cdot \Omega), \quad (1.1)$$

where  $I, \theta$  are  $N$ -dimensional vectors of the action-angle variables;  $\Omega$  is the  $M$ -dimensional vector of the driving frequencies;  $n, m$  are integer vectors of dimensions  $N$  and  $M$ , respectively, and  $\varepsilon$  stands for a small perturbation parameter. The dot in expressions like  $n \cdot \theta$  denotes the scalar product. Below we shall consider the simpler case of a completely integrable and nondegenerate unperturbed system whose Hamiltonian  $H_0(I)$  depends on the full set of  $N$  actions only.

Hamiltonian (nondissipative) dynamics is always determined by resonances (see, e.g., Refs. [1, 2]) corresponding to particular terms in the perturbation (1.1). The condition for a primary resonance with unperturbed frequencies (1.3) reads:

$$\omega_{nm} \equiv n \cdot \omega(I) + m \cdot \Omega \approx 0. \quad (1.2)$$

In the case of linear oscillations all the frequencies are fixed as parameters of the system which is either in or off resonance independent of initial conditions. However, for nonlinear oscillations

\*chirikov@inp.nsk.su

with the action-dependent frequencies

$$\omega(I) = \partial H_0(I) / \partial I, \quad (1.3)$$

condition (1.2) determines resonance surfaces (zones) in the phase space, that is, the system is always in resonance for some initial conditions. On the other hand, nonlinearity stabilizes the impact of a (sufficiently weak) perturbation, ensuring bounded oscillations even for resonant initial conditions. This is precisely due to non-isochronous oscillations (1.3). In one degree of freedom such a nonlinearity is necessary and sufficient to destroy the oscillation isochronism. The generalization of that for several degrees of freedom is the necessary condition for determinant to be nonzero everywhere,

$$\left| \frac{\partial^2 H_0}{\partial I^2} \right| \neq 0. \quad (1.4)$$

In this case the system is called nondegenerate. This allows, in particular, the transformation from action to frequency space. In the latter, the resonance structure is especially simple and transparent, as the resonant surfaces (1.2) become planes.

Another condition for the nonlinear stabilization is the requirement for the quadratic form associated with the matrix  $\partial^2 H_0 / \partial I^2$  to be sign-definite or, geometrically, for surfaces  $H_0(I) = \text{const}$  to be convex [10]. The latter condition is a weaker one as it may include higher polynomial forms. Both conditions are only sufficient [10, 11].

The above conditions also ensure the absence of strong instability ( $\sim \varepsilon$ ), due to a quasilinear (isochronous) resonance [1], especially when several ( $r$ ) independent resonance conditions (1.2) are simultaneously satisfied. The latter is called multiple ( $r$  — fold) nonlinear resonance. However, a weak instability caused by nonresonant ( $\omega_{nm} \neq 0$  for given initial conditions) terms in the perturbation series (1.1) is possible, and it is just the AD we are going to discuss in detail. Moreover, this weak instability is a typical phenomenon of nonlinear oscillations, since it occurs for almost any perturbation, particularly one that is arbitrarily weak, of a completely integrable system. The only restriction is the action space dimension  $d_a$ , which must be larger than that of the invariant torus ( $d_a > d_t = 1$ ) [3]. The torus is an absolute barrier for the motion trajectory, which can only bypass it but never go through. For a driving perturbation ( $M > 0$  in Eq. (1.1)) the minimum number of degrees of freedom is, thus,  $N_{min} = 2$ , but in the conservative case ( $M = 0$ ) it is  $N_{min} = 3$ , since the trajectory is constrained to follow an energy surface.

Even these minimal restrictions are not absolute, since they apply to the strong nonlinearity (1.4) only when the effect of resonant perturbation is small ( $\Delta I / I \sim \sqrt{\varepsilon} \ll 1$ ). In case of linear  $H_0(I)$  (the harmonic oscillator)  $N_{min}$  is smaller by 1 [12].

At least three perturbation terms in the series (1.1) are necessary for AD. We shall call each of these terms a resonance (for the appropriate initial conditions of the motion). A single resonance retains the complete integrability of the unperturbed system. The interaction of even two resonances results in the formation of narrow chaotic layers around the unperturbed separatrices of both resonances [13–15], but, the chaotic motion remains confined within a small domain of the layer. Only the combined effect of at least two driving resonances gives rise to diffusion along the layer of the first, guiding, resonance if  $N \geq N_{min}$  holds (see Ref. [1] for details).

In the first approximation (1.2) the driving perturbation terms are nonresonant ( $\omega_{nm} \neq 0$ ), but the final effect is due to the secondary resonances between the driving perturbation and

the slow phase oscillation on the guiding resonance. This is a particular case of the general rule that all the long-term effects in nonlinear oscillations are due to some resonances. For the problem in question the principal parameter is the ratio

$$\lambda = \frac{|\omega_{nm}|}{\omega_g}, \quad (1.5)$$

where  $\omega_g \sim (\varepsilon|V_g|)^{1/2}$  is the frequency of small phase oscillations at the center of the guiding resonance, and where  $V_g$  is the Fourier amplitude of the corresponding perturbation term. For a weak perturbation ( $\varepsilon \rightarrow 0$ ) the parameter  $\lambda \gg 1$  is big, and thus the effect of the driving resonances is a high-frequency one. In fact, this is equivalent to a low-frequency (adiabatic) perturbation. Hence we use the term inverse adiabaticity [14]. The symmetry between the standard and inverse adiabaticity is especially clear in a conservative system, i.e., for the interaction of coupling resonances. Indeed, in this case the resonance interaction results in energy exchange between the guiding and driving resonances. While for the former the perturbation is a high-frequency one (inverse adiabaticity), for the latter it is low-frequency (standard adiabaticity).

For an analytic perturbation the effect in both cases is exponentially small in the adiabaticity parameter  $\lambda$  (1.5), namely [1, 14]:

$$D \sim e^{-\pi\lambda} \sim w_s^2, \quad (1.6)$$

where  $D$  is the local dimensionless diffusion rate in the action  $I$  within a chaotic layer and where  $w_s \sim |\Delta H_0|/\varepsilon V_g$  stands for the dimensionless layer width (for a more accurate estimate see Ref. [14]). Notice that the effect (1.6) is of a nonperturbative nature, since  $\lambda \sim \varepsilon^{-1/2}$  (see Eq. (1.5)).

This is the simplest resonant mechanism of AD. In particular models the accuracy of such a three-resonance approximation was found to be within a factor of 2, provided that the perturbation is not too weak, i.e., the adiabaticity parameter  $\lambda$  is not very big [1] (see also Section 3 below).

