Journal of Statistical Physics, Vol. 71, Nos. 1/2, 1993

Theory of Fast Arnold Diffusion in Many-Frequency Systems

B. V. Chirikov[1] and V. V. Vecheslavov[1]

Received May 19, 1992

A previous conjecture by the authors about a new regime of Arnold diffusion with a power-law dependence of the diffusion rate on perturbation strength is confirmed by detailed theoretical evaluation. A new effect of slow (logarithmic) dependence of the power-law exponent on the perturbation parameter is conjectured. The theory developed seems to allow for a new interpretation of the recent extensive numerical experiments on Arnold diffusion in a particular many-dimensional model of Kaneko and Konishi even in the presence of some global chaos.

KEY WORDS: Arnold diffusion; dynamical chaos; nonlinear resonances; KAM integrability; numerical experiments.

1. INTRODUCTION: ARNOLD DIFFUSION

One of the most beautiful phenomena in Hamiltonian dynamics is the so-called *Arnold diffusion* (AD), a peculiar universal instability of many-dimensional nonlinear oscillations.[1,2] This global instability had been predicted by Arnold,[3] while its chaotic nature was discovered in refs. 4, 5, and 1 and further studied in detail in refs. 6–9 and 2.

First, we briefly recall the AD mechanism, which is related to the interaction of nonlinear resonances. Consider the Hamiltonian

$$H(I, \theta, t) = H_0(I) + \varepsilon \sum_{m,n} V_{mn}(I) \exp(im \cdot \theta + itn \cdot \Omega) \tag{1.1}$$

where I, θ are N-dimensional vectors of the action–angle variables; Ω is the M-dimensional vector of driving frequencies; m, n are integer vectors of N

[1] Budker Institute of Nuclear Physics, 630090 Novosibirsk, Russia.

0022-4715/93/0400-0243$07.00/0 © 1993 Plenum Publishing Corporation

and M dimensions, respectively; and ε stands for a small perturbation parameter. The dot in expressions like $m \cdot \theta$ denotes the scalar product.

A primary resonance

$$\omega_{mn} \equiv m \cdot \omega(I) + n \cdot \Omega = 0 \tag{1.2}$$

is called *nonlinear* if the unperturbed frequencies

$$\omega(I) = \frac{\partial H_0(I)}{\partial I} \tag{1.3}$$

depend on actions I, or to be more precise, if the determinant

$$\left| \frac{\partial^2 H_0}{\partial I^2} \right| \neq 0 \tag{1.4}$$

is nonzero.

For nonlinearity to stabilize the resonance perturbation, the quadratic form of the matrix $\partial^2 H_0/\partial I^2$ must be sign definite or, geometrically, the surface $H_0(I) = \text{const}$ must be convex.[14] The latter condition is a weaker one, as it may include higher polynomial forms. Both conditions are sufficient only,[14,15] but in physical applications it is an unimportant restriction (see below).

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

Even these minimal restrictions are not absolute, being related to the *strong nonlinearity* only of (1.4) when the effect of resonant perturbation is small ($\Delta I/I \sim \sqrt{\varepsilon} \ll 1$). In case of linear $H_0(I)$ (harmonic oscillator) $N_{\min}$ is less by one,[10] but we are not going to discuss this special case here.

At least three perturbation terms in series (1.1) are necessary for AD. We shall call each of these terms resonance (for the appropriate initial conditions of the motion). A single resonance retains the complete integrability of the unperturbed system. The interaction of two resonances already results in the formation of narrow chaotic layers around the unperturbed separatrices of both resonances. Yet the chaotic motion remains confined within a small domain of the size of a layer's width, as the direction of motion in the layer typically does not coincide with that of the layer itself. Only the combined effect of at least two *driving resonances* with incommensurable frequencies provides the diffusion along the layer of the first, *guiding*, resonance if $N \geqslant N_{\min}$.

In the first approximation (1.2) the driving perturbation terms are nonresonant ($\omega_{mn} \neq 0$). Yet 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

$$\lambda = \frac{|\omega_{mn}|}{\omega_g} \tag{1.5}$$

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 ($\varepsilon \to 0$), the parameter $\lambda \gg 1$ is big, and thus the driving perturbation is a high-frequency one. In effect this is equivalent to a low-frequency (adiabatic) perturbation. Hence the term *inversed adiabaticity* we use.[11]

For an analytic perturbation the effect in both cases is exponentially small in the *adiabaticity parameter* λ of (1.5), namely[11,13]

$$\frac{D}{D_0} \sim e^{-\pi\lambda} \sim w_s^2 \tag{1.6}$$

where D is the diffusion rate, $D_0 \sim \varepsilon^2 V_g^2/\omega_g \sim (\varepsilon V_g)^{3/2}$, and w_s stands for a dimensionless width of the chaotic layer. Notice that the effect (1.6) is of a nonperturbative nature, as $\lambda \sim \varepsilon^{-1/2}$.

