Doklady Physics, Vol. 47, No. 4, 2002, pp. 261–266. Translated from Doklady Akademii Nauk, Vol. 383, No. 4, 2002, pp. 472–477.

PHYSICS

Fractal Structure of a Colloidal Aggregate

A. Yu. Zubarev and A. O. Ivanov

Presented by Academician V.P. Scripov October 30, 2001

Received November 19, 2001

Colloid systems are common in nature and are actively used in current technologies. Over the last century, aggregate stability and aggregation kinetics have been among the fundamental subjects in studies of colloids. Traditional approaches to the theoretical analysis of aggregation processes date back to Smoluchowski's pioneering study [1] and are based on solving the set of kinetic equations of coagulation. A large number of studies (see, e.g., [2, 3] and the references cited therein) were devoted to the analytical and numerical study of the set of Smoluchowski kinetic equations. Exact solutions were found for constant kinetic coefficients; and the self-similar behavior of solutions was investigated for various dependences of these coefficients on the sizes of aggregates, including behavior under the conditions of sedimentation and in shear flows.

At the same time, special features of the internal structure of aggregates that appeared received insufficient attention in these approaches. Similar processes of clusterization are also observed in molecular systems, e.g., chemical deposition in supersaturated solutions, the growth of crystals in supercooled melts, etc. In certain cases, the resulting aggregates have a branching treelike structure characterized by spatial self-similarity. These objects were named fractal clusters. The developed methods of computer simulation [4–8] made it possible to reveal the main regularity of the fractal structure of these clusters: the total number N of particles (molecules) involved in the cluster depends on its dimension Σ by the power law $N(\Sigma) \sim \Sigma^{d_f}$. The exponent d_f was named the fractal dimension of the cluster, and its determination is the prime objective of numerous theoretical and computer investigations.

Until now, information about the structure of fractal clusters has been primarily acquired by computer simulation. The configurations of a system of particles with a specified mechanism of motion (Brownian or free-molecular), the type of particle interaction, and the kinetics of particle coalescence are reproduced by numerical methods. When time evolution is simulated by the multiple step-by-step variation of the configuration, the distribution of particles inside clusters is obtained in the form of a highly branching structure. Statistical averaging over the ensemble of clusters makes it possible to determine their fractal dimension depending on the mechanism of particle motion and the type of coagulation kinetics. The concept of this method dates from the model of diffusion-limited aggregation (DLA) [5], which considers the alternate motion of solitary particles and their addition to a cluster that has already been formed. Later, this approach was generalized to more complicated types of aggregation (see, e.g., [4, 6–8]. The main result of this approach is the fundamental conclusion that the fractal dimension is a universal characteristic whose value depends on the space dimension: $d_f \approx 2.5$ for three-dimensional bulk structures and $d_f \approx 1.7$ for two-dimensional surface formations. The available theoretical models are applied to describe the features of the fractal structure of computer-designed clusters with an *a priori* specified fractal branching structure. At the same time, the application of computer simulation results to particular types of aggregates, e.g., colloid aggregates, faces a number of problems, primarily because a computer experiment cannot actually reproduce the simultaneous diffusion motion of a large number of particles in a colloid system. Moreover, it is very difficult to trace the effect of particular physical and chemical conditions realized in a colloid on the internal structure of aggregates by computer methods. For this reason, it is important to develop analytical models of the aggregation of colloid dispersions. These models combine the abilities to acquire information about the internal structure of colloid aggregates and to determine the regularities of the evolution of an ensemble of aggregates.

This study is devoted to developing the corresponding model of the growth of an individual colloid aggregate formed due to the attachment of solitary particles. Contrary to the above theories, our model involves no prior assumptions on the internal structure of the cluster. The model also takes into account that the diffusion motion of particles inside the cluster occurs in constrained conditions, and combination and recombination are complicated due to the presence of an excluded volume formed by the particles of the aggregate frame. Since the diffusion motion of colloid particles is rather slow, we analyzed the case where the growth of an

Ural State University,
pr. Lenina 51, Yekaterinburg, 620083 Russia

1028-3358/02/4704-0261$22.00

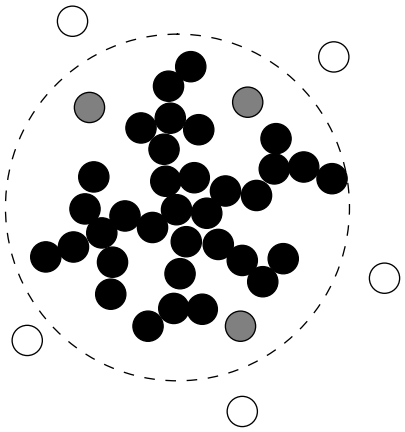


Fig. 1. Colloid fractal aggregate. The black, gray, and white circles are aggregate-frame particles, free *n*-class particles, and free σ-class particles, respectively. The dashed line is the boundary of the aggregate.

aggregate was limited by the rates of the diffusion supply of particles to its boundaries. Under these conditions, the attachment of free particles and the detachment of frame particles can be internally balanced. A fundamentally new physical result is that the growing aggregate is characterized at this dynamic balance by a power (!) dependence of the concentration of bound particles on a spatial coordinate that is a basic property of the fractal cluster. The fractal-dimension value of $d_f = 2.5$ obtained analytically is independent of the particular physical and chemical parameters of the colloid system and coincides with experimental data [9–13] and computer results for three-dimensional clusters in the classical DLA mode.

The distributed model of the growth of a colloid aggregate is based on using an aggregated-particle concentration φ averaged over angles. This concentration depends on time t and distance r from the formal center of the aggregate. For three-dimensional fractal clusters, this concentration depends on the coordinate according to the power law $\varphi(r) \sim r^{d_f - 3}$. This function carries incomplete information about the degree of aggregate branching. However, a nonuniform distribution $\varphi(t, r)$ points to the looseness of the aggregate and discloses the regularities of the evolution of its internal structure in the course of its growth. Therefore, the model can be formulated on the basis of the concept of coexisting interpenetrating continua [14] under the following simplifying assumptions.

(i) Colloid particles can be divided into three classes: particles bound in the aggregate frame and free particles inside and outside the aggregate. The structure of the cluster is described by the volume concentration $\varphi(t, r)$ of aggregated particles coexisting with solitary particles of concentration $n(t, r)$ inside the cluster (Fig. 1). Outside the aggregate, there are only solitary particles whose concentration is $\sigma(t, r)$. For close interparticle bonds, the hydrodynamic mobility of aggregated particles can be neglected compared to the mobility of solitary particles. Since we consider quantities averaged over angular variables, all the densities depend only on time and the radial coordinate r.

(ii) The aggregate occupies a spherical spatial region of radius $\Sigma(t)$ corresponding to the bound particle most distant from the center (Fig. 1). The motion of the formal boundary $\Sigma(t)$ is determined by the balance between the rate of diffusion supply of solitary σ particles from the system bulk to the cluster surface and the rate of their attachment to outer aggregated particles, i.e., the rate of transition of particles from class σ to class φ at the surface $r = \Sigma(t)$.

(iii) We consider a loosely concentrated colloid system; therefore, the interaction between solitary σ particles can be neglected. The evolution of the structure of a randomly formed cluster is determined by the kinetics of the attachment of solitary n particles to φ particles of the aggregate frame, i.e., the transition of particles from class n to class φ inside the aggregate region $r < \Sigma(t)$. The probability of this transition is proportional to the product of the densities φn and is much higher than the probability of the coagulation of solitary n and σ particles. Therefore, the formation of doublets, triplets, etc. of free particles can be neglected. The aggregation is assumed to be reversible; therefore, the detachment of particles from the aggregate frame is also taken into account. The probability of the reverse transition from class φ to classes n and σ can be very low, but this process always takes place in real systems.

(iv) Inside a cluster, solitary and bound particles are considered as interpenetrating coexisting continua. Therefore, it is convenient to describe the $n \longrightarrow \varphi$ (combination) and $\varphi \longrightarrow n$ (recombination) transitions by mass-exchange terms in the diffusion equations.

According to general thermodynamic principles, the diffusion flow of solitary n particles inside the aggregate region is determined by the degree of inhomogeneity of their chemical potential μ:

$$\mathbf{j} = \beta n \nabla \mu, \quad \mu = kT \ln\left(\frac{n}{1 - \varphi/\varphi_m}\right), \qquad (1)$$

where β is the hydrodynamic mobility of solitary colloid particles. This form of the chemical potential μ implies that the combination of particles is associated with the presence of a deep, narrow minimum in the energy of interparticle interaction. Since this potential well is narrow, we can ignore the interparticle interaction when particles are not in contact. When they come in contact, a solitary particle can be attached to the aggregate frame with a certain probability. The attachment probability depends naturally on the parameters of the energy barrier between particles. On the other hand, the aggregated particles in a certain representative volume occupy a part that is inaccessible to solitary n particles. Therefore, the chemical potential μ must include these constrained conditions. Relationship (1) has the form of the simplest van der Waals approxima-

tion, and $\varphi_m \approx 0.4$–0.6 is the concentration of a random close packing.