As  $\lambda \rightarrow \infty$  the higher-order resonances with large harmonics numbers  $|n_i|, |m_j| \rightarrow \infty$  come into play. Even though their amplitudes  $V_{nm} \sim \exp(-\sigma k)$  drop exponentially, where  $k = \sum |n_i| + \sum |m_j|$ , the detunings  $|\omega_{nm}|$  also rapidly decrease. The operative resonances which control the diffusion have been roughly identified in Refs. [1, 15] by minimizing the expression

$$-\ln D \equiv E \sim k + \lambda(k) \gtrsim \lambda_0^{1/L} \quad (1.7)$$

with respect to  $k$ . Here  $\lambda_0 = \omega_0/\omega_g$ ,  $\omega_0$  stands for a characteristic oscillation frequency, and the following diophantine estimate was used:

$$\omega_{nm} \sim \frac{\omega_0}{k^{L-1}}. \quad (1.8)$$

The most important parameter in Eq. (1.7),

$$L = N + M - r, \quad (1.9)$$

is the number of linearly independent (incommensurate) unperturbed frequencies on an  $r$ -fold resonance. We shall call  $L$  the resonance dimension (in frequency space). Actually, Eq. (1.9)

gives the maximum dimension when all  $L$  independent frequencies contribute to the driving resonances, which may be termed the full resonances. There are also partial resonances which depend on a smaller number of frequencies  $\tilde{L} < L$ . Even though there are only a few of the latter, they are crucially important for the new AD regime which is the main subject of this paper (Section 5).

The estimate (1.7), which represents another AD mechanism, seems to agree with numerical data [7, 14]. On the other hand, Nekhoroshev rigorously proved [10] an upper bound of the form (1.7) but with a different exponent ( $M = r = 0$ ):

$$L \leq L_N = \frac{(3N - 1)N}{4} + 2. \quad (1.10)$$

Even for the minimum dimensions  $N = 3$  this upper bound  $L_{max} = 8$  considerably exceeds the estimate (1.9):  $L = 2$  ( $r = 1$ ). The difference grows as  $N \rightarrow \infty$ . Even though this discrepancy is not a direct contradiction inasmuch as Eq. (1.10) is the upper bound, it constitutes a problem: what would be the origin of the difference between the two estimates?

Recently, this problem has been resolved by Lochak [11] who rigorously proved a more efficient Nekhoroshev-type estimate with the exponent (1.9) (for  $M = 0$  but any  $r$ ). The explanation is that Lochak assumed convexity of the unperturbed Hamiltonian  $H_0(I)$  given above, whereas Nekhoroshev's proof holds under a weaker condition of the so-called steepness of  $H_0$ . From the physical point of view this difference appears to be insignificant. At least, we are not aware of any example of a steep but non-convex  $H_0$ .

Both the diffusion rate and the measure of chaotic component ( $\sim w_s$ , see Eq. (1.6)) are exponentially small in the perturbation in the limit  $\varepsilon \rightarrow 0$ , hence the term KAM integrability [14] referring to the Kolmogorov–Arnold–Moser theory which proves the complete integrability for most initial conditions as  $\varepsilon \rightarrow 0$ . This partial integrability, or better to say almost-integrability, is as good as the approximate adiabatic invariance. Notice, however, that the complementary set of initial conditions supporting AD — the so-called Arnold web — is everywhere dense, as is the set of all resonances (1.2), any one of which can be a guiding resonance. Also, the variation is exponentially slow in action  $I$  only while the variation in oscillation constant (for the unperturbed motion) phase  $\theta_0$  is much faster, with a characteristic time of order the inverse Lyapunov exponent,  $\theta_0 \sim \omega_g / |\ln w_s| \sim T_w^{-1}$ , where  $T_w$  is the oscillation period in the chaotic layer (see Eq. (2.2) below).

Both rigorous estimates are valid asymptotically, for sufficiently small  $\varepsilon$  only. For example, Lochak requires [11] ( $L \gg 1$ )

$$\varepsilon < \varepsilon_L \sim \left( \frac{\sigma^2}{L} \right)^{2L^2}, \quad (1.11)$$

where  $\sigma$  is some average decay rate of the perturbation amplitudes. This is very small perturbation, and the problem arises of estimating the diffusion rate in the intermediate asymptotic region:  $\varepsilon_L \ll \varepsilon \ll 1$ , or  $1 \ll \lambda_0 \ll \lambda_L$ . This problem was first addressed in Refs. [14], where a new regime of diffusion, called the fast Arnold diffusion (FAD), was conjectured from some preliminary results of numerical experiments. Two characteristics of the new regime as contrasted to the far-asymptotic AD (1.11) are as follows:

- (i) the dependence of the diffusion rate on the adiabaticity (perturbation) parameter  $\lambda_0$  (1.7) is a power law rather than exponential, and

(ii) the diffusion rate does not depend on the resonance dimension  $L$ , in particular, on the number of degrees of freedom  $N$  (cf. Eq. (1.7)).

Precisely this behavior has been observed in numerical experiments with another multidimensional model [16]. However, the authors of [16] have given a different interpretation of their numerical results. Instead, we tried to reconcile the same results with our new diffusion mechanism [17]. Unfortunately, both interpretations remain somewhat ambiguous because the perturbation in those numerical experiments was not sufficiently small to reach any asymptotic behavior where the theoretical estimates were expected to hold true. To resolve this ambiguity we continued numerical and theoretical studies with the same model but using a much weaker perturbation. In this paper we report on our first results and present their theoretical explanation.

## 2. MODEL AND NUMERICAL EXPERIMENTS

Following Refs. [16,17] we make use here of the same model with Hamiltonian

$$H(x, p, t) = \frac{|p|^2}{2} - K \sum_{i=1}^{N+1} \cos(x_{i+1} - x_i) \delta_1(t) \quad (2.1)$$

and periodic boundary conditions ( $x_{N+2} = x_1$ ;  $p_{N+2} = p_1$ ) where  $p, x$  are action-angle variables,  $\delta_1(t)$  stands for the  $\delta$ -function of period 1, and  $K \rightarrow 0$  is small perturbation parameter. Notice that this model has  $N$  degrees of freedom due to the additional motion integral  $\sum p_i = \text{const}$ . The unperturbed frequencies  $\omega_i = p_i$  are equal to the action variables, and the energy surfaces  $H_0(p) = |p|^2/2 = \text{const}$  are spheres, and hence are strictly convex with unit determinant (1.4). The driving perturbation in the form of periodic «kicks» is not important for the diffusion but greatly simplifies numerical experiments as it allows the use of a (multidimensional) map rather than differential equations of motion.

Even though this model does not immediately represent by itself a physical system, it is very convenient for the studies of subtle nonlinear phenomena like AD. The emerging theory can, then, be applied to real physical problems, such as the stability of the Solar System [18] or of charged particles in magnetic fields in plasma devices, accelerators and colliding beams [15, 19].