In particular models the accuracy of this three-resonance approximation was found to be within a factor of two, provided the perturbation was not too weak, that is, the adiabaticity parameter λ is not very big.

As $\lambda \to \infty$, the higher-order resonances with $|m|$, $|n| \to \infty$ come into play. Even though their amplitudes $V_{mn} \sim \exp[-\sigma(|m| + |n|)]$ drop exponentially, the detunings $|\omega_{mn}|$ also rapidly decrease. The group of operative resonances which control the diffusion can be singled out by minimizing the expression[1,4,12]

$$-\ln \frac{D}{D_0} \equiv E \sim k + \lambda(k) \gtrsim \lambda_0^{1/L} \tag{1.7}$$

with respect to k, where $k=|m|+|n|$; $\lambda_0=\omega_0/\omega_g$, with ω_0 as a characteristic oscillation frequency, and the following Diophantine estimate is used:

$$\omega_{mn} \sim \frac{\omega_0}{k^{L-1}} \tag{1.8}$$

The most important parameter here,

$$L = N + M - r \tag{1.9}$$

is the number of linearly independent (incommensurable) unperturbed frequencies on an r-fold resonance, that is, when r independent resonance conditions (1.2) are simultaneously fulfilled.

Estimate (1.7) seems to agree with numerical data.[7,11,13] On the other hand, Nekhoroshev rigorously proved the upper bound of type (1.7) but with a different exponent[14] ($M=0$):

$$L_{\max} = \frac{3(N-1)N}{4} + 2 \tag{1.10}$$

Even for the minimal dimension $N=3$ this upper bound $L_{\max}=6.5$ considerably exceeds the estimate (1.9): $L=2$ ($r=1$). The difference becomes increasingly large as $N \to \infty$. Even though this discrepancy is not a direct contradiction, as Eq. (1.10) is the upper bound, it constitutes a problem of what would be the origin of the deviation between the two estimates.

Recently, this problem has been resolved by Lochak,[15] who rigorously proved a more efficient Nekhoroshev-type estimate with exponent (1.9) (for $M=0$ but any r). The point is that Lochak assumed convexity of the unperturbed Hamiltonian $H_0(I)$ explained 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 nonconvex H_0.

Both the diffusion rate as well as the measure of the chaotic component [$\sim w_s$, see Eq. (1.6)] are exponentially small in the perturbation $\varepsilon \to 0$: hence the term *KAM integrability*[11] referring to the Kolmogorov–Arnold–Moser theory, which proves the complete integrability for most initial conditions as $\varepsilon \to 0$. Such a partial integrability, or better to say almost integrability, is as good as the approximate adiabatic invariance. Notice, however, that the complementary set of the initial conditions supporting AD is everywhere dense, as is the set of all resonances (1.2), any one of which can be a guiding resonance. Instructively, the chaotic trajectory is not ergodic even though it approaches arbitrarily close to any point of phase space (energy surface) because of a small measure of the chaotic component.

Both rigorous estimates are valid asymptotically, for sufficiently small ε only. For example, Lochak requires[15]

$$\varepsilon < \varepsilon_L \sim \left(\frac{\sigma_0^2}{L}\right)^{2L^2} \tag{1.11}$$

where σ_0 is the decrease rate of the Fourier amplitudes for an analytic perturbation, and $L \gg 1$ is assumed. This is a very small perturbation, and the problem arises to estimate the diffusion rate in the intermediate region: $\varepsilon_L \ll \varepsilon \ll 1$. This problem was first addressed in refs. 11 and 13, where a new regime of diffusion, which can be called fast Arnold diffusion (FAD), was conjectured from some preliminary results of numerical experiments. Below we consider this interesting and surprising phenomenon in detail and compare our theoretical predictions with recent extensive computer simulations.[16–19]

2. A MORE ACCURATE ESTIMATE OF AD RATE

In a simple theory[1,4,7,12] the Nekhoroshev-type estimates are obtained by minimizing expressions like Eq. (1.7) (cf. first of refs. 15). To begin with, consider a simpler auxiliary problem with $M = 0$ (conservative system) and $r = 0$ (nonresonant frequency vector ω), hence $L = N$. There are two possibilities for the specification of the vector m norm:

$$k = \sum_{i=1}^{N} |m_i| \tag{2.1a}$$

$$|m|^2 = \sum_{i=1}^{N} m_i^2 \tag{2.1b}$$

where m_i are (integer) components of the vector m. The first one is more suitable to describe the exponential decay of perturbation amplitudes

$$V_m \approx V_0 e^{-\sigma k} \tag{2.2}$$

where σ is some average decay rate. This corresponds to the first term of the minimized sum in Eq. (1.7), which is $2\sigma k$ as $D \sim V_m^2$.