The equation of the diffusion of free particles inside the aggregate region has the form

$$\frac{\partial n}{\partial t} = -\mathrm{div}\mathbf{j} - \frac{\partial \varphi}{\partial t} = D\nabla\left[\left(1 - \frac{\varphi}{\varphi_m}\right)\nabla\left(\frac{n}{1-\varphi/\varphi_m}\right)\right] - \frac{\partial \varphi}{\partial t}, \quad 0 \le r < \Sigma(t), \tag{2}$$

where $D = \beta kT$ is the diffusivity of colloid particles and the last term on the right-hand side corresponds to the mass exchange between free and frame particles ($n \rightleftarrows \varphi$ transitions). These transitions depend naturally on the local probabilities for the particles to overcome the energy barriers and on densities n and φ:

$$\frac{\partial \varphi}{\partial t} = an\varphi - b\varphi(\varphi_m - \varphi), \quad a, b \approx \mathrm{const}, \quad 0 \le r < \Sigma(t). \tag{3}$$

The local kinetic coefficients a and b of attachment and detachment, respectively, are determined by the real physical and chemical properties of the colloid system. The structure of the last term in Eq. (3) (recombination) is particularly remarkable. The number of particles detached per unit time in a local representative volume must be proportional to the number of aggregate-frame particles in this volume; therefore, the intensity of the $\varphi \longrightarrow n$ transition must be proportional to the concentration φ. On the other hand, the probability of detachment must be a decreasing function of the number of bonds between φ particles and must depend on a fraction of the free space in the local neighborhood of every aggregated particle. In the region of close packing where $\varphi \approx \varphi_m$, the recombination process is impossible. Relationship (3) uses the simplest dependence in which the number of particles locally detached from the aggregate is proportional to the product $\varphi(\varphi_m - \varphi)$.

In the space outside an aggregate, the distribution of the solitary-particle concentration is described by the ordinary diffusion equation with the natural boundary condition of inexhaustibility of the system (for $r \longrightarrow \infty$):

$$\frac{\partial \sigma}{\partial t} = D\Delta\sigma, \quad r > \Sigma(t), \quad \sigma(\infty) = \sigma_\infty. \tag{4}$$

The following conditions must be satisfied at the aggregate boundary $r = \Sigma(t)$:

(i) chemical-potential continuity for solitary particles

$$\ln\left(\frac{n}{1-\varphi/\varphi_m}\right) = \ln\sigma, \quad r = \Sigma(t); \tag{5}$$

(ii) the balance of diffusion flows of particles

$$D\frac{\partial \sigma}{\partial r} - D\left(1 - \frac{\varphi}{\varphi_m}\right)\frac{\partial}{\partial r}\left(\frac{n}{1-\varphi/\varphi_m}\right) = (n + \varphi - \sigma)\frac{d\Sigma}{dt}, \quad r = \Sigma(t); \tag{6}$$

(iii) the equation of boundary motion

$$\varphi\frac{d\Sigma}{dt} = \nu[a\sigma\varphi - b\varphi(\varphi_m - \varphi)], \quad r = \Sigma(t), \tag{7}$$

where the boundary $\Sigma(t)$ is assumed to be displaced due to attaching (detaching) particles by a value equal to ν (measured in centimeter) on average. This value can be considered as being approximately equal to the particle radius. Conditions (6) and (7) indicate that the aggregate growth (an increase in the value Σ with time) is determined by the balance between the diffusion supply of particles to the boundary [Eq. (6)] and the kinetics of their attachment to outer particles of the aggregate [Eq. (7)].

Thus, the model under consideration reduces to a system of coupled nonlinear equations with an unknown moving boundary. The model involves three natural time scales. The evolution of the internal structure of the aggregate is determined by Eq. (3) and is characterized by the aggregation time $\tau_a = \frac{1}{a}$. The solutions of diffusion equations (2) and (4) relax at the time $\tau_D = \frac{\Sigma^2}{D}$ of establishing the steady-state density profile. Finally, the motion of the boundary is characterized by a growth time of $\tau_\Sigma = \frac{\Sigma}{d\Sigma/dt}$. The general theory of problems with a mobile boundary [15] indicates that the time relaxation of a solution is primarily determined by the growth time τ_Σ. For sufficiently large and slow-growing aggregates, the inequality $\tau_a, \tau_D \ll \tau_\Sigma$ is valid. In other words, it is possible to set $\frac{\partial \varphi}{\partial t} \approx 0$ and $\frac{\partial \sigma}{\partial t} \approx 0$ in Eqs. (3) and (4). This property means that combination and recombination are dynamically balanced inside an aggregate, and the distribution of the solitary-particle concentration outside the cluster can be considered in the quasistationary approximation.

For further analysis, it is convenient to introduce the function $\rho = \frac{n}{1-\varphi/\varphi_m}$ corresponding to the concentration of solitary n particles that is determined for a representative-volume fraction free of aggregate-frame particles. Furthermore, the boundary condition $\varphi(t, 0) = \varphi_m$ at the center of an aggregate ($r \longrightarrow 0$) is quite reasonable and natural. This condition implies that the aggregate begins to grow from a small randomly formed, quasi-dense nucleus with the concentration φ_m of ran-

dom close packing. Under the above assumptions, the model includes the following equations and conditions:

$$\frac{\partial}{\partial t}\left[\rho\left(1-\frac{\varphi}{\varphi_m}\right)\right] = D\nabla\left[\left(1-\frac{\varphi}{\varphi_m}\right)\nabla\rho\right]-\frac{\partial\varphi}{\partial t}, \quad 0 \le r < \Sigma(t); \tag{8}$$

$$\frac{\partial\varphi}{\partial t} = a\varphi\left(1-\frac{\varphi}{\varphi_m}\right)(\rho - K) \approx 0, \quad K = \frac{b}{a}\varphi_m \ll 1; \tag{9}$$

$$\begin{aligned}\varphi(t, r=0) &= \varphi_m, \quad 0 \le r < \Sigma(t),\\ \Delta\sigma &= 0, \quad r > \Sigma(t), \quad \sigma(\infty) = \sigma_\infty;\end{aligned} \tag{10}$$

$$\rho = \sigma, \quad r = \Sigma(t); \tag{11}$$

$$D\frac{\partial\sigma}{\partial r} - D\left(1-\frac{\varphi}{\varphi_m}\right)\frac{\partial\rho}{\partial r} = \varphi\left(1-\frac{\rho}{\varphi_m}\right)\frac{d\Sigma}{dt}, \quad r = \Sigma(t); \tag{12}$$

$$\frac{d\Sigma}{dt} = \nu a\left[\sigma - K\left(1-\frac{\varphi}{\varphi_m}\right)\right], \quad r = \Sigma(t). \tag{13}$$

It is easy to note that principal kinetic equation (9) has the homogeneous solution

$$\varphi(r) = \varphi_m = \text{const}, \tag{14}$$

for which diffusion equation (8) reduces to the identity. This means that, after a quasi-dense nucleus has been formed, the aggregate grows like a liquid drop in a supersaturated vapor. Inside the aggregate space, there are only bound particles and therefore the *n*-particle concentration vanishes. Consequently, the function ρ remains a constant found from boundary conditions (11)–(13):

$$\begin{aligned}\rho(t, r) = \rho(\Sigma) = \sigma(\Sigma) &= \frac{D\sigma_\infty}{D + a\nu\varphi_m\Sigma},\\ \frac{d\Sigma}{dt} &= \frac{a\nu D\sigma_\infty}{D + a\nu\varphi_m\Sigma}.\end{aligned} \tag{15}$$

The above assumption $\tau_a \ll \tau_\Sigma$ concerning the time scales is evidently satisfied:

$$\frac{\tau_a}{\tau_\Sigma} = \frac{d\Sigma/dt}{a\Sigma} = \frac{\nu D\sigma_\infty}{\Sigma(D + a\nu\varphi_m\Sigma)} \ll 1,$$

because the lowest value of the aggregate dimension Σ cannot be less than the particle radius ν and a weakly concentrated system is considered. The second condition $\tau_D \ll \tau_\Sigma$ is also satisfied. Homogeneous drop-shaped aggregates are formed during phase stratification of colloids as a result of violating thermodynamic stability [2].

Using the standard approaches of the theory of stability to small perturbations, we obtain

$$\begin{aligned}\varphi(t, r) &= \varphi_m - \varphi_1(r)\exp(\alpha t),\\ \alpha &= a\varphi_m[K - \rho(\Sigma)]\\ &= \frac{a\varphi_m[a\nu\varphi_m K\Sigma - D(\sigma_\infty - K)]}{D + a\nu\varphi_m\Sigma},\end{aligned} \tag{16}$$

from which it follows that there is a region of sizes $0 < \Sigma < R_0$ where solution (14), (15) is stable because the increment α of increasing perturbations is negative. For large sizes $\Sigma > R_0$, the homogeneous solution becomes unstable ($\alpha > 0$). A critical value R_0 is determined from the condition $\alpha = 0$ and is equal to

$$R_0 = \frac{D}{\nu aK}\frac{\sigma_\infty - K}{\varphi_m}. \tag{17}$$

Of course, this value is meaningful only for $\sigma_\infty > K$, because growth of the aggregate is possible only in a system whose concentration σ_∞ exceeds the concentration K for dynamic balance between combination and recombination. Correspondingly, the difference $\sigma_\infty - K$ plays the role of supersaturation of the medium.