In previous work the diffusion in multidimensional models like (2.1) was studied only down to  $K \sim 0.1$  [16, 9]. For such perturbation levels and large  $N$  a considerable part of phase space becomes globally chaotic, which obscures the AD effect. Even though the combined action of AD and global diffusion is an interesting problem which is important for applications [1, 15], here we mainly wanted to understand the mechanism of AD itself. To this end we went down as far as to  $K \sim 10^{-6}$  with up to  $N = 15$  degrees of freedom. Realization of this program has required essential modification of the problem itself. This is because direct computation of the diffusion rate quickly becomes prohibitively long as  $K \rightarrow 0$ , especially since a multiple computation precision is required for such a small  $K$ . To overcome this technical difficulty we have taken a different approach [14], namely, computing the chaotic layer width  $w_s$  and recalculating the diffusion rate from a relation like (1.6). Of course, this make sense for a model with  $N \geq N_{min}$  degrees of freedom (Section 1). In this way we have managed to reach (for another model) adiabaticity parameter values of  $\lambda_0 \approx 50$  with an ordinary computer, as compared to  $\lambda_0 \approx 10$  only for a direct diffusion calculation on a Cray supercomputer [7]. In

the model (2.1) this would roughly correspond to  $K \sim \lambda^{-2} \sim 4 \cdot 10^{-4}$  and  $10^{-2}$ , respectively, and  $N = 2$  only.

In the present work we go further, and give up the calculation of the diffusion rate altogether. Instead, we are studying numerically and developing the theory of the chaotic layer only. This proves sufficient to understand the mechanism of AD as well, since both are essentially determined by the same higher-order adiabaticity parameter (1.5) and exponent in Eq. (1.7). Then, all we need in numerical experiments is to compute the oscillation period  $T(w_s)$  inside the chaotic layer of a guiding resonance, and recalculate the layer width  $w_s$  using the simple relations [1]:

$$\omega_g T_{min} = \ln \frac{32}{w_s}, \quad \omega_g T_{av} = \ln \frac{32}{w_s} + 1, \quad (2.2)$$

where  $T_{min}$ ,  $T_{av}$  are the shortest and average periods, respectively. The two values are in a reasonable agreement,  $\langle \ln(w_{min}/w_{av}) \rangle = 0.31$ , within the rms fluctuations  $\Delta \ln(w_{min}/w_{av}) = \pm 0.39$ , and both underestimate the full layer width. This is because the diffusion at the layer edge is very slow, so that the 100 oscillation periods used in numerical experiments were insufficient to reveal the whole layer. A crude estimate [14] yields the expected correction factor of order 2. No such correction was introduced into the numerical data, but it will be discussed below in Section 3.

A primary coupling resonance  $\omega_1 \approx \omega_2$  with phase oscillation frequency  $\omega_g = \sqrt{2K}$  has been chosen as the guiding resonance. Correspondingly,  $p_1 \approx p_2 \approx p_g$  while other  $p_i$  ( $i = 3, \dots, N + 1$ ) were taken at random (mod  $2\pi$ ). For the trajectory to be inside the layer the initial value of the guiding resonance phase was taken to be approximately  $\psi_1 = x_1 - x_2 \approx \pi$ . However, for small  $K$  the exact position of the layer had to be located numerically prior to computation of  $w_s$  by a special searching part of the code. The computation was performed for 7 values of  $N = 2, 3, 4, 5, 7, 9, 15$  with the same initial conditions for a single trajectory.

The results are summarized in Figs. 1 and 2. The lower bound of  $w_s \sim 10^{-22}$  was determined by the computation precision (about 30 decimal places). The values of the principal model parameter — the number of independent unperturbed frequencies, or the resonance dimension  $L = N + M - r = N$  — are also indicated. Notice that under the particular conditions of the numerical experiments the resonance dimension is equal to the number of degrees of freedom of the model because the driving perturbation is periodic ( $M = 1$ ), and guiding resonance is simple ( $r = 1$ ).

The most striking feature of the empirical data is the qualitatively different behavior for  $L = 2$  which was observed already in [16]. The rest of the data show no systematic dependence on  $L$ , but rather big fluctuations which rapidly increase with  $\lambda$ .

### 3. SMALL- $\lambda$ LIMIT: A SIMPLE DYNAMICAL THEORY

To lowest order in the small perturbation parameter  $K$  we can only consider the primary driving resonances which are explicitly present in the original Hamiltonian (2.1). Then, the problem is very similar to that studied in Ref. [1] apart from a different expression for the kinetic energy. First, we transform the variables for the two degrees of freedom which determine the guiding resonance:  $x_1, x_2, p_1, p_2 \rightarrow \psi_1, \psi_2, I_1, I_2$  where

$$\psi_1 = x_1 - x_2, \quad \psi_2 = x_1 + x_2, \quad p_1 = I_1 + I_2, \quad p_2 = I_2 - I_1. \quad (3.1)$$

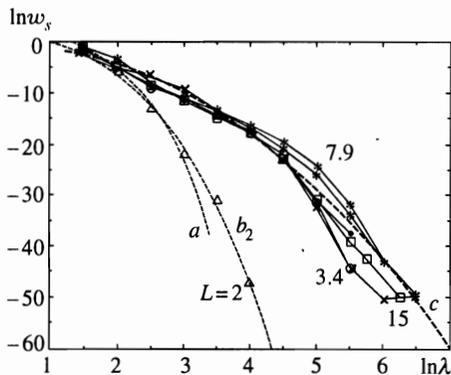


Fig. 1

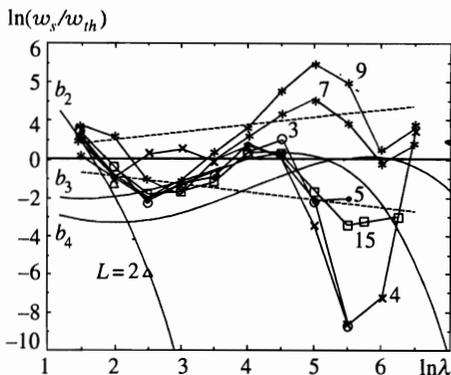


Fig. 2

Fig. 1. Summary of numerical data for the model (2.1). Broken solid lines connecting various symbols show computed values of  $w_s$  as a function of the adiabaticity parameter  $\lambda \equiv 1/\sqrt{K}$  and the resonance dimension  $L = N$  indicated by the numbers. Dotted lines represent the theory: (a) small- $\lambda$  limit, one fitting parameter, Eq. (3.5); (b) large- $\lambda$  limit for  $L = 2$ , two fitting parameters, Eq. (4.9); (c) intermediate asymptotics, three fitting parameters, Eq. (5.8)

Fig. 2. The same data as in Fig. 1, with respect to the theoretical dependence  $w_{th}(\lambda)$ , Eq. (5.8) (curve c in Fig. 1). Thin solid curves  $b_L$  represent the first three members of the family  $w_s(\lambda, L)$ , Eq. (4.9) (cf. Fig. 3). Two dashed lines show rms  $w_s$  fluctuations (5.11)