To estimate detunings $\omega_{nn} = |m \cdot \omega|$ in the second term the norm (2.1b) is more convenient. At average

$$\langle k \rangle \approx |m| \left(\frac{2N}{\pi}\right)^{1/2} \tag{2.3}$$

for an isotropic distribution of vectors m.

Writing detuning as $|m \cdot \omega| = |\omega| m_\omega$ with $m_\omega \ll 1$ the modulus of the m component along the vector ω, consider the layer of width $2m_\omega$ and radius $|m|$ perpendicular to the vector ω. Its volume in m space

$$v_\omega \approx \frac{m_\omega}{\pi} \left(\frac{2}{e}\right)^{1/2} \left(\frac{2\pi e}{N-1}\right)^{N/2} |m|^{N-1} \tag{2.4}$$

is approximately equal to the number of vectors m whose detunings obey $|m \cdot \omega| < |\omega| m_\omega$. Notice that asymptotic (in N) relations like Eq. (2.4) and similar ones below are actually fairly accurate down to $N = 2$. The minimal detuning is found from the condition $v_\omega \approx 1$, and the whole exponent in Eq. (1.7) to be minimized with respect to $|m|$ takes the form [see also Eqs. (1.6), (2.2), and (2.3)]

$$\begin{aligned} E(\lambda_0, N, |m|) &= \left(\frac{8N}{\pi}\right)^{1/2} \sigma |m| + \pi^2 \left(\frac{e}{2}\right)^{1/2} \lambda_0 \left(\frac{N-1}{2\pi e}\right)^{N/2} |m|^{1-N} \\ &\lesssim \sigma_1 \lambda_1^{1/N} N \left(\frac{N}{N-1}\right)^{1/2} \left(\frac{N-1}{\sqrt{N}}\right)^{1/N} \approx \sigma_1 \left(N + \frac{1}{2}\right) \lambda_1^{1/N} \end{aligned} \tag{2.5}$$

Here $\lambda_0 = |\omega^0|/\omega_g$, ω^0 is the frequency vector on the guiding resonance, and

$$\sigma_1 = \frac{2\sigma}{\pi\sqrt{e}} \approx 0.4\,\sigma; \qquad \lambda_1 = \frac{\pi^{3/2}}{2\sigma_1} \lambda_0 \approx 2.8\,\frac{\lambda_0}{\sigma_1} \tag{2.6}$$

The important new feature of the estimate (2.5) as compared to previous ones[1,4,7,12,15] is the additional factor $(N + 1/2)$ in the exponent E, which qualitatively changes the dependence $D(\lambda_1, N)$, as we shall see below.

Before we explain this new behavior, we need to account for M and r. The first is straightforward because in all the above estimates the total number of independent frequencies matters only, regardless of their nature, internal (ω) or driving (Ω). Hence, $L = N + M$ ($r = 0$ so far) should be substituted for N, and $|\omega^0| \to |\omega^0 + \Omega|$. One minor difference arises for the models described by maps. Then, for one driving frequency $\sigma = 0$, as the map means a δ-function time dependence, and the additional factor is reduced by one: $L + 1/2 \to L - 1/2$.

A more difficult problem is $r \neq 0$, that is, a resonant ω. This is always the case for AD, which proceeds just along resonances, so that $r \geqslant 1$. Then r frequencies can be excluded in the expression for detunings ω_{mn} using r resonance conditions (1.2). Hence, in estimates (2.4), and (2.5), $L = N + M - r$ should be taken for N. However, the first term in Eq. (2.5) becomes, generally, very complicated because after exclusion of r frequencies the new components m_i become some combinations of the old ones, and the simple expresion (2.2) for the perturbation amplitudes no longer holds. The difference is the greater, the larger are the multiplicity r and the vector m^g of the guiding resonance. On the other hand, the fastest (at average) diffusion corresponds to minimal $r = 1$ (simple resonances which occupy the largest part of the phase space as compared to multiple resonances with $r > 1$) and to minimal $|m^g| \sim 1$ with maximal V_g and ω_g. For this reason we neglect below these complications, and retain the above estimates with $L = N + M - r$ instead of N.