From the structure of kinetic equation (9), it follows that there is one more solution,

$$\rho(r) = K = \sigma(\Sigma), \tag{18}$$

for which Eq. (8) also reduces to the identity. In this case, the particle concentration in the aggregate depends only on the spatial coordinate. The explicit form of this function is found from Eqs. (10)–(13):

$$\begin{aligned}\varphi(r) &= \sqrt{\frac{\varphi_m D(\sigma_\infty - K)}{a\nu K r}},\\ \frac{d\Sigma}{dt} &= \sqrt{\frac{Da\nu K}{\varphi_m}\frac{(\sigma_\infty - K)}{\Sigma}}.\end{aligned} \tag{19}$$

In this case, the condition of internal dynamic equilibrium is evidently satisfied:

$$\frac{\tau_a}{\tau_\Sigma} = \sqrt{\frac{D\nu K(\sigma_\infty - K)}{a\varphi_m\Sigma^3}} = K\frac{\nu}{\Sigma}\sqrt{\frac{R_0}{\Sigma}} \ll 1.$$

For $\Sigma = R_0$, the growth rates (15) and (19) coincide with each other in both modes. For $\Sigma < R_0$, solution (18), (19) is meaningless, because $\varphi(r) > \varphi_m$. For $\Sigma > R_0$, the cluster-growth rate $\frac{d\Sigma}{dt}$ (19) is higher than rate (15) (see Fig. 2). Therefore, homogeneous solution (14), (15) becomes unstable at the point $\Sigma = R_0$ and goes over to inhomogeneous solution (18), (19). The latter solution is characterized by a higher velocity of the boundary and is stable, which corresponds completely to the general theory of the stability of solutions of parabolic problems with a mobile boundary.

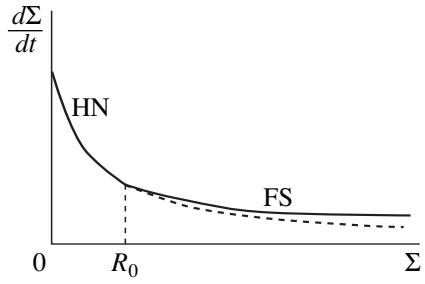


Fig. 2. The rate $\frac{d\Sigma}{dt}$ of aggregate growth vs. its size Σ. For $0 < \Sigma < R_0$, solution (15) corresponding to the homogeneous nucleus (HN) of the aggregate is stable. For $\Sigma > R_0$, solution (19) corresponding to the fractal structure (FS) is stable. The dashed line is the unstable HN solution.

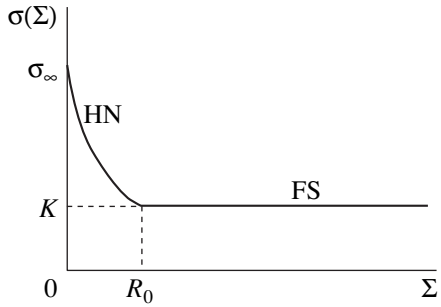


Fig. 3. The boundary concentration σ(Σ) of solitary particles vs. the size Σ of an aggregate.

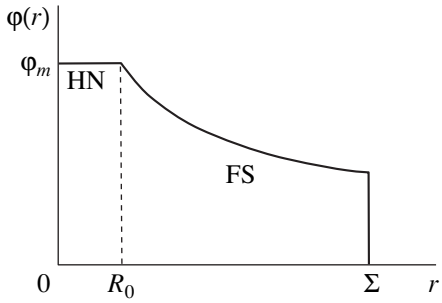


Fig. 4. The concentration φ of aggregated particles as a function of the distance r from the cluster center. For $0 < r < R_0$, there is a quasi-dense homogeneous nucleus (HN). For $R_0 < r < \Sigma$, the function $\varphi(r)$ is defined by Eq. (20) corresponding to the fractal structure (FS) of an aggregate.

It should be noted that the boundary value of the solitary-particle concentration σ(Σ) also continuously goes over from solution (15) to dependence (18) at the point $\Sigma = R_0$ (see Fig. 3).

An analysis of the internal structure of a cluster for $\Sigma > R_0$ leads to the conclusion that the center is occupied by a homogeneous dense nucleus with size R_0 (see Fig. 4). In the region around this nucleus, the aggregate becomes looser. In this region, the concentration of aggregated particles decreases according to a power law (!):

$$\varphi(r) = \varphi_m\left(\frac{R_0}{r}\right)^{1/2}, \quad R_0 \le r \le \Sigma. \tag{20}$$

In other words, the aggregate in this region is characterized by a fractal structure. As the aggregate grows, the total number of particles N in it also increases by a power law:

$$N(\Sigma) \sim \Sigma^{2.5}, \quad \Sigma > R_0. \tag{21}$$

Thus, the value $d_f = 2.5$ analytically determined for the fractal dimension of an aggregate is universal and independent of the particular physical and chemical conditions in a colloid system, which are specified by the parameters a, b, D, φ_m, σ_∞, and ν. This value is extremely close to the known experimental data: $d_f = 2.56 \pm 0.03$ [9] for aggregates of the immunoglobulin-type proteins IgG, 2.52 ± 0.05 [10] and 2.4 ± 0.1 [11] for aggregates of colloid quartz, 2.32 ± 0.05 [12] for particles of calcium carbonate, and 2.5 [13] for porphyrin aggregates. In addition, this result virtually coincides with the fractal dimension for three-dimensional DLA clusters: $d_f = 2.51 \pm 0.06$ [5, 6] and 2.495 ± 0.005 [7]. The last circumstance is nontrivial, because the DLA mode does not imply internal dynamic balance between combination and recombination.

It should be noted that the proposed quasiequilibrium self-similar solution is valid not only under the condition of dynamic equilibrium $\tau_a \ll \tau_\Sigma$, but also under the condition of slow growth $\tau_D \ll \tau_\Sigma$. Therefore, this solution can be considered as being asymptotically valid in the zero order with respect to the small value $\frac{\tau_D}{\tau_\Sigma} = \frac{\Sigma}{D}\frac{d\Sigma}{dt} \ll 1$ corresponding to the relative rate of cluster growth. In principle, it is possible to obtain the solutions in the next orders of perturbation theory by using asymptotic methods. This will likely lead to small deviations of the fractal dimension from $d_f = 2.5$.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 01-02-96430ural, 00-02-17731a, 01-02-16072a, and 01-01-00058a), by the Ministry of Education of Russian Federation (project no. E00-3.2-210), and by the U.S. Civilian Research and Development Foundation for the Independent States of the Former Soviet Union (grant no. REC-005).

REFERENCES

1. M. Von Smoluchowski, Z. Phys. Chem. Stoechiom. Verwandtschaftsl. **92**, 129 (1917).

2. W. B. Russel, D. A. Saville, and W. R. Schowalter, *Colloidal Dispersions* (Cambridge Univ., Cambridge, 1989).
3. B. V. Derjaguin, *Theory of Stability of Colloids and Thin Films* (Consultants Bureau, New York, 1989; Nauka, Moscow, 1986).
4. B. M. Smirnov, *Physics of Fractal Clusters* (Nauka, Moscow, 1993).
5. T. A. Witten and L. M. Sander, Phys. Rev. B **27**, 5686 (1983).
6. P. Meakin, Phys. Rev. A **27**, 1495 (1983).
7. S. Tolman and P. Meakin, Phys. Rev. A **40**, 428 (1989).
8. T. Vicsek, *Fractal Growth Phenomena* (World Sci., Singapore, 1992).
9. J. Feder, T. Jossang, and E. Rosenqvist, Phys. Rev. Lett. **53**, 1403 (1984).
10. S. K. Sinha, T. Freltoft, and J. Kjems, *Aggregation and Gelation* (North-Holland, Amsterdam, 1984).
11. J. Kjems and T. Freltoft, *Scaling Phenomena in Disordered Systems* (Plenum, New York, 1985).
12. C. Allain, M. Cloitre, and M. Wafra, Phys. Rev. Lett. **74**, 1478 (1995).
13. F. Mallamace, M. Micali, and S. Trusso, Phys. Rev. Lett. **76**, 4741 (1996).
14. R. I. Nigmatullin, *Foundations of the Mechanics of Heterogeneous Media* (Nauka, Moscow, 1978).
15. É. M. Kartashov, *Analytical Methods in the Theory of Solid State Thermal Conduction* (Vysshaya Shkola, Moscow, 1985).

Translated by V. Bukhanov

Doklady Physics, Vol. 47, No. 4, 2002, pp. 267–270. Translated from Doklady Akademii Nauk, Vol. 383, No. 4, 2002, pp. 478–481.