In this approximation the momentum satisfies  $I_2 \approx \psi_2$ , and all  $p_i \approx x_i$  for  $i \geq 3$  are constant and determine the frequencies of the driving resonances. The unperturbed motion on the separatrix of the guiding resonance is given by

$$\psi_1(t) = 4 \operatorname{arctg}(e^{\omega_g t}) - \pi, \tag{3.2}$$

where the frequency of the phase oscillation is  $\omega_g = \sqrt{2K}$ . As the interaction in the original Hamiltonian (2.1) is local, only the two degrees of freedom directly coupled to the guiding resonance contribute to the driving perturbation in the chaotic layer. The full set of driving resonances remains formally infinite because of the external perturbation  $\delta_1(t)$  of frequency  $\Omega = 2\pi$  but the effect of most of them is exponentially small due to the large detuning  $\omega_{nm}$  (see Eqs. (1.5) and (1.6)). Consequently, one can retain a single driving resonance only with minimal detuning:

$$\omega_d = \min |p_g - p_d + s\Omega|, \tag{3.3}$$

where  $p_d = p_3, p_{N+1}$  and  $s = 0, \pm 1$ . In this approximation the Hamiltonian takes the form  $H = H_0(I_1, \psi_1) + V(\psi_1, t)$ , where

$$H_0 = I_1^2 - K \cos \psi_1, \quad V \approx -K \cos(\psi_1/2 - \omega_d t + \phi), \tag{3.4}$$

and  $\phi$  is some constant phase.

Now, we can apply the standard method for deriving the separatrix map and the layer width (see Refs. [1, 13] for details):

$$w_s = \Delta H_0/K \approx 4\pi f \lambda_0^2 \exp(-\pi\lambda_0/2), \tag{3.5}$$

where  $\Delta H_0$  is the layer width in energy,  $\lambda_0 = \omega_d/\omega_g = \lambda\omega_d/\sqrt{2}$ , and  $\lambda \equiv 1/\sqrt{K}$ . Besides the usual approximations for such evaluations an additional factor  $f \sim 1$  shows up for the model (2.1) because the relative perturbation  $|V/H_0| \sim 1$  is not small. In the particular case  $N = 1$ , which reduces to the well studied standard map, this factor  $f \approx 2.15$  was found in numerical experiments [1], and later confirmed with a much better accuracy in [20]:  $f = 2.255\dots$  The best theoretical value recently derived is  $f \approx 2.14$  [21]. Uncertainty in this factor limits the theoretical accuracy of relation (3.5). It is partly balanced by an underestimated layer width, and also by a factor of 2 [14] as discussed above. Hence the factor  $f = f_{th}/f_n$  in Eq. (3.5) is actually the ratio of a theoretical  $f_{th}$  to the correction  $f_n = w_\infty/w_s$  of empirical  $w_s$  value (for 100 oscillation periods in our case) to obtain the true value  $w_\infty$  for infinitely many periods.

In this small- $\lambda$  region the width  $w_s$  does not depend on  $N$  (Fig. 1) because the original interaction is local. However, this region is rather narrow. A comparison of numerical data for  $L = 2$  with theory (3.5) (the dotted line  $a$ ) is presented in Fig. 1. The value of  $f = 0.64$  was obtained from the three leftmost points in Fig. 1 ( $\ln \lambda = 1.5 - 2.5$ ) with rms deviation from the theory (3.5)  $\Delta \ln w_s = \pm 0.53$ . Assuming the empirical correction  $f_n = 2$  [14] gives  $f_{th} = 1.3$ , which is rather different from that in the standard map.

#### 4. LARGE- $\lambda$ LIMIT: STATISTICAL ESTIMATES

For large  $\lambda$  the layer width, as well as the AD rate, progressively exceeds the simple estimate (3.5) (Fig. 1). This was noticed already in the first numerical experiments on AD [1]. Apparently, it is somewhat strange, at the first glance, effect of higher-harmonic driving resonances, even though they are much weaker. Generally, such resonances are present in the original Hamiltonian (1.1), and their amplitudes  $V_{nm}$  are explicitly given. However, in the model (2.1) under consideration here this is not the case, and the higher perturbation harmonics show up only in higher orders of the perturbation expansion with respect to small perturbation parameter  $K \ll 1$ . The mechanism for generating higher-harmonic terms is related to the modulation of each unperturbed frequency  $p_i$  by any other degree of freedom. In particular, this general mechanism transforms the original local interaction between degrees of freedom in the system into a global one. Approximately, the higher-order amplitudes  $V_n \sim K^n = \exp(n \ln K)$ , and their decay rate  $\sigma$  (per freedom) can be assumed in the form [17]:

$$\sigma = \ln(A/K) \tag{4.1}$$

with some constant  $A$  depending on a particular shape of the perturbation. In our model (2.1) the leading higher terms roughly correspond to  $A \sim 2$ , which we will use below. Notice that the amplitudes do not depend on the external perturbation harmonic  $m$ , since it is a  $\delta$ -function.

A counterbalance to the weaker higher perturbation terms is the smaller  $\lambda$  (1.5) due to the smaller detuning  $\omega_{nm}$  (1.2). Generally, the dependence  $\omega_{nm}(n, m, \omega)$  is very complicated, with wild fluctuations, and exact evaluation of a higher-order perturbation is practically impossible and even useless beyond a few first terms [21]. However, the leading dependence can be found as follows (see, e.g., Refs. [22, 23] and also Refs. [1, 15, 17]):

$$\omega_{nm} = \frac{\Omega}{q^{L-1}} F_{nm}(\omega), \tag{4.2}$$

where  $q = \langle |n_i| \rangle$  is average absolute value of the components of integer vector  $n$  and now the new function  $F_{nm}$  describes the fluctuations only. The latter are quite big, which is the main obstacle for reliable estimates. In some special cases the function  $F_{nm} = F_0$  is simply a constant. For example, for the case  $L = 2$  and frequency ratio  $R = \omega/\Omega = (\sqrt{5} - 1)/2$  («the most irrational» real number) we have  $1/F_0 = R + 1/R = \sqrt{5}$ . Generally, only a sort of statistical estimate can be obtained by setting  $F_{nm}(\omega) \approx F_f \approx \text{const}$  to some average value to be fitted from numerical data.