Now let us discuss the consequences of the new factor $(L + l)$ $(l \sim 1)$ in the exponent E of (2.5). It decreases the diffusion rate as L grows, and thus counteracts the increase in D due to the dependence $\lambda_1^{1/L}$. For any pair $L_1 < L_2$ there is a certain value of $\lambda_1 = \lambda_1^*$ at which both diffusion rates coincide. Namely

$$\lambda_1^* = \left(\frac{L_2 + l}{L_1 + l}\right)^{L_1 L_2/(L_2 - L_1)} \tag{2.7}$$

If $\lambda_1 < \lambda_1^*$, the rate $D(L_1) > D(L_2)$, and vice versa.

For a given model with certain L the driving resonances may include any number of independent frequencies $\tilde{L} \leqslant L$. Then, for a given λ_1, the particular $\tilde{L}(\lambda_1)$ should be found which provides the maximal diffusion rate. In this way we would obtain a broken line which is the envelope of the family of curves $E(\lambda_1, \tilde{L})$ of (2.5). Interestingly, the existence of such a family of intersecting curves could be inferred already (but was missed) from the validity of the three-resonance approximation[1,2,6–8,11,13] (see Section 1), which corresponds to $\tilde{L} = 1$, while $L = 2$ for $N = N_{\min}$ $(r = 1)$.

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

$$\frac{\partial E}{\partial \tilde{L}} = E \cdot \left(\frac{1}{\tilde{L} + l} - \frac{\ln \lambda_1}{\tilde{L}^2}\right) = 0 \tag{2.8}$$

Whence $(\lambda_1 \gg 1,\ \tilde{L} \gg l)$

$$\tilde{L}(\lambda_1) \approx \ln \lambda_1 + l = \ln \lambda_2; \qquad \lambda_2 = \lambda_1 e^l \tag{2.9}$$

and

$$E(\lambda_1, L) \approx \sigma_1 e \ln \lambda_3; \qquad \lambda_3 = \lambda_1 e^{2l} \tag{2.10}$$

Hence [see Eq. (1.7)]

$$D(\lambda_1) = D_0 \lambda_3^{-e\sigma_1} \tag{2.11}$$

and the dependence of the AD rate on the adiabaticity parameter λ_1 becomes a *power law* provided $\tilde{L} \leqslant L$, or

$$\lambda_2 \lesssim e^L \tag{2.12}$$

that is, for a not too weak perturbation. This border is, of course, much higher (in ε) than that in the rigorous theory [cf. Eq. (1.11)].

Apart from some numerical factors, the main result of this paper, Eq. (2.11), confirms our previous estimate.(11,13)

Since now the dependence $D(\varepsilon)/D_0$ is a power law (2.11), the quantity(13)

$$D_0 \sim \frac{\varepsilon^2 V_g^2}{\omega_g} \sim |\varepsilon V_g|^{3/2} \tag{2.13}$$

is equally important, where the subscript g refers, as usual, to the guiding resonance.

We shall call this regime the *fast Arnold diffusion* (FAD). Within the domain (2.12) the diffusion rate does not depend on L, but asymptotically, as $\lambda \to \infty$, the Nekhoroshev-like dependence

$$D(\lambda_1, L) \approx D_0 \exp[-\sigma_1(L + l)\lambda_1^{1/L}] \tag{2.14}$$

is recovered.

Interestingly, in the degenerate case when $H_0(I) = \omega \cdot I$ (Section 1), the AD rate is also a power law, but for another reason.(10,13)

3. NUMERICAL EXPERIMENTS

The asymptotic estimate (2.14) was well confirmed numerically(7,11,13) with $\sigma_1 \approx 2.8$ for the particular model used ($N = 2$; $M = 1$; $r = 1$). A restricted set of preliminary data(11,13) seems to agree also with the theoretical prediction (2.11) for the intermediate AD regime ($N = 2$; $M = 2$; $r = 1$).

Recently the results of extensive numerical studies of AD were reported(16–19) for the model with Hamiltonian

$$H(x, p, t) = \frac{|p|^2}{2} - K \sum_{i=1}^{N} \cos(x_{i+1} - x_i)\, \delta_1(t) \tag{3.1}$$

and periodic boundary conditions ($x_{i+N}=x_i$; $p_{i+N}=p_i$), where p, x are action–angle variables, $\delta_1(t)$ stands for the δ-function of period 1, and we have slightly changed the notations in refs. 16–19: $2\pi x_i \to x_i$, and $2\pi p_i \to p_i$. Notice that the number of freedoms in this model is $N-1$ due to the additional motion integral $\sum p_i = \text{const}$, hence the principal parameter $L=N-1$ ($M=r=1$).