PHYSICS

Behavior of Critical Exponents in the Immediate Vicinity of a Critical Point for Nonideal Systems: The Second Crossover

D. Yu. Ivanov

Presented by Academician V.P. Skripov October 29, 2001

Received November 1, 2001

In the classic Ginzburg study dated 1960 [1], applicability limits of the Landau theory for second-order phase transitions were determined. In fact, it was shown that, on the basis of the ratio between the correlation energy and the volume energy, a temperature region near a critical point may be indicated where the role of fluctuations cannot be disregarded. In this region, classical theories of the Landau type (van der Waals theory for liquids, Weiss theory for magnets, and Bragg–Williams theory for binary alloys) no longer adequately describe the situation. At present, it is well known that, as the system being considered approaches a critical point, its classical (mean-field) behavior gives way to an Ising-type renormalization-group (RG) fluctuation behavior. The position of such a transition, if it exists, is specified by the Ginzburg criterion [1]. Thus, such a transition (it is often referred to as a crossover) from the classical type of behavior to the Ising-type behavior divides the region near the critical point into two parts. For the reasons substantiated in the analysis given below, we will call it the first crossover.

The history of studying critical phenomena and second-order phase transitions goes back about 200 years. During the last 40 years, the main efforts of both theorists and experimentalists were primarily aimed at comprehensively investigating singularities in the behavior of matter in the region of fully developed large-scale fluctuations near a critical point. As is well known, these efforts culminated in the formulation of the modern theory of critical phenomena by K. Wilson, who was awarded a Nobel prize for this in 1982. Despite these considerable achievements, there remain as-yet-unresolved problems first indicated by Ginzburg about 30 years ago (see, e.g., [2]). These problems are associated with the behavior of systems whose inhomogeneities are caused by the presence of walls, flows, external fields, etc. From the most general standpoint, we can state that, in this case, we are dealing with critical phenomena in nonideal systems or in systems affected by the action of various physical fields.[1]

The first question to be answered in this connection is whether the behavior of an real system changes (and if it does change, what is the character of these changes) as it moves deeply into the fluctuation region. The goal of the present study is to demonstrate that there is a positive answer to this question, but this answer is quite paradoxical.

It is well known that, as the system approaches a critical point, the development of large-scale fluctuations is accompanied by a continuous increase in the susceptibilities of the critical system, in particular, susceptibilities to various external perturbations. On this basis, we can assume that, eventually, the effect of various disturbing factors (fields), which are insignificant under the usual conditions, will lead, in the presence of continuously and indefinitely growing system susceptibility, to a deformation (suppression) of fluctuations. As a consequence of this fact, the mean-field (classical) behavior in the system [3] must be restored. We call this transition, which occurs in the inverse direction, i.e., from the Ising-type of behavior to the classical type, the second crossover.

In this paper, we present certain results obtained from our comparative analysis of available experimental and theoretical investigations that support the above concept [4]. As far back as the mid-1970s, we performed a precision $p\rho T$ experiment with pure SF_6 (99.9995% purity) in the immediate vicinity of the critical point. In the conditions of the same experiment, we found for the first time a trend of three static critical exponents, namely, β for the coexistence curve, γ for the isothermal compressibility in the single-phase region, and δ for the critical-isotherm, towards their classical values (see, e.g., [5–7]). In these papers, the changes found were for the first time attributed to gravity. The determination accuracy for the state parameters (including critical parameters) in those studies, per-

St. Petersburg State University of Refrigeration and Food Technologies,
ul. Lomonosova 9, St. Petersburg, 191002 Russia
E-mail: dmitri.ivanov@pobox.spbu.ru

[1] By a field, we imply here all possible additional disturbances, such as the gravitational and Coulomb fields, surface forces, shear stresses, turbulence, and the presence of boundaries.

1028-3358/02/4704-0267$22.00

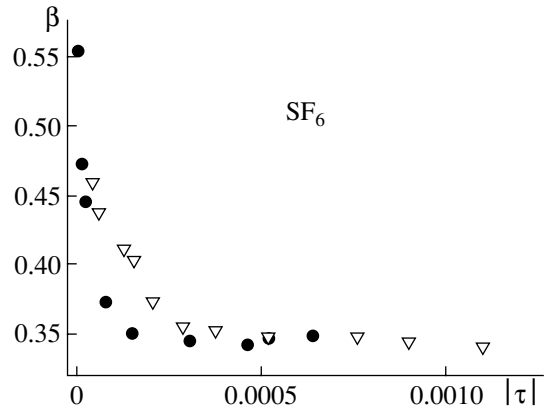


Fig. 1. Behavior of the critical exponent for the coexistence curve near a critical point ($\tau = \frac{T - T_c}{T_c}$): experimental points (●) are plotted according to data of [5]; the piezometer height is 8 mm, and the displayed data correspond only to the liquid branch; experimental points (▽) are plotted according to data of [8]; the piezometer height is 30 mm, and the displayed data correspond to both branches.

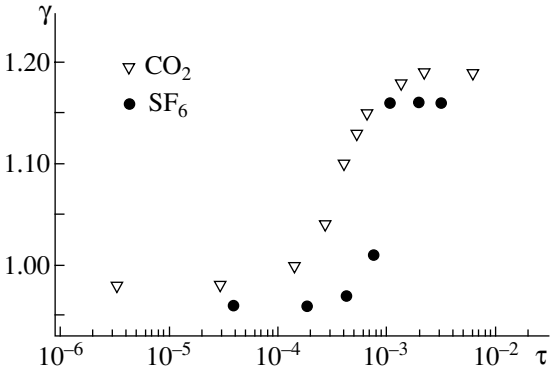


Fig. 2. Variation of critical exponents for the isothermal compressibility in the single-phase region, at the approach to a critical point. The experimental points (●) and (▽) are plotted according to data of [7, 15] and [9]; and piezometer heights are 8 and 30 mm, respectively.

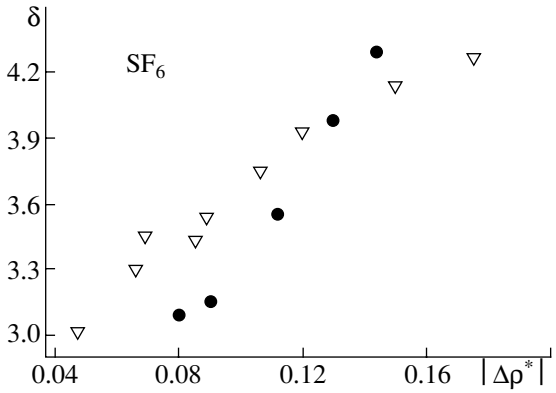


Fig. 3. Variation of the critical exponent for the critical isotherm in the single-phase region at the approach to the critical point in the density scale. The experimental points (●) and (▽) plotted according to data of [6] (the piezometer height is 8 mm; the displayed data correspond only to the liquid branch) and [8] (the piezometer height is 30 mm; the displayed data correspond to both branches).

formed at a unique setup in the laboratory headed by I.P. Krichevskiĭ, was at a metrological level. It was $\pm 2 \times 10^{-4}$ K in the proper temperature scale, ±0.001% in the pressure scale, and ±0.02% in the density scale. The critical parameters were determined independently by visually observing the appearance and disappearance of the two-phase state of matter in a piezometer of a constant–variable volume [5, 6]. About 800 experimental points obtained were concentrated in a narrow temperature range ($-0.3 < T - T_c < 1.3$ K, where T_c is the critical temperature) and density range ($|\Delta\rho^*| \le 0.15$, where $\Delta\rho^* = \frac{\rho - \rho_c}{\rho_c}$, with ρ_c being the critical density) near the critical point. As a result, it was established [5–7] that, within the ranges under investigation, there are own "far" and "near" regions. In the "far" region, the critical exponents β, γ, and δ had values close to the Ising ones, whereas in the "near" region, they again acquired values characteristic of the mean-field (classical) behavior (Figs. 1–3). For a long time, these papers were the only studies in the world scientific literature where such behavior was found and attributed to the effect of gravity. Now, the situation has substantially changed.

In study [8], which appeared in 1992, the immediate vicinity of the SF_6 critical point was investigated anew with the help of a completely automated, specially designed high-precision $p\rho T$ setup. In that study, our results and their interpretation [5–7] were at last independently corroborated. Later, in addition to SF_6 [5–9], the fact that, in the immediate vicinity of the critical point, the critical exponents change towards their classical values under the effect of gravity was also observed for CO_2 [9]. The investigations of the critical behavior in the presence of shear flows, which were performed in [10, 11], led to conceptually the same result as in our experiments with a pure liquid in the presence of gravity. Indeed, the critical exponents also changed from the Ising values to the mean-field ones. This was shown theoretically in the fundamental study [10] based on the RG approach and experimentally for an aniline–cyclohexane binary mixture in study [11], which was published simultaneously with [10]. In [12], in which the intrinsic gravitational effect had been studied theoretically, it was shown that, near a critical point, gravity changes the local properties of a liquid, thereby modifying the very nature of a phase transition. Using light scattering, it was shown in [13] that, in a guaiacol–glycerin solution with the addition of a small amount of water, the critical exponent ν of the correlation radius changes from nearly RG values to values characteristic of the mean-field behavior as the system approaches the double critical point. The Monte Carlo simulation performed in [14] revealed that, upon taking into account dimensional effects within the two-dimensional Ising model, the critical exponent β takes the RG value

$\beta = 1/8$ far from the critical point and the classical value $\beta = 1/2$ near the critical point.