Now a particular term of the higher-order perturbation takes a form similar to Eq. (3.4):

$$V_n \sim \exp(-q\sigma(L - 1)) \cos\left(\frac{q\psi_1}{2} - \omega_{nm}t + \phi_{nm}\right), \tag{4.3}$$

where the factor  $L - 1$  is less by 1 than the full number of frequencies because of the  $\delta$ -function in the Hamiltonian (2.1), as discussed above. Assuming again that the term (4.3) provides the main contribution to the formation of the chaotic layer, which seems to be plausible owing to the big detuning fluctuations, we arrive, analogously to Eq. (3.5), at the following estimate for the layer width:

$$w_s \sim (2e\lambda_n/q)^q \exp(-E(n)). \tag{4.4}$$

Here the principal exponent is (cf. Eq. (1.7))

$$E(n) = q\sigma(L - 1) + \frac{\pi\lambda_n}{2}, \quad \lambda_n = \frac{\omega_{nm}}{\omega_g} \approx \lambda_0 \frac{F_f}{q^{L-1}}, \tag{4.5}$$

where  $\lambda_0 = \Omega/\omega_g = \lambda\Omega/\sqrt{2}$ , and  $\lambda \equiv 1/\sqrt{K}$  (Fig. 1).

The minimum of  $E(n)$  is ( $\Omega = 2\pi$ )

$$E_{min} = \sigma^p L \Lambda^{1/L}, \quad \Lambda = \frac{\pi^2}{\sqrt{2}} F_f \lambda, \quad p = 1 - \frac{1}{L}, \tag{4.6}$$

and is reached at  $q \approx q_0$ , where

$$q_0^L \approx \frac{\Lambda}{\sigma}, \quad \frac{\lambda_n}{q_0} \approx \frac{2\sigma}{\pi}. \tag{4.7}$$

The latter relation shows that the factor  $(\lambda_n/q)$  in Eq. (4.4) approximately reduces to a constant  $\sigma \rightarrow \sigma_L$  which renormalizes the amplitude decay rate, where

$$(L - 1)\sigma_L \approx (L - 1)\sigma - \ln \sigma - \ln \frac{4}{\pi} - 1 > 0. \tag{4.8}$$

The latter inequality is a necessary condition for the validity of these approximate relations. This condition is satisfied for sufficiently large original  $\sigma$ , or small  $K$  (see Eq. (4.1)).

Finally, the approximate relation for the layer width in this limit reads

$$\ln w_s \approx A_f - b(L)\sigma_L^p L \Lambda^{1/L}. \tag{4.9}$$

This theoretical dependence is also shown in Fig. 1 (curve  $b_2$ ) for  $L = 2$  and fitted values  $A_f = 5.42$ , and  $F_f = 0.34$  for the detuning parameter in Eq. (4.6). The rms deviation for 5 points ( $\ln \lambda = 2 - 4$ ) is  $\Delta \ln w_s = \pm 0.71$ . While the average detuning  $F_f$  has a reasonable value, the factor  $A_f$  seems too big (see next section). Apparently, this discrepancy characterizes the accuracy of our statistical estimates. The additional parameter  $b(L) = 1$  was set equal to unity for  $L = 2$ , and will be discussed in detail in section 5 below.

For bigger  $L$  the behavior is completely different, and this is our most interesting result to be described in the next section.

5. INTERMEDIATE ASYMPTOTICS: FAST ARNOLD DIFFUSION

The crucial change in the dependence  $w_s(\lambda)$  stems from the factor  $L - 1$  in the expression for the exponent  $E(n)$  (4.5). The effect of this factor was previously missed in Refs. [1, 15] (cf. Eq. (1.7)). Indeed, it leads to a nonmonotonic dependence  $w_s(L)$  according to Eq. (4.9). The latter was derived from optimization with respect to the average harmonic number  $q$  among the driving resonances with the maximal dimension  $L = N$  only (see Eq. (4.2)). Meanwhile, there are also resonances of lower dimension with  $\tilde{L} < L$ . Hence we need the second optimization, now with respect to  $\tilde{L}$ , as was first done in [14] (see also [17]). First, we explain the idea of optimization for a simple example (cf. Eq. (4.9))

$$w_s = \exp(-L \lambda^{1/L}) . \tag{5.1}$$

The new factor  $L$  decreases the layer width as  $L$  grows, and thus counteracts the increase in  $w_s$  due to the dependence  $\lambda^{1/L}$ . For any pair  $L_1 < L_2$  there is a certain value of  $\lambda = \lambda^*$  at which both  $w_s$  values coincide,

$$\lambda^* = (L_2/L_1)^{L_1 L_2 / (L_2 - L_1)} . \tag{5.2}$$

For  $\lambda < \lambda^*$  we have  $w_s(L_1) > w_s(L_2)$  and vice versa. Thus, for a given  $\lambda$  the particular  $\tilde{L}(\lambda)$  should be found which maximizes  $w_s$ . In this way we would obtain a broken line which is the envelope of the family of curves  $w_s(\lambda, \tilde{L})$ . Interestingly, the existence of such a family of intersecting curves could already be inferred (but was missed) from the validity of approximation (3.5) [1, 2, 6, 7] which corresponds to  $\tilde{L} = 1$ .

For  $L \gg 1$  a smooth approximation to the envelope is found from the local condition

$$\frac{dw_s}{d\tilde{L}} = -w_s \lambda^{1/\tilde{L}} \left( 1 - \frac{\ln \lambda}{\tilde{L}} \right) = 0 , \tag{5.3}$$

hence we obtaine the optimal value

$$\tilde{L}_0(\lambda) = \ln \lambda \tag{5.4}$$

and

$$w_{max}(\lambda) = w_s(\tilde{L}_0) = \lambda^{-e} , \tag{5.5}$$

where  $e = \exp(1)$ . Thus, the dependence of the layer width on the adiabaticity parameter becomes a power law, provided that  $\tilde{L}_0 \leq L$ , or

$$\lambda \leq \lambda_L = e^L , \tag{5.6}$$

i.e., for a not-too-weak perturbation. This border is, of course, much higher (in  $\epsilon$ ) than that in the rigorous theory (cf. Eq. (1.11)). We term (5.6) the intermediate asymptotic region as contrasted to the far asymptotic limit for the reversed inequality. The former is always bounded from above but rapidly grows with  $L$ , and may be arbitrarily large as  $L \rightarrow \infty$ .

We call this regime fast Arnold diffusion (FAD). Within the domain (5.6) the layer width (and diffusion rate) does not depend on  $L$  but for any fixed  $L$  and  $\lambda \rightarrow \infty$ , the Nekhoroshev-like dependence (4.9) is recovered asymptotically.