The main numerical results, confirmed by our own computations, are shown in Fig. 1, reproduced from ref. 17. Apart from a strong perturbation, they look, indeed, like a power-law dependence with the exponent $\gamma \approx 6.6$ almost independent of N (besides $N=3$, see below). The authors also fit the data to a power law, but obtained smaller γ values that, moreover, depended on N, because they included the intermediate region of large K where the diffusion mechanism might be different. Essentially they tried to fit the asymptotic Nekhoroshev-like dependence (2.14), but failed because the principal exponent ($1/L$) turned out to be independent of N. Also, they studied a different, global, model where the perturbation couples all the freedoms, unlike the local model (3.1) with the pair interaction only. However, the results are essentially the same (cf. Figs. 1 and 2 in ref. 17). Below we restrict ourselves to the model (3.1), and present a different interpretation of these numerical results.[17]

The strongest guilding resonances for AD are $p_i = \omega_i = \omega_{i+1}$ (for any $i=1,..., N$), whose frequency is $\omega_g = (2K)^{1/2}$. Due to periodicity in p_i and symmetry with respect to $p_i = \pi \bmod 2\pi$, the average $\langle \omega_i^0 \rangle = \pi/2$ (for the random initial conditions used in the computation), and $\langle |\omega^0 + \Omega|^2 \rangle \approx$

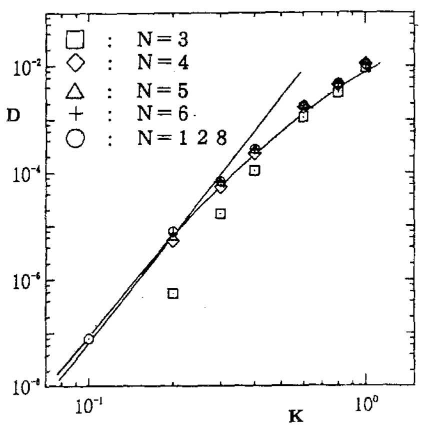

Fig. 1. Diffusion rate D vs. perturbation parameter K in model (3.1) (after ref. 17). The straight line shows a power-law dependence with exponent $\gamma = 6.6$ (our fit); the curve is the theoretical dependence for FAD, (5.2)–(5.4), with $A=75$; $l=-0.24$.

$(\pi/2)^2 L + (2\pi)^2$. From Eqs. (2.11) and (2.13), $\gamma = 3/2 + e\sigma_1/2 \approx 6.6$ (numerically) and $\sigma_1 \approx 3.8$, which is comparable with the previous result mentioned above ($\sigma_1 \approx 2.8$; $L = 2$) as well as with factor π for $L = 1$, (1.6).

The full numerical dependence can be approximately described near minimal $K \sim 0.1$ as (straight line in Fig. 1)

$$D(K) \approx \alpha K^{\gamma} \tag{3.2}$$

with $\alpha \approx 0.3$, $\gamma \approx 6.6$. The theoretical value for the factor α is found from Eq. (2.11), where we should put $\langle \omega_i^0 \rangle = \pi/2$ for $|\omega^0 + \Omega|$ in the expression for λ_0 [see Eqs. (2.6), (2.10)]. This is because the additional factor $\sqrt{L}$ is absorbed in the complicated dependence $E(L)$ and does not change the final asymptotic estimate (2.5). In the case of $L \gg 1$ the term $\Omega = 2\pi$ can also be dropped from the expression for $|\omega^0|$. Assuming $D_0 \approx K^{3/2}/4\pi^2$ [see Eq. (2.13)] with an additional factor $4\pi^2$ due to the different notations explained above, we have

$$\alpha_T \approx \frac{1}{4\pi^2} \left(\frac{4\sqrt{2}\sigma_1}{\pi^{5/2} e^{2l}} \right)^{e\sigma_1} \tag{3.3}$$

From the empirical $\alpha \approx 0.3$ and $\sigma_1 \approx 3.8$ we obtain $l \approx 0$ instead of the expected -0.5, which is not a big difference if $L \gg 1$.

Anomalous behavior for $N = 3$ ($L = 2$) also can be explained, at least qualitatively, by the existence of the FAD border (2.12) with

$$K_b \approx \frac{\pi^5 e^{2l}}{32\sigma_1^2 (\lambda_2^*)^2} \tag{3.4}$$

To reach agreement with the empirical $K_b \approx 0.3$ (Fig. 1), we would need $l \approx 1.6$, which seems too big, while for $l = 0$ the estimated border $K_b \approx 0.01$ is too small. Apparently, the asymptotic relations we use are too poor for such a small value of $L = 2$. One way to improve the agreement is to take the opposite limit $|\omega^0 + \Omega| \approx \Omega = 2\pi$ instead of $\pi/2$ in Eq. (3.4), which gives $K_b \approx 0.2$. Another difficulty is related to a very complicated (and unusual) motion structure for the particular conditions of the numerical experiments.[16–19] We proceed with a preliminary discussion of such a structure and some of its consequences.