In our opinion, the above results can be considered weighty arguments in favor of the basic idea of this study. The fact that the mean-field critical behavior in the immediate vicinity of a critical point (second crossover) is observed for a broad range of systems subjected to perturbations of various physical origins indicates that this phenomenon is quite universal. This behavior was discovered by the author and his coworkers in 1974 for the case of gravity.

We now dwell on the data displayed in Figs. 1–4, where all the dependences illustrating the behavior of various critical exponents in the second-crossover region were plotted on the basis of our data and data available from studies of other authors. From Figs. 1 and 2, we can see that the position of the second crossover for the critical exponents β and γ depends on the height of the piezometer used. This is an additional confirmation of the gravitational nature of these changes. By comparing the two figures in question, we also notice that the positions of this crossover for the exponents β and γ, in terms of the relative proximity to the critical point in the temperature scale, differ by approximately two orders of magnitude. The observed isothermal compressibility is more sensitive to the effects of various perturbations (in particular, to the effect of gravity). This seems to be quite natural from the physical standpoint. It is worth noting that, apparently, the positions of the first crossover can differ significantly for different thermodynamic properties of the same system, although the Ginzburg criterion based on the anomaly in the heat capacity does not suggest this *a priori* [1]. Figure 3 demonstrates that, in real systems, a transition to classical values of the critical exponents can occur as the systems approach a critical point not only in the temperature scale, but also in the density scale. In the author's opinion, the second crossover in the behavior of the critical exponents β, γ, and δ (Figs. 1–3) is due to the intrinsic gravitational effect. A specific reason for the variation of the exponent ν (Fig. 4) was not established in [13]. By analogy with the effect of the gravitational field (Figs. 1–3), we may assume that the Coulomb field of water dipoles played the role of a disturbing factor in that experiment. In any case, the evolution of the critical exponent of the correlation radius is such that it fits well into the proposed pattern. Thus, we have grounds to believe that such behavior is inherent not only in static critical exponents [7], but represents a universal property of the immediate vicinity for a critical point of real systems [3, 4].

Thus, we can conclude that, in the general case, the universal behavior of real nonidealized systems near a critical point is the following. In addition to the first crossover (its position is determined by the Ginzburg criterion) from the mean-field behavior to the Ising one, the universal behavior can also include the second

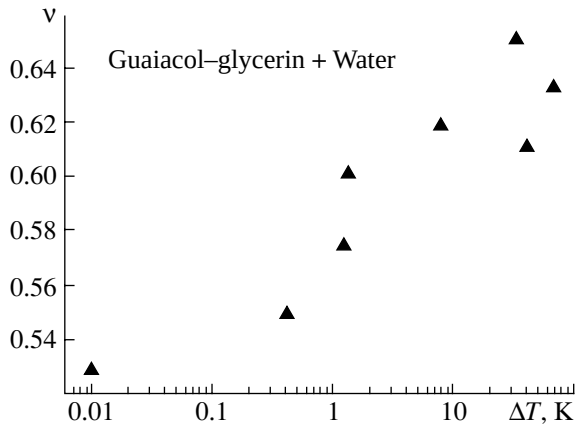


Fig. 4. Critical exponent for the correlation radius as a function of the distance from the double critical point (in the temperature scale) in the presence of water. The displayed data are plotted according to data of [13].

crossover in the immediate vicinity of the critical point. The second crossover proceeds in the inverse direction, namely, from the RG type of behavior to the classical type. Therefore, the immediate vicinity of a critical point again becomes the region of the mean-field type of behavior [3, 4].

In conclusion, the following comment is worth noting. It is usually assumed that the gravitational field has an adverse effect upon experiments in the immediate vicinity of a critical point, for example, by distorting the shapes of the curves and values of the critical exponents so that they cease to be true ones. However, all the aforesaid suggests that the existence of an additional factor in the form of an effect of the gravitational field or another arbitrary field is not, in fact, adversarial. On the contrary, this makes it possible to reveal new features in the critical behavior. It is only necessary to adopt a different point of view. Specifically, it should be assumed that the values of the critical exponents are not distorted by the effect of these fields but in fact are obtained in the presence of these fields, i.e., that they are valid just for these conditions. In the author's opinion, such an approach radically changes the situation, making it possible to obtain deeper insights into the nature of critical phenomena.

REFERENCES

1. V. L. Ginzburg, Fiz. Tverd. Tela (Leningrad) **2**, 2031 (1960) [Sov. Phys. Solid State **2**, 1824 (1960)].
2. V. L. Ginzburg, in *Physics in XX Century: Development and Progress* (Nauka, Moscow, 1984), pp. 281–330.
3. D. Yu. Ivanov, in *Proceedings of XIVth European Conference on Thermophysical Properties, Lyon, France, 1996*, p. 463.
4. D. Yu. Ivanov, Doctoral Dissertation in Mathematical Physics (St. Petersburg, 2001).

5. D. Yu. Ivanov, L. A. Makarevich, and O. N. Sokolova, Pis'ma Zh. Éksp. Teor. Fiz. **20**, 272 (1974) [JETP Lett. **20**, 121 (1974)].
6. D. Yu. Ivanov and L. A. Makarevich, Dokl. Akad. Nauk SSSR **220**, 1103 (1975).
7. D. Yu. Ivanov and V. K. Fedyanin, Preprint No. P4-8430, OIYaI (Joint Institute for Nuclear Research, 1974).
8. W. Wagner, N. Kurzeja, and B. Pieperbeck, Fluid Phase Equil. **79**, 151 (1992).
9. N. Kurzeja, Th. Tielkes, and W. Wagner, Int. J. Thermophys. **20**, 531 (1999).
10. A. Onuki and K. Kawasaki, Ann. Phys. **121**, 456 (1979).
11. D. Beysens, M. Gbadamassi, and L. Boyer, Phys. Rev. Lett. **43**, 1253 (1979).
12. J. V. Sengers and van J. M. J. Leeuwen, Int. J. Thermophys. **6**, 545 (1985).
13. S. V. Krivokhizha, O. A. Lugovaya, I. L. Fabelinskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **103**, 115 (1993) [JETP **76**, 62 (1993)].
14. A. Z. Panagiotopoulos, Int. J. Thermophys. **15**, 1057 (1994).
15. L. A. Makarevich, O. N. Sokolova, and A. M. Rozen, Zh. Éksp. Teor. Fiz. **67**, 615 (1974) [Sov. Phys. JETP **40**, 305 (1974)].

Translated by T. Syromyatnikova

Doklady Physics, Vol. 47, No. 4, 2002, pp. 271–274. Translated from Doklady Akademii Nauk, Vol. 383, No. 4, 2002, pp. 482–485.

PHYSICS

Composition Self-Modulation in Epitaxy of III–V Compounds and Multiplicity of Self-Organization Ways Due to Variations of Deviations from Equilibrium

K. S. Maksimov and S. K. Maksimov

Presented by Academician Yu.A. Osip'yan October 10, 2001

Received October 16, 2001

The nonequilibrium self-modulation in composition [1, 2] and/or the nonequilibrium atomic ordering [2] that arise in the process of epitaxy of SiGe and ternary III–V compounds are of interest not only for applications (e.g., in nanoelectronics), but also (as we hope to show) for understanding the general laws of self-organization. These semiconductor alloys are characterized by a positive mixing energy [2]. Atomic ordering arises because the mixing energy changes sign at the crystal surface due to surface reconstruction. This hypothesis is corroborated by correlations between the parameters of reconstruction and atomic ordering [2]. Self-modulation is explained by the fact that the mixing energy at the crystal surface remains positive [3–7]. The contradiction between the models of ordering and self-modulation is especially substantial when both structures coexist in the same bulk of a material [8–10].

To explain self-modulation, three models have been proposed. According to the first model, where this phenomenon is attributed to the minimization of elastic energy [4], the direction of modulation coincides with an in-plane direction $\langle 001 \rangle$ of the highest elastic compliance [4] (lateral phase separation [2]). The model is supported by numerous results [2]. In the second theory, which explains self-modulation by a specific fit of atoms of components into surface steps, the direction of modulation coincides with the growth direction and the periodicity is equal to the height of these steps [5]. This type of modulation (vertical phase separation [2]) was revealed for II–VI compounds [5] and for SiGe [6].

In numerous cases, modulation cannot be explained in the framework of the above models, since (1) its direction is out of the substrate plane [1, 2] and, coinciding with the growth direction, can differ from $\langle 001 \rangle$ [10, 11], and (2) the modulation period can attain 100 nm [1, 9, 11]. Self-modulation with anomalous parameters is explained only in the autocatalysis model in which it is caused exclusively by the energy of interatomic interactions (mixing energy) [3, 7, 13]. If the mixing energy is positive, a surface enriched with atoms of a certain species adsorbs mainly atoms of the same species (autocatalysis). In the diffusion autocatalysis model, modulation can appear under the conditions [7]

$$q_i = \alpha_i \pm \beta_i \eta \text{ and } |\alpha_i| < |\beta_i|.$$

Here, q_i is the adsorption coefficient of the ith component and $\eta = (C_1 - C_2)/(C_1 + C_2)$, where C_i is the concentration of the ith component at the growth surface.