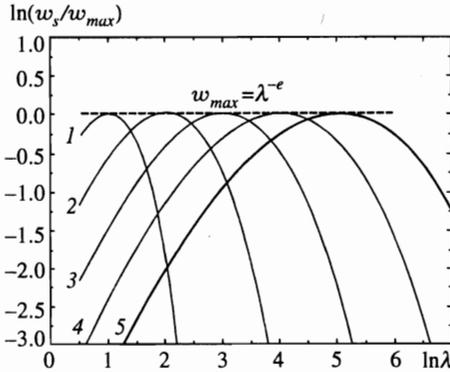


Fig. 3. A scheme of the family  $w_s(\lambda, \tilde{L})$ , for  $\tilde{L} = 1 \dots 5$  as indicated, with maximal  $\tilde{L} = L = 5$  which form the smooth power-law dependence (5.5) shown by dotted straight line

In Fig. 3 the power-law mechanism is illustrated, for the simple example (5.1), by plotting the family of curves  $\ln(w_s(\lambda, \tilde{L})/w_{max})$  which are tangent to the line of maximal  $w_{max}(\lambda)$  (5.5) up to the largest  $\tilde{L} = L = 5$ .

For the more realistic asymptotic relation (4.9) the optimization is more complicated because of the additional dependence on  $L$  via  $\sigma_L^p$ . That can be partly removed by approximate renormalization:  $\Lambda_0 \rightarrow \Lambda_0/\sigma$ . For  $L \gg 1$  the remaining dependence (4.8) is weak and can be neglected, at least in evaluating the optimal  $\tilde{L}_0$ , which now becomes (cf. Eq. (5.4))

$$\tilde{L}_0(\lambda) \approx \ln(\Lambda/\sigma). \tag{5.7}$$

However, we retain the more accurate value of  $\sigma_L$  (4.8) in the final expression:

$$\ln w_s \approx A_f - b_f e \left[ \sigma \ln \left( \frac{\Lambda}{\sigma} \right) - \ln \left( \frac{4\sigma}{\pi} \right) - 1 \right], \tag{5.8}$$

which is the main result of our studies. It is compared with numerical data in Fig. 1 (curve  $c$ , see also Fig. 2). Besides two fitting parameters previously used in Eq. (4.9) (curve  $b_2$  in Fig. 1), which now take somewhat different values:  $A_f = -1.05$ , and  $F_f = 0.4$ , we have to introduce the third one:  $b_f = 0.29$ . The fitting of empirical data has been performed for  $N = 5, 7, 9, 15$  only. We excluded data for  $N = 3, 4$  as they seem to violate the condition (cf. Eq. (5.6))

$$\Lambda \leq \Lambda_L = \frac{\pi^2}{\sqrt{2}} F_f \lambda_L \approx \sigma e^L \tag{5.9}$$

for  $\ln \lambda \gtrsim 5$  (see Figs. 1 and 2). Using the above fitted value for  $F_f = 0.4$ , and Eq. (4.1) for  $\sigma = \ln(2/K) = \ln(2\lambda^2)$  we obtain from Eq. (5.9)  $\ln \lambda_3 \approx 4.2$  and  $\ln \lambda_4 \approx 5.5$ . While the first value is close to the empirical one, the second is too large. The origin of this discrepancy is not completely clear, but it may be caused by fluctuations. Apparently, the latter are mainly related to the detuning function  $F_{nm}(\omega)$  which fluctuates with both the harmonic numbers and the set of frequencies for different  $L$ . Interestingly, while the optimal harmonic number  $q_0$  increases with  $\lambda > \lambda_L$  (4.7) it remains approximately constant,

$$q_0 \approx e \approx 3, \tag{5.10}$$

in the whole FAD region (5.9). This follows directly from Eqs. (4.7) and (5.7). Surprisingly, the above asymptotic relations remain reasonably good in spite of the relatively small  $q_0$  value

(Figs. 1 and 2). Notice, however, that the number of resonances  $\sim q_0^{\tilde{L}_0} = \Lambda/\sigma$  still increases with  $\lambda$ .

Detuning fluctuations in  $F_f$  were calculated from the numerical data using the relation (see Eq. (5.8))

$$\frac{d \ln w_s}{d \ln F_f} = -b_f e \sigma_L \approx -b_f e (0.7 + 2 \ln \lambda) , \tag{5.11}$$

which gives for the rms dispersion

$$\langle \Delta \ln F_f \rangle_4 = 0.18, \quad \text{and} \quad \langle \Delta \ln F_f \rangle_6 = 0.25. \tag{5.12}$$

The first value is the average over 4 cases with  $N = 5, 7, 9, 15$  as in the main fitting; for the second  $N = 3, 4$  are also included. The latter value is used in Fig. 2 for rms fluctuations  $\Delta \ln w_s$  according to Eq. (5.11).

The accuracy of our theory does not allow for a reasonable estimate of the factor  $A_f \approx -1$  in the main relation (5.8), whose value is considerably smaller than  $A_f \approx 5$  in Eq. (4.9). However, the value of a new fitting parameter  $b_f = 0.29$ , which we had to introduce in Eq. (5.8) instead of  $b(2) = 1$  in Eq. (4.9), is a problem for the theory. It is impossible to fit the data for large  $L$  with the latter value or vice versa, i.e., with  $b(2) = 0.3$ , as in Eq. (5.8) but for  $L = 2$ , unless one assumes the value  $F_f = 3$  in Eq. (4.9) instead of 0.3, which seems too big. In any event, something happens in going from  $L = 2$  to  $L \geq 3$ , which is obvious from the data in Fig. 1. To reconcile these data with the above theory one needs to assume a drop either in the parameter  $b$  from 1 to 0.3 (with approximately the same  $F_f \approx 0.4$ ) or in the parameter  $F_f$  from 3 to 0.4 (with approximately the same  $b \approx 0.3$  still to be explained anyway). Actually, the value  $F_f = 3$  for  $L = 2$  would contradict the rigorous upper bound  $F_f \leq 1$  [22]. So we have to understand the first possibility above.

In [17], using a somewhat different approach, the following expression has been derived for the parameter  $b$  in the relation (2.11), similar to Eq. (5.8) above:  $b \approx 1/\pi\sqrt{e} = 0.19$ . This value is close to the present empirical one,  $b_f = 0.29$ . However, the former does not fit the far asymptotic expression (4.9) for  $L = 2$ , as discussed above.

A qualitative explanation of the decrease in  $b(L)$  with  $L$  could be related to an underestimate of the perturbation Fourier amplitudes in Eq. (4.3). Indeed, we assumed that the amplitudes decay independently for each degree of freedom (factor  $L - 1$ ). However, the higher harmonics may arise in the perturbation series not individually but in groups, thus decreasing the effective parameter  $L$  or  $\sigma$ . The former possibility is excluded by the assumed expression (4.2) for detuning. Hence we guess the effective amplitude decay rate in the form  $\sigma \rightarrow b \sigma$  with empirical  $b \approx b_f \approx 0.3$ .