4. FAD AND GLOBAL CHAOS

Usually, AD is assumed to proceed through a set of narrow chaotic channels along resonance surface (1.2), so that the relative measure of the chaotic component $\mu_s \ll 1$ is very small. This is certainly true for a sufficiently weak perturbation ($K \to 0$). However, in the numerical experiments under discussion[16–19] this is apparently not the case. According to the data in ref. 18, $\mu_s \approx 1$ for $K \geqslant 0.1$ and $N \geqslant 5$, that is, the phase space is almost entirely chaotic, just opposite to the expected structure. Our own numerical experiments with model (3.1) (to be published elsewhere) confirm this important conclusion.

Does it make any sense in such a complicated structure of motion to single out a very specific AD within narrow chaotic layers? We think it does, and the reason is that the global chaos for sufficiently small $K \ll 1$ is apparently provided by the overlap of high-order resonances which produce very inhomogeneous and rather slow diffusion for large N (see ref. 20 and below). In these conditions the main diffusion flow can still follow chaotic layers of the strongest primary resonances as was assumed in the above estimates. This picture was qualitatively confirmed[19] by observation of a "clustered" motion, that is, the correlation between different freedoms of the model. Such "clusters" have a finite lifetime because the motion trajectory both enters (forming a cluster) and leaves (destroying a cluster) chaotic layers. In the special case of $N = 3$ some primary resonances are clearly seen in Fig. 1a of ref. 18 (see also the end of this section). The described structure of the motion is further confirmed by our measurements of the maximal Lyapunov exponent, which was found to scale as $\Lambda \sim \omega_g = (2K)^{1/2}$ and hence to correspond to the main primary resonances. These seem, indeed, to play a special role in spite of global chaos.

The presence of global chaos apparently explains the surprising independence of the average diffusion rate on initial conditions.[16,17] Then, a more realistic estimate for $D(K)$ requires some averaging over the phase space. Since the details of the motion structure are unknown as yet, we made use of a very crude approximation, multiplying the diffusion rate (2.11) within a layer by its width w_s [see Eq. (1.6)] as a relative measure μ_A of the chaotic component supporting AD:

$$\mu_A \sim w_s \sim \left(\frac{D}{D_0}\right)^{1/2} \tag{4.1}$$

Whence [see Eq. (2.11)]

$$D \to \langle D \rangle \approx D_0 \lambda_3^{-3\sigma_1 e/2} \tag{4.2}$$

Thus, the main effect of averaging is the renormalization of the parameter $\sigma_1 \to 3\sigma_1/2$. For a given numerical exponent $\gamma = 6.6$ the new value is $\sigma_1 = 2.5$. From the numerical value $\alpha \approx 0.3$ we have the parameter

$l = -0.23$, which is much closer to the expected $l \approx -0.5$. With $|\omega^0 + \Omega| \approx \Omega$ the border value of $K_b \approx 0.3$ is also well in agreement with numerical data (cf. Section 3).

The global chaos apparently entails a peculiar transient behavior. The diffusion rate (2.11) or (2.14) is reached only on a rather long time scale, $t_D \approx 1/20D$,[16] while for a shorter time the formally computed $D \sim t^{-\beta}$ decreases, roughly but not exactly as a power law. Figure 2 shows that the exponent β increases as K decreases, approaching $\beta = 1$ for $K \to 0$. The latter means, actually, the absence of diffusion, that is, some stationary oscillations. This effect is well known and commonly used as a check for the diffusive character of the motion.[7] However, the difference $(1 - \beta)$ may indicate some slow diffusion due to the global chaos. Indeed, in numerical experiments we observed transitions between primary resonances. Remarkably, this diffusion is not only slow, but its rate drops in time, which could explain why eventually AD, also fairly slow, dominates in spite of the presence of global chaos.

Another interesting transient regime of diffusion occurs on a very short time scale, $\sim K^{-1/2}$, which is the oscillation period on a guiding resonance. This regime is seen in the data of ref. 16 (but was not discussed there), and it is especially clear in the scaled variables of Fig. 2, which represents the results of our computations. A peculiar feature of this regime is the maximal diffusion rate $D = K^2/4\pi^2$ apparently due to uncorrelated initial phases x_i used in the computation.[16] The correlation is building up due to the