Although this model explains self-modulation with anomalous parameters, it involves a number of controversial points and contradicts the models of self-organization epitaxial growth, which are considered as being approved [2].

First, the mechanism is developed as a kinetic one: it is associated with the transport of components, and the key moment in this transport is the transfer of the atoms of components from the mother phase onto the crystal surface; i.e., the model is realized in spatially inhomogeneous media, but it has not been substantiated energetically and the problem of producing entropy has not been investigated [14].

Second, the mixing energy is thought to give rise to and sustain the phenomenon. Stratification in composition arises when a minimum of the Gibbs potential corresponds to a homogeneous solid solution. However, the mixing energy is only one of the components of the Gibbs potential. Therefore, the statement that one of the terms of the Gibbs potential under certain conditions is responsible for the formation of structures that do not correspond to the minimum of the Gibbs potential should be substantiated.

Third, the model is applicable under the following conditions [13]: (1) desorption can be neglected, (2) adsorption-induced variations in composition of the surface layer of the mother phase are not compensated by diffusion, and (3) the coefficient β_i of the term responsible for adsorption depending on composition exceeds the coefficient α_i of the term characterizing

Moscow Institute of Electronic Technology (Technical University), Moscow, 103498 Russia

1028-3358/02/4704-0271$22.00

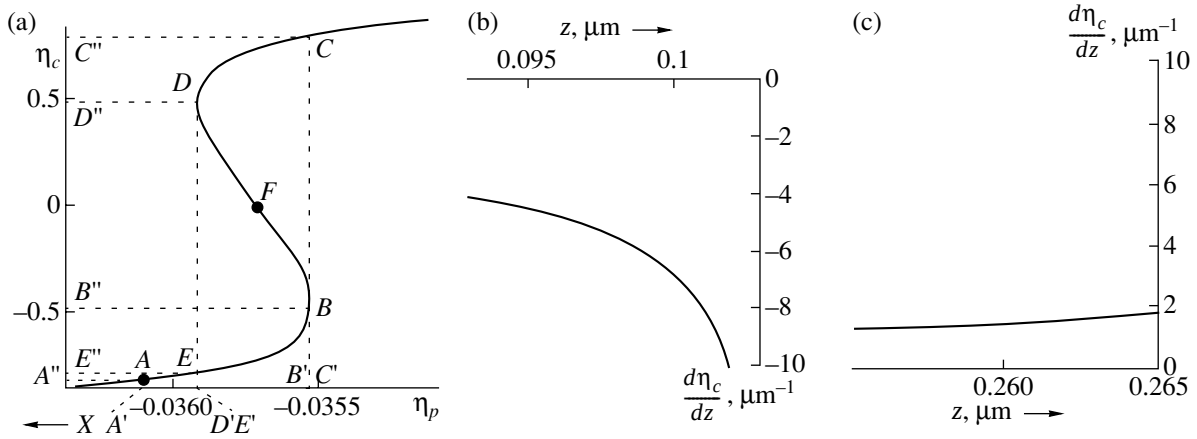


Fig. 1. Autocatalysis-model results for (a) the composition η_c of a crystal vs. the parent-phase composition η_p near the point *F*, (b) the composition gradient near the boundary when passing from a *Y*-enriched stratum to an *X*-enriched stratum, and (c) the composition gradient near the opposite boundary.

adsorption independent of composition. However, the realizability of these conditions in real epitaxial processes is not evident.

Fourth, in the first and second models, the phenomenon is attributed to factors that are independent of the mixing energy. Therefore, it is not inconceivable that these models can be made to agree with the reconstruction model of atomic ordering, but the autocatalysis model cannot be made to agree with the reconstruction model.

Fifth, according to data available in published papers, self-organization arises when at least three components participate in a reaction [15] and the coefficients of autocatalysis are nonlinear [3, 15]. However, in the model under discussion, self-organization arises when a process involves only two reactants and the adsorption coefficients depend linearly on surface composition. These contradictions are possibly illusory, since the model is realized in a spatially inhomogeneous medium where nonlinearity is caused by discontinuities of functions at the growth surface [8]. Nevertheless, this possibility should be confirmed.

Several efforts have been made to prove the validity of the autocatalysis model [13, 15]. Here, the following two conclusions are of interest [13]: (1) the modulation amplitude is independent of the composition, and the mean composition of an epitaxial layer is provided by variations in the thickness of the composition strata; and (2) variations in composition have a number of features explained below in Fig. 1.

Let us consider the features of self-modulation in a binary alloy *X*–*Y*. The crystal composition as a function of the parent-phase composition has a certain point *F* near which there is a region of several-valued (sigmoid) dependence [7, 13]. Let the parent-phase composition be set by a point *A* (e.g., the *X* component dominates). When an *X*-enriched stratum is formed, the composition of this phase changes from the composition specified by the point *A* to the composition specified by the point *B* and the effect of diffusion on the surface-layer composition is the converse of the effect of adsorption. For the same parent-phase composition and the formation of a *Y*-enriched stratum, the parent-phase composition varies from the point *C* to the point *D* and all the processes modify the composition of the surface layer in one direction [13]. Therefore, the interfaces between strata have different structures. In the growth direction, the interfaces between *X*- and *Y*- enriched strata are virtually stepwise. In the direction from *Y*-enriched strata to *X*-enriched strata, regions where the composition varies rapidly precede stepwise interfaces (Fig. 1). Observation of these interfaces proves that the autocatalysis model fits the process adequately.

If electron images are obtained from an object inclined in such a way that the electrons are sequentially scattered from volumes having different compositions (lattice parameters), intensity distribution in the images is determined by the interference of waves scattered on each of the interfaced volumes. Depending on the structure of the boundaries between strata, four types of images are possible [15]. When the composition changes only on boundaries and only stepwise, a superposition of two intensity oscillations, displacement fringes and moiré fringe, is observed. When continuous variation in composition precedes a stepwise boundary, only the moiré fringe appears. When the main portion of variations in composition are continuous, the boundary can be observed as a single light or dark fringe. When the strata are separated by a region where composition varies continuously, the boundary may be invisible.

In this study, the autocatalysis model is verified by using the boundary images lying on each side of the point *F*. We used foils with the (110) plane and self-modulated $GaAs_{1-x}P_x$/(001)GaAs samples that were obtained by chloride–hydride vapor-phase epitaxy

technology [1] where x varied continuously from 0.4 to 0.52. To obtain interference patterns, we applied [211] projections. Figure 2 shows an intensity profile corresponding to a microphotograph for a $02\bar{2}$ reflection where the growth direction is from left to right.

With a rise in the GaP content, the extent of the AB strata increases, which makes it possible to determine the dominant component in them and the types of boundaries. Boundaries between GaAs- and GaP-enriched strata are labeled by the letter A and opposite boundaries, by the letter B. In the left and central regions of the profile, an interference oscillation labeled by (▽) is observed on the A images. The oscillation has a moiré nature, since its period is independent of variations in diffraction conditions. The period is also independent of the parent-phase composition. Simultaneously with variations in the extent of the AB and BA strata, the intensity of the B images increases, whereas the interference oscillation on the A images weakens. On the right side of the photograph, the character of the images changes: the A** image has no interference oscillation; however, intensity oscillations labeled (◇) appear on the B* image. The oscillation on the B* image is characterized by a virtually constant amplitude, whereas the amplitude of the oscillation on an A* image decreases from right to left. The characters of weakening of the oscillations are different because the excitation errors change in different directions when passing through the boundaries. The localization of the point F within the section A*B* is corroborated by the extent of the section. The changes in boundary images when passing the point F can be explained only by variations in their structures, since these images are observed under the same diffraction conditions. They fully correspond to the autocatalysis model, which is therefore corroborated.

The self-modulation under investigation arises in GaAsP [1], GaInP [2, 11], InAsSb [8], InGaAs [2, 9], and GaAlAs [2, 11] in vapor- [1, 2] and liquid-phase [11] epitaxy and in the MOS hydride process [2, 8–10]. Therefore, it cannot be stimulated by the particular conditions of processes. Epitaxy is characterized by step-flow growth attributed to the motion of surface steps. However, the characteristic lifetime of a step for the growth rates in use is less than the oscillation times in the composition of the surface layer by several orders of magnitude. The only cause of modulation is the energy balance of interatomic interactions at the crystal–parent-phase interface. This balance is characterized by the mixing energy. The mixing energy can stimulate the appearance of structures that do not correspond to the minimum of the Gibbs potential.

There are only indirect reasons for the realizability of condition $|\alpha_i| < |\beta_i|$. The relative contribution of the mixing energy to the Gibbs potential increases with a decrease in temperature. We can assume that the difference $|\beta_i| - |\alpha_i|$ also increases with a decrease in temperature, i.e., with an increase in its nonequilibrium. In equilibrium, the rates of direct and inverse processes are identical. Therefore, the complete exclusion of desorption and the restricted role of diffusion mean that the model under discussion corresponds to substantial deviations from the conditions of equilibrium crystallization. In the model proposed in [5, 6], the segregation of identical atoms is equilibrium at the surface of a crystal, but it is not equilibrium in the bulk of the crystal. The model developed in [4] corresponds to small deviations from equilibrium. The validity of the models proposed in [2, 5, 6] and the model under discussion makes it possible to conclude that there are different ways of self-organization caused by variations in fluctuations from the equilibrium state.