A different value of  $b = 1$  for  $L = 2$  is also explained in this way because in that case only a single oscillation frequency remains. However, another important question arises: is the new factor  $b(L)$  a constant for  $L \geq 3$  or does it change still further with  $L$ ? In other words, is FAD really independent of  $N$ ? Our empirical data seem to confirm such independence. Even though there are quite big fluctuations for large  $\lambda$  they do not reveal any systematic variation of  $w_s$  with  $L$ . This is especially clear from Fig. 2 where the difference between the numerical data and the theory is shown. Moreover, the theory explains even a small dip in the dependence  $w_s(\lambda)$  around  $\ln \lambda = 3$ . This results from a deviation of the approximate smoothed envelope (5.8) from the family of curves  $w_s(\lambda, \tilde{L})$ , three of which are shown in Fig. 2 (for  $\tilde{L} = 2, 3, 4$ , cf. Fig. 3) as calculated from Eq. (4.9) with the factor  $b(2) = 1$ , and  $b(3) = 0.29$ .

If the above hypothesis is true a new fitting is required, because the renormalization  $\sigma \rightarrow b\sigma$  would result in more complicated expressions than just a single factor in Eqs. (5.8) and (4.9). By doing so we have found that  $w_s(\lambda)$  according to Eq. (5.8) changed negligibly after some changes in the fitting parameters:  $A_f = -0.88$ ,  $b_f = 0.28$ ,  $F_f = 0.21$  which appear to be reasonable also. A larger change  $\delta A_f \approx 1$  occurs in the family of curves Eq. (4.9) for  $\tilde{L} > 2$ , and their agreement with the smooth envelope (5.8) worsens owing to the approximate relation (5.7). To keep the above estimates more self-consistent we neglect all these minor changes, and retain the above relations with a single parameter  $b_f = 0.29$  for  $L > 2$ . In any event, the relations, which are approximate anyway, are much simpler in this form.

Interestingly, half of the data in Fig. 1 ( $\ln \lambda \leq 4$ ,  $L > 2$ ) also fit a simple power law with exponent 6.3, which is very close to the value 6.6 obtained in [16, 17] around  $\ln \lambda \approx 2$ . However, for larger  $\ln \lambda > 4$  the deviation from such a simple dependence (it would be a straight line in Fig. 1) progressively increases in accordance with the theory (5.8).

## 6. DISCUSSION

We have performed detailed investigations into fast Arnold diffusion, a new regime of AD when the diffusion rate depends on the perturbation strength  $\varepsilon = K$ , for the models (1.1) and (2.1) respectively, or on the adiabaticity parameter  $\lambda \sim 1/\sqrt{\varepsilon} \sim 1/\sqrt{K}$  as a power law (5.8) rather than an exponential like Eq. (4.9).

We made use of a specific model (2.1) which is relatively simple and very convenient for numerical experiments with arbitrary number of degrees of freedom  $N$  but, at the same time, is rather difficult for theoretical analysis. This is because the model represents the limiting case of the local interaction between degrees of freedom. Not only between two degrees of freedom, which would model a pair interaction in a broad class of physical systems, but even further restricted to the coupling between two nearest-neighbor degrees of freedom in a chain. Moreover, the coupling is harmonic, so that only three-frequency primary resonances (for the two degrees of freedom and for the driving perturbation) with harmonic numbers  $n = \pm 1$  show up in the original Hamiltonian (2.1) independent of  $N$ . As a result, the higher-harmonic multifrequency resonances, which make the principal contribution to AD, arise only in higher-order perturbation terms, which makes the theory very difficult from the beginning. We circumvented this difficulty by a plausible and simple conjecture (4.1) for the decay rate of the high-order perturbation amplitudes. However, to reconcile the empirical data with the theory we had, later on, to further modify this conjecture by introducing the additional factor  $b(L)$  into our main relations (4.9) and (5.8). Even though we suggest in section 5 a qualitative explanation for  $b(L) \neq 1$ , the origin of this additional dependence is not yet completely clear, and it constitutes an open question in our theory. Apparently, this is related to the specific Hamiltonian (2.1) as discussed above.

The factor  $b = 0.29$ , assumed to be constant for  $L = N > 2$ , is one of the three fitting parameters in our main theoretical relation (5.8) for the FAD. As explained in section 2, we actually computed and calculated the chaotic layer width  $w_s$  related to the diffusion rate via estimate (1.6). The second fitting parameter  $F_f = 0.4$ , which describes the detuning fluctuations  $\omega_{nm}$  (1.2), also cannot be evaluated in the present state of the theory but was found numerically to have a plausible value. Finally, the third fitting parameter  $A_f$  remains completely out of theoretical reach and simply characterizes the global accuracy of the theory. We remind that all our estimates except the simplest one (3.5) are of a statistical nature, owing to the large detuning

fluctuations. Within this accuracy and fluctuations, the agreement between the empirical data and the theory as presented in Figs. 1 and 2 can be regarded as satisfactory, especially taking into account the big range of  $w_s$  variation, almost 22 orders of magnitude!

Surprisingly, all this huge range corresponds to the intermediate asymptotic region ( $1 \ll \lambda \ll \lambda_L$ , see Eq. (5.9)) with FAD, starting even at a relatively small  $L = N \geq 5$ . Even for  $L = 3$  and 4 the FAD range is apparently of a comparable size, and only for the minimal  $L = 2$  does the far (exponential) asymptotic ( $\lambda \gg \lambda_L$ ) behavior clearly show up. As already discussed in section 5, the sharp change in  $w_s(\lambda)$  from  $L = 2$  to  $L = 3$  is precisely due to the «mysterious» parameter  $b$ , which drops by a factor of 3. Unfortunately, this does not allow us to reach the far asymptotic limit and to confirm the exponential dependence (4.9) on  $N$  for  $\lambda > \lambda_L$  beyond minimal  $L = 2$ . Meanwhile, this would be important to decide on the different interpretation of  $N$ -independent diffusion for large  $N$  in [16]. The authors of the latter conjectured that this independence is a result of the local interaction in the model (2.1). This is contrary to our theory but not as yet to the direct empirical evidence. At the moment we can only remark that their reference to Wayne's theory [24] for the same model is irrelevant. Indeed, Wayne proved a long  $N$ -independent stability for very special, nonresonant, initial conditions (theorem 1.1), whereas AD occurs within chaotic layers only, i.e., also for highly specific but resonant initial conditions. Thus, the former theory is related to a global chaos rather than to KAM integrability with its peculiar Arnold web of chaotic layers.

In the case of a global interaction (1.1) with strong nonlinearity (1.4) and uniform amplitude decay rate our theory remains valid, and even becomes simpler as  $\sigma = \text{const}$ . However, the numerical experiments would be much more difficult for large  $N$ . On the other hand, both the FAD range (5.9) and the diffusion rate there depend generally on the number of incommensurate unperturbed frequencies  $L = N + M - r$  (1.9), which may be large if  $M$ , the number of driving perturbation frequencies, is large.

Fast Arnold diffusion should not be confused with the much faster diffusion in degenerate systems or those with nonconvex energy surfaces (section 1). In the latter case the diffusion mechanism is completely different. Apparently, this sort of diffusion was recently observed in numerical experiments with the classical model of the hydrogen atom in crossed electric and magnetic fields [25].