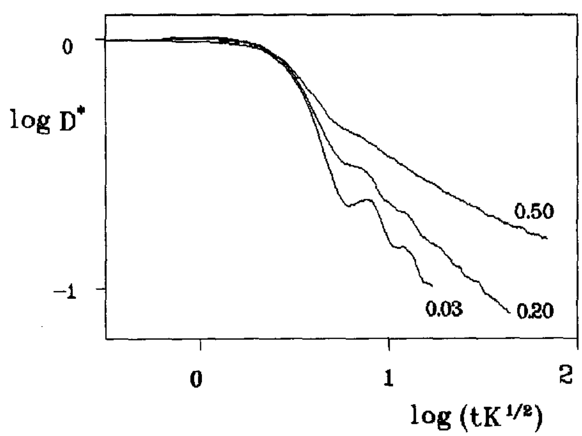

Fig. 2. Transient diffusion in model (3.1): normalized diffusion rate $D^* = 4\pi^2 D/K^2$ vs. $K^{1/2}t$; $N = 128$; random initial conditions; numbers on the curves are K values; logarithms are decimal.

resonant oscillation with frequency $\omega_s = (2K)^{1/2}$, which also affects the later behavior (Fig. 2). This is another manifestation of transient clusters[19] as well as of special role of strong primary resonances.

5. A NEW CONJECTURE

An important parameter controlling the AD rate, both the intermediate fast one, (2.11), as well as the asymptotic one, (2.14), is σ, the rate of perturbation Fourier amplitude decay. At first glance it is determined by the perturbation function $V(I, \theta, t)$ [see Eq. (1.1)]: $\sigma = \sigma_0$. Yet, for a broad class of dynamical systems with a finite number of perturbation Fourier harmonics this is not the case, and formally $\sigma_0 = \infty$. Clearly, the whole spectrum of the perturbation arises in higher orders of the perturbation theory and has the form of a polynomial series of the original perturbation. Hence, $V_{mn} \sim K^k$, where k is defined in Eq. (2.1a). Comparing with Eq. (2.2), we can assume that

$$\sigma = \ln \frac{A}{K} \tag{5.1}$$

where A is roughly some constant independent of K and N. Generally, the latter is certainly not true, but as we shall see, it seems to be a reasonable approximation, at least in a restricted domain of K values.

Consider first the FAD as described by Eq. (4.2). For the model under consideration, (3.1), assumption (5.1) leads to the expression

$$\langle D \rangle \approx \frac{K^{3/2}}{4\pi^2} \left(\frac{4e^{-2l}}{3e\pi^{3|2}} \cdot \frac{q}{\lambda_0} \right)^q \tag{5.2}$$

where the exponent q now slowly depends on the perturbation:

$$q = \frac{3\sqrt{e}}{\pi} \ln \frac{A}{K} \approx 1.57 \ln \frac{A}{K} \tag{5.3}$$

From the empirical $\sigma_1 = 2.50$ (for $K = 0.1$) the new parameter $A = 65$, but a better fitting to numerical data is achieved for $A = 75$. The latter value would give $\sigma_1 = 2.56$ and the power-law exponent $\gamma = 6.7$, which is also compatible with numerical data (Fig. 1). In both cases the empirical value is $l \approx -0.24$. Using $\lambda_0 = (\pi/2)/(2K)^{1/2}$, as explained in Section 3, we arrive at the final very simple relation:

$$\langle D \rangle \approx \frac{K^{(3+q)/2}}{4\pi^2} \left(\frac{q}{7.6} \right)^q \tag{5.4}$$

This theoretical dependence is also shown in Fig. 1 by the solid curve. The agreement with numerical data is surprisingly good in the whole range of K values (the average discrepancy is 13% only). In particular, the new relation (5.4) well describes the deviation from the approximate power-law dependence (straight line in Fig. 1) which was our starting point.

The new quantity A is, of course, a sort of fitting parameter. Yet this is not a purely empirical fitting, but rather a physically meaningful one even though we have as yet no idea about how one could estimate this parameter theoretically. In contrast, the second empirical parameter $l = -0.24$ turns out to be close to the expected value $l \approx -0.5$.

Without averaging (4.2), the original estimate (2.11) can also be fitted to the numerical data but with quite different $l \approx -0.03$ and $A \approx 950$ (the average accuracy is about 25%).

We think that these results can be considered as a preliminary confirmation of the new conjecture on the variation of the parameter $\sigma(K)$. Moreover, this effect appears to be typical if the initial perturbation is represented by a truncated Fourier series, so that its parameter $\sigma_0 = \infty$. For a finite σ_0 the effect in question is restricted to a sufficiently large perturbation when $\sigma(K) < \sigma_0$, or

$$K > Ae^{-\sigma_0} \tag{5.5}$$

For further studies of this effect, numerical experiments with a much weaker perturbation are most important, which, however, would require increasingly heavy computation. The optimal region is near the border (2.12), where the diffusion rate is maximal (for a given K) while N is minimal, both factors decreasing the computation time.