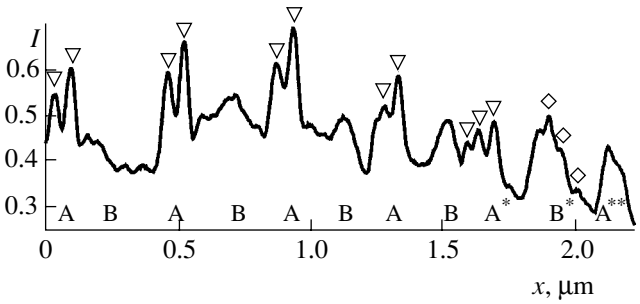


Fig. 2. Intensity profile on the image of a GaAsP self-modulated layer near the point F.

The validity of the autocatalysis model testifies to the inapplicability of the Hanusse theorem and to the possibility of self-organization when the absorption coefficients depend linearly on the surface composition in spatially inhomogeneous media. Self-organization caused exclusively by the positive mixing energy when only two components participate in a reaction and the dependence of the adsorption coefficients is linear was possibly of considerable importance in the appearance of living matter, since this self-organization promotes the segregation of compounds that were the elementary units of this matter.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 02-02-17521.

REFERENCES

1. S. K. Maksimov and E. N. Nagdaev, Dokl. Akad. Nauk SSSR **245**, 1369 (1979) [Sov. Phys. Dokl. **24**, 297 (1979)].
2. A. Zunger and S. Mahajan, *Handbook of Semiconductors* (Elsevier, New York, 1994), Vol. 3, pp. 1399–1513.
3. O. D. Chernavskaya, D. S. Chernavskiĭ, and R. A. Suris, in *Structural Defects and the Methods of Analysis, and Influence of the Properties of Crystals and Films*, Ed. by

S. K. Maksimov (Mosk. Inst. Élektronnoĭ Tekhniki, Moscow, 1982), pp. 3–16.

4. I. P. Ipatova, V. G. Malyshkin, A. A. Maradudin, *et al.*, Phys. Rev. B **57**, 12968 (1998).
5. A. Barabasi, Appl. Phys. Lett. **70**, 764 (1997).
6. P. Venezuela, J. Tersoff, Floro, *et al.*, Nature (London) **397**, 678 (1999).
7. S. K. Maksimov, K. S. Maksimov, and É. A. Il'ichev, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 412 (1996) [JETP Lett. **63**, 433 (1996)].
8. T. E. Seong, G. R. Booker, and A. G. Norman, Appl. Phys. Lett. **64**, 3593 (1994).
9. T. E. Seong, A. G. Norman, and G. R. Booker, J. Appl. Phys. **75**, 7852 (1994).
10. S. K. Maksimov, Kristallografiya **42**, 157 (1997) [Crystallogr. Rep. **42**, 145 (1997)].
11. S. K. Maksimov, L. A. Bondarenko, and A. S. Petrov, Fiz. Tverd. Tela **24**, 628 (1982) [Sov. Phys. Solid State **24**, 355 (1982)].
12. K. C. Hsieh, K. Y. Hsieh, J. L. Hwang, *et al.*, Appl. Phys. Lett. **68**, 1790 (1996).
13. S. K. Maksimov and K. S. Maksimov, Kristallografiya **46**, 942 (2001) [Crystallogr. Rep. **46**, 868 (2001)].
14. G. Nicolis and I. Prigogine, *Self-Organization in Non-equilibrium Systems* (Wiley, New York, 1977).
15. K. S. Maksimov and S. K. Maksimov, Izv. Ross. Akad. Nauk, Ser. Fiz. **65**, 1260 (2001).

Translated by Yu. Vishnyakov

Doklady Physics, Vol. 47, No. 4, 2002, pp. 275–280. Translated from Doklady Akademii Nauk, Vol. 383, No. 4, 2002, pp. 486–491.

PHYSICS

On Satellite-Based Detection of Ultrahigh-Energy Cosmic Rays by the Radio Method

V. A. Tsarev and V. A. Chechin

Presented by Academician E.L. Feĭnberg October 22, 2001

Received October 24, 2001

1. Within the last 35–40 years, a few tens of events were recorded by large-scale cosmic ray (CR) detectors, which testified to the presence in the CR flux of particles with macroscopic energies (higher than 5×10^{19} eV $\approx$ 8 J), so-called ultrahigh-energy cosmic rays (UHECR) [1]. The existence of such particles is inconsistent with the CR-spectrum cutoff caused by the Greisen–Zatsepin–Kuz'min (GZK) effect [2, 3]. This discrepancy has posed a number of fundamental questions concerning the sources of such particles and the mechanisms of their production and propagation [1–3]. To explain the presence of UHECR, there is a need either to change the traditional concepts of astrophysical sources of ultrahigh-energy particles or to assume the existence of supermassive particles whose decays result in the appearance of UHECR [4] (see also references in [1]). The most radical assumption is associated with a possible deviation from the standard special relativity theory at very high values of Lorentz factors [5]. In order to thoroughly investigate UHECR and verify various models proposed for their explanation, both the amount of statistical data and the measured energy range should be considerably increased. However, the UHECR flux is extremely low: it is at the level of one particle per km^2 per century for $E \sim 10^{20}$ eV. Thus, to obtain a statistically significant sampling of events within a reasonable time, dedicated large-scale detectors with a huge aperture are required. Currently, such detectors are planned or are even under construction. They are oriented to the detection of either charged (shower) particles or photons (fluorescent light) produced in extensive air showers (EAS) initiated by UHECR in the atmosphere. The size of the observation area is expected to be 6000 km^2 for the ground-based "Auger observatory" and about 10^5 km^2 for the detection of fluorescent light from a satellite.

In [6, 7], the possibility of detecting UHECR using satellite-based radio antennas was indicated. It is assumed that suitable receivers will detect radio emission from nearly horizontal EAS initiated by UHECR in the Earth's atmosphere. In this paper, we present comprehensive quantitative estimates concerning a radio signal of an EAS at large distances from it and compare this signal with the radio noise in the satellite's orbit. Our analysis shows the feasibility of satellite-based detection of UHECR by the radio method.

Radio emission from electromagnetic showers in a medium was predicted by Askaryan in 1961 [8], and Jelly in 1965 [9] was the first to experimentally observe radio pulses correlated with an EAS. Later on, experiments in this field were carried out in many laboratories and radio pulses were detected in a wide frequency range from a few megahertz to a few hundred megahertz. In particular, the possibility of dealing with only radio antennas (without coincidence with shower arrays or optical arrays intended for EAS detection) was demonstrated. Nevertheless, up to now, the radio method has not been as efficiently used for detecting EAS as the method of recording optical Cherenkov radiation. The main reason is evidently the following. An intense radio signal that can be reliably detected is generated only by the highest-energy EAS. However, the flux of such EAS is too low, and in order to detect them, we must observe a very large area. As is shown below, the problem of radio detection of EAS can be solved by observing them from space. In this case, the possibility arises of receiving EAS signals from more than one million km^2 of atmosphere surface area. Thus, detection of EAS with $E_{\mathrm{EAS}} \geq 10^{20}$ eV becomes statistically possible. As we show in the next sections, signals from such ultrahigh-energy EAS are sufficiently intense that they can be detected at large distances from satellites and reliably separated from the background.

2. Two mechanisms are mainly responsible for the coherent radio emission by EAS. One of them is related to the excess negative charge produced as a result of shower-particle interactions with atomic electrons:

$$\gamma + e^-_{at} \to \gamma + e^-, \quad e^+ + e^-_{at} \to e^+ + e^-,$$

$$e^+ + e^-_{at} \to \gamma + \gamma, \quad e^- + e^-_{at} \to e^- + e^-.$$

Lebedev Institute of Physics, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117924 Russia

1028-3358/02/4704-0275$22.00

The relative magnitude of the excess $\eta = \frac{N_- - N_+}{N_- + N_+}$ is estimated as $\eta = 0.1$–0.2 [8]. Here, N_+ and N_- are the numbers of positively and negatively charged particles in an EAS. The other source of radiation is related to the separation of positively and negatively charged particles in an EAS caused by the geomagnetic field. As a result of this separation, a transverse electric-dipole moment (and a transverse current) arises. The motion of this dipole (and current) through the atmosphere generates Cherenkov-type radiation even in the absence of the excess charge [10].