In the present study we have chosen one of the strongest primary resonances as guiding, with amplitude  $V_g = V_1 \sim K$  (section 2). In case of a high-harmonic guiding resonance ( $V_g = V_n$ ,  $n \gg 1$ ) the main effect would be a tremendous drop in the diffusion rate due to the exponential rise of the adiabaticity parameter with  $q$  (see Eq. (4.5)):

$$\lambda_n \sim \exp\left(\frac{\sigma}{2}Lq\right) \sim \exp\left(\frac{\sigma}{2}L\rho^{1/L}\right), \quad (6.1)$$

where  $\rho(n) \sim q^L$  is the density of the guiding resonances in the Arnold web with harmonic numbers up to  $q$  (cf. Eq. (4.2)). Hence, the diffusion rate in the intermediate asymptotic region drops exponentially with  $q$  or  $\rho$ , Eq. (5.8), and even as a double exponential in the far asymptotic limit!

In conclusion, our present studies confirm the previous conjecture and preliminary empirical data [14,17] concerning a new regime of fast Arnold diffusion. Moreover, we have found that in multifrequency systems ( $L \gg 1$ ), in particular, large ones ( $N \gg 1$ ), the FAD range in the perturbation (5.9) is fairly big, so that this regime appears to be typical, in a sense, and might be important in various applications.

We are very grateful to our colleagues in the University of Milano at Como for the hospitality during the essential part of this study. This work was partially supported also by the Russia Foundation for Fundamental Research (grant 95-01-01047).

## References

1. B. V. Chirikov, *Phys. Reports* **52**, 263 (1979).
2. A. Lichtenberg and M. Lieberman, *Regular and Chaotic Dynamics*, Springer (1992).
3. V. I. Arnold, *Dokl. Akad. Nauk SSSR* **156**, 9 (1964).
4. B. V. Chirikov, *Studies in the Theory of Nonlinear Resonance and Stochasticity*, preprint INP-267, Novosibirsk (1969) (Engl. trans., CERN Trans, 71-40, 1971).
5. G. V. Gadiyak, F. M. Izrailev, and B. V. Chirikov, *Proc. 7th Intern. Conf. on Nonlinear Oscillations*, Akademie-Verlag, Berlin (1977), Vol. II,1, p. 315.
6. J. Tennyson, M. Lieberman, and A. Lichtenberg, *AIP Conf. Proc.* **57**, 272 (1979); M. Lieberman, *Ann. N.Y. Acad. Sci.* **357**, 119 (1980).
7. B. V. Chirikov, J. Ford, and F. Vivaldi, *AIP Conf. Proc.* **57**, 323 (1979).
8. T. Petrosky, *Phys. Rev. A* **29**, 2078 (1984).
9. B. Wood, A. Lichtenberg, and M. Lieberman, *Physica D* **71**, 132 (1994).
10. N. N. Nekhoroshev, *Usp. Mat. Nauk* **32**, (6), 5 (1977).
11. P. Lochak, *Phys. Lett. A* **143**, 39 (1990); *Usp. Mat. Nauk* **47**, (6), 59 (1992); P. Lochak and A. Neishtadt, *Chaos* **2**, 495 (1992).
12. A. A. Chernikov, R. Z. Sagdeev, and G. M. Zaslavsky, *Physica D* **33**, 65 (1988); A. Lichtenberg and B. Wood, *Phys. Rev. A* **39**, 2153 (1989); B. V. Chirikov and V. V. Vecheslavov, in *From Phase Transitions to Chaos*, ed. by G. Györgyi et al, World Scientific (1992), p. 273.
13. N. N. Filonenko, R. Z. Sagdeev, and G. M. Zaslavsky, *Nuclear Fusion* **7**, 253 (1967); G. M. Zaslavsky and N. N. Filonenko, *Zh. Eksp. Teor. Fiz.* **54**, 1590 (1968); G. M. Zaslavsky and B. V. Chirikov, *Usp. Fiz. Nauk* **105**, 3 (1971); A. Rechester and T. Stix, *Phys. Rev. Lett.* **36**, 587 (1976); D. Escande, *Phys. Reports* **121**, 163 (1985); V. F. Lazutkin, *Dokl. Akad. Nauk SSSR* **313**, 268 (1990); V. V. Afanasiev et al., *Phys. Lett. A* **144**, 229 (1990).
14. B. V. Chirikov and V. V. Vecheslavov, *How Fast is Arnold Diffusion?* Preprint INP 89-72, Novosibirsk (1989), B. V. Chirikov and V. V. Vecheslavov, in *Analysis etc*, ed. by P. Rabinowitz and E. Zehnder, Academic Press (1990), p. 219.
15. B. V. Chirikov, *Fiz. Plasmy* **4**, 521 (1978); *Proc. Roy. Soc. Lond. A* **413**, 145 (1987).
16. K. Kaneko and T. Konishi, *Phys. Rev. A* **40**, 6130 (1989); T. Konishi, *Suppl. Prog. Theor. Phys.* **98**, 19 (1989); T. Konishi and K. Kaneko, *J. Phys. A* **23**, L715 (1990).
17. B. V. Chirikov and V. V. Vecheslavov, *J. Stat. Phys.* **71**, 243 (1993).
18. *Resonances in the Motion of Planets, Satellites and Asteroids*, ed. by S. Ferraz-Mello and W. Sessin, Univ. de Sao Paulo, Brazil (1985); J. Wisdom, *Icarus* **72**, 241 (1987); J. Laskar, *Nature* **338**, 237 (1989); *Icarus* **88**, 266 (1990); *Astron. Astrophys.* **287**, L9 (1994).
19. B. V. Chirikov, *Priroda* **7**, 15 (1982); *Nonlinear Dynamics Aspects of Particle Accelerators*, Lecture Notes in Physics, Vol. 247, Springer (1986); A. Lichtenberg, *Phys. Fluids B* **4**, 3132 (1992).
20. V. F. Lazutkin et al., *Physica D* **40**, 235 (1989).
21. V. V. Vecheslavov, *Zh. Eksp. Teor. Fiz.* **109**, 2208 (1996).
22. A. Ya. Khinchin, *Continued fractions*, Fizmatgiz, Moskva (1961) (in Russian); J. Cassels, *An Introduction to Diophantine Approximation*, Cambridge University Press (1957); W. Schmidt, *Diophantine Approximations*, Lecture Notes in Mathematics **785**, Springer (1980).
23. B. V. Chirikov, *Chaos, Solitons and Fractals* **1**, 79 (1991).
24. E. Wayne, *Comm. Math. Phys.* **104**, 21 (1986).
25. J. von Milczewski, G. Dierksen, and T. Uzer, *Phys. Rev. Lett.* **76**, 2890 (1996).