6. CONCLUDING REMARKS

The present theory confirms the previous conjecture[11,13] that in the intermediate domain ($1 \ll \lambda_2 \lesssim e^L$) of moderately weak perturbation in a many-frequency system ($L \gg 1$) the AD rate decays approximately as a power law in the adiabaticity parameter λ_1, (2.11). In this sense the diffusion is "fast" as compared to the asymptotic ($\lambda \to \infty$) exponential decay (2.14). Generally, only the total number of unperturbed frequencies both internal and driving $(N + M)$ matters, provided the number of freedoms is $N \geq N_{\min}$. A surprising feature of FAD is independence of L, (1.9), which was a puzzling observation in recent numerical experiments.[16,17] In the theory presented above there is no L dependence at all, as the diffusion rate is determined by some $\tilde{L} < L$, (2.9). Numerically, a slight dependence exists which is about, on average, 20% between $N = 4$ and 128, but only 2%

between $N = 6$ and 128. Apparently this is related to a smooth transition between the two AD regimes instead of a sharp border (2.12) in our simplified theory.

Another interesting result is that a very specific AD persists and, moreover, dominates, as $t \to \infty$, in the presence of some global chaos. Apparently this is explained by the very complicated structure of the latter,[16,17] where the diffusion rate drops in time. However, the complete picture of such a structure is far from clear. In particular, the appropriate procedure of phase-space averaging for the diffusion rate (4.2) is not known. The whole AD structure is very complicated, indeed.[15]

Detailed analysis of numerical data[16,17] has led us to a new conjecture: the parameter σ, (2.2), controlling the high-order Fourier components of the perturbation and, hence, the diffusion rate, is not a constant as was assumed before, but slowly depends on the perturbation, (5.1). Numerical data seem to confirm this new effect, but additional numerical experiments for a weaker perturbation are certainly desirable.

Restricted by inequality (5.5), this is not a generic phenomenon, as typically $\sigma_0 \neq 0$. Yet, in the intermediate region of FAD, it may be very important.

The theory presented above is rather simple—perhaps, oversimplified—but we hope that it catches the decisive features of this beautiful phenomenon of resonance interaction, Arnold diffusion.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to P. Lochak for stimulating discussions, and to T. Konishi and K. Kaneko for making available to us their important results prior to publication. The computation was performed on a PC with a fast transputer module. We are very grateful to G. A. Aksenov, V. V. Shilo, and T. B. Bolshakov for assistance with this new technique.

REFERENCES

1. B. V. Chirikov, *Phys. Rep.* **52**:263 (1979).
2. A. Lichtenberg and M. Lieberman, *Regular and Stochastic Motion* (Springer, 1983).
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) [English translations, CERN Transactions, 71-40, (1971)].
5. G. V. Gadiyak, F. M. Izrailev, and B. V. Chirikov, in *Proceedings 7th International Conference 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).
7. B. V. Chirikov, J. Ford, and F. Vivaldi, *AIP Conf. Proc.* **57**:323 (1979).

8. M. Lieberman, *Ann. N. Y. Acad. Sci.* **357**:119 (1980).
9. T. Petrosky, *Phys. Rev. A* **29**:2078 (1984)
10. A. A. Chernikov, R. Z. Sagdeev, and G. M. Zaslavsky, *Physica D* **33**:65 (1988); B. V. Chirikov and V. V. Vecheslavov, The structure of a weakly nonlinear resonance, preprint INP 91-92, Novosibirsk (1991).
11. B. V. Chirikov and V. V. Vecheslavov, KAM integrability, in: *Analysis etc.* P. Rabinowitz and E. Zehnder, eds. (Academic Press, 1990), p. 219.
12. B. V. Chirikov, *Fiz. Plasmy* **4**:521 (1978).
13. B. V. Chirikov and V. V. Vecheslavov, How fast is the Arnold diffusion? Preprint INP 89-72, Novosibirsk (1989).
14. N. N. Nekhoroshev, *Usp. Mat. Nauk* **32**(6):5 (1977).
15. P. Lochak, *Phys. Lett. A* **143**:39 (1990); Canonical perturbation theory via simultaneous approximation, *Usp. Mat. Nauk* **47**:59 (1992); P. Lochak and A. Neishtadt, *CHAOS* **2**:495 (1992).
16. K. Kaneko and T. Konishi, *Phys. Rev. A* **40**:6130 (1989).
17. T. Konishi and K. Kaneko, *J. Phys. A* **23**:L715 (1990).
18. T. Konishi, *Suppl. Prog. Theor. Phys.* **98**:19 (1989).
19. T. Konishi and K. Kaneko, Clustered motion in symplectic coupled map systems, preprint DPNU-91-54 Nagoya (1991).
20. B. V. Chirikov, *Chaos, Solitons and Fractals* **1**:79 (1991).