Initially, we consider radiation caused by excess charge and use the long-wave approximation and the usual relations of radiation theory for a system of moving charges. Then we write out the following expressions for the Fourier components of the magnetic $\mathbf{H}_\omega$ and electric $\mathbf{E}_\omega$ fields of a volume charge at large (i.e., much longer than the characteristic dimensions of an EAS) distances R at an angle θ to the EAS axis:

$$|\mathbf{H}_\omega| = |\mathbf{E}_\omega| = \frac{2\pi e\nu}{c^2}\frac{|\mathbf{L}_\omega(\theta)|}{R}.$$

Here, ν is the radiation frequency, c is the speed of light, e is the electron charge, and $\mathbf{L}_\omega(\theta)$ is the four-dimensional Fourier component of the particle-current density (of the electron excess):

$$\mathbf{L}_\omega(\theta) = \int dt\iiint d\mathbf{r}[\mathbf{n}\mathbf{j}_-(\mathbf{r}, t)]\exp(i\omega t - i\mathbf{k}\mathbf{r}).$$

In this formula, $\mathbf{k} = \frac{\mathbf{n}\omega}{c}$ is the radiation wave vector in the direction $\mathbf{n}$ and the integration is carried out over both the volume of the EAS and the entire time of its existence.

The current density has the following form:

$$\mathbf{j}_-(\mathbf{r}, t) = N_-(t)\mathbf{v}(\mathbf{r}, t)n(\mathbf{r}).$$

Here, $n(\mathbf{r})$ and $\mathbf{v}(\mathbf{r}, t)$ are the density (normalized to unity) of particles and their velocity in the EAS disk, respectively, and $N_-(t) \equiv N_e$ is the number of particles (corresponding to the electron excess) in the EAS.

For qualitative estimates, the function $\mathbf{v}(\mathbf{r}, t)$ can be considered to be constant:

$$|\mathbf{v}(\mathbf{r}, t)| \approx |\mathbf{v}| \approx c.$$

Under this assumption, the EAS disk moves as a whole with a constant velocity $\mathbf{v}$, i.e., $\mathbf{r} \approx \mathbf{r}' + \mathbf{v}t$. Here, $\mathbf{r}'$ is a radius vector with respect to the EAS-disk center. Then, $|\mathbf{L}_\omega(\theta)|$ acquires the form of a product of the total track length L_- for all particles of the electron excess in the EAS and two form factors F_t and F, both of which tend to unity in the limit of very long wavelengths:

$$|\mathbf{L}_\omega(\theta)| = L_-\sin\theta|F_t(\omega[1 - \beta\cos\theta])F(k)|,$$

$$L_- = v\int dtN_-(t),$$

$$L_-F_t(\omega[1 - \beta\cos\theta])$$

$$= v\int dtN_-(t)\exp(i\omega t[1 - \beta\cos\theta]),$$

$$F(k) = \iiint d\mathbf{r}n(\mathbf{r})\exp(-i\mathbf{k}\mathbf{r}).$$

In the long-wave approximation, the Fourier component of the radiation field is proportional to the frequency ν and to the total track length L_- of electron-excess particles in an EAS, i.e., $|H_\omega| \backsim \nu\sin\theta\frac{L_-}{R}$. The field shape for shorter wavelengths is determined by the form factors $F(k)$ and $F_t(\omega[1 - \beta\cos\theta])$. The spatial form factor $F(k)$ cuts the radiation spectrum at wavelengths on the order of the projection length $l_\perp$ of the EAS disk onto the radiation direction. For the time-dependent form factor $F_t(\omega[1 - \beta\cos\theta])$, the analogous parameter is on the order of

$$l_t \approx vt_{EAS}[1 - \beta\cos\theta],$$

where t_{EAS} is the EAS development time. It is worth noting that these standard formulas are related only to distances R that are much longer than the EAS development length vt_{EAS}, so that $vt_{EAS} \gg l_\perp$.

As a simplest model for the EAS space–time shape, we consider the Gaussian distribution

$$n(\mathbf{r}) = \frac{\exp\left(-\frac{\rho^2}{\sigma_\perp} - \frac{z^2}{\sigma_\parallel}\right)}{\sigma_\perp\sigma_\parallel^{1/2}\pi^{3/2}}, \quad (1)$$

$$N_-(t) = N_{max}\exp\left\{-\frac{v^2(t - t_{max})^2}{\sigma_t}\right\}. \quad (2)$$

Here, ρ and z are the transverse and longitudinal coordinates of the EAS disk. The EAS parameters $\sigma_\perp$, $\sigma_\parallel$, and σ_t only slightly depend on the EAS energy and are known reasonably well from experiments. Thus, we find

$$L_- = N_{max}(\pi\sigma_t)^{1/2}, \quad F \times |F_t| = \exp\left(-\frac{k^2\sigma_\theta}{4}\right),$$

$$\sigma_\theta = \sigma_\perp\sin^2\theta + \sigma_\parallel\cos^2\theta + \sigma_t[1 - \beta\cos\theta]^2.$$

To take into account the Cherenkov radiation, we should substitute β by βn_a, where $n_a = 1.0003$ is the

atmospheric refractive index. It is worth noting that the results obtained with the Gaussian parametrization are in good agreement with those obtained with the standard Gaisser–Hillas parametrization for longitudinal EAS development.

We now estimate the quantities that enter into the above formulas.

(i) For definiteness, we consider an EAS with an energy $E \sim 10^{20}$ eV and assume that $\eta = 0.2$. Then, the number of excess negatively charged particles in the EAS is $N_{max} \approx 2 \times 10^{10}$. It is worth noting that, in general, η depends on time and coordinates taken within the EAS disk and only the product $j_-(r, t) = \eta(r, t)j(r, t)$ enters into all the above formulas, where $j(r, t)$ is the current density for all particles in the EAS. Approximation expressions (1) and (2) corresponding to this product describe reasonably well the results of Monte Carlo calculations for the electron-excess distribution in the EAS. With allowance for this remark, we will speak, for brevity, about a distribution of all particles in the EAS, bearing in mind that only a certain constant part η of these particles contributes to the radiation field.

(ii) The parameter σ_t that determines the temporal development of an EAS slightly depends on the energy: $\sigma_t \approx 4\lambda_{rad} v t_{max}$, where λ_{rad} is the radiation length (in air and under normal conditions, $\lambda_{rad} \approx 0.35$ km). The distance to the maximum of EAS development $v t_{max} = l_{max} \approx 2.3\lambda_{rad} \ln\left(\frac{E}{10^8 \text{ eV}}\right)$ [10]. Then, in the context of this model, for $E \sim 10^{20}$ eV, $\sigma_t^{1/2} \approx 10\lambda_{rad} \approx 3.5$ km and $l_{max} \approx 30\lambda_{rad} \approx 10$ km. Correspondingly, $L_- \approx N_{-max}(\pi\sigma_t)^{1/2} \approx \eta\left(\frac{E}{10^9 \text{ eV}}\right) \times 18\lambda_{rad} \approx 1.3 \text{ km} \left(\frac{E}{10^9 \text{ eV}}\right)$ (in air).

(iii) The disk dimensions are reasonably well known from experiments and slightly depend on the age of an EAS. We assume that $\sigma_\perp^{1/2} = 40$ m and $\sigma_\parallel^{1/2} = 4$ m.

(iv) Typical electron energies in an EAS are on the order of 100 MeV, so that the Lorentz factor is $\gamma = (1 - \beta^2)^{-1/2} \approx 200$.

The frequency dependence for the magnitude of the electric-field Fourier component, which is calculated with the above-specified parameters, is shown in Fig. 1 by dashed lines. The results correspond to the energy $E = 10^{20}$ eV; to the observation angles $\theta = 0.03$, 0.1, and 0.5; and to the distance $R = 1000$ km. Owing to the standard relativistic factor $(1 - \beta\cos\theta)$ with $\beta \approx 1$, the space–time EAS "dimension" $\sigma_\theta^{1/2}$ sharply increases with the emission angle. This results in a rapid decrease in the maximum radiation frequency. Since emission angles are chosen to be rather small, such a radiation spectrum can be observed in the atmosphere from satellites only for EAS that are close to the horizon.

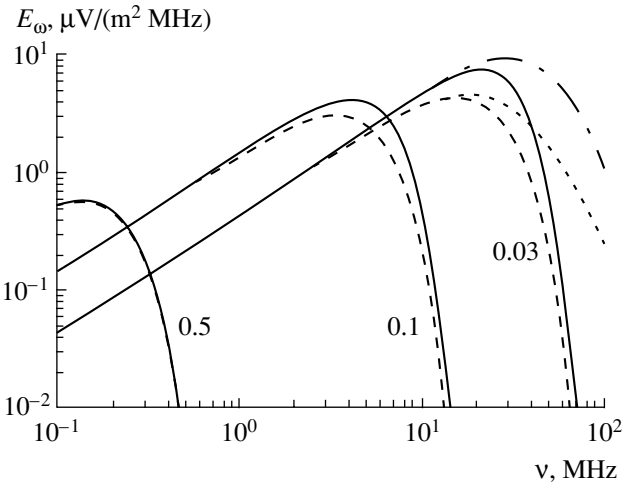


Fig. 1. Frequency dependence for the Fourier component of the electric-field intensity in different models. The distance from the EAS is 1000 km; the energy $E = 10^{20}$ eV; and the emission angles $\theta = 0.03$, 0.1, and 0.5. Solid and dashed lines correspond to the cases with and without the dipole-charge separation taken into account, respectively (for a shower disk with a Gaussian shape). Dashed-dotted and dotted lines correspond to the cases with and without dipole separation taken into account, respectively (for a shower disk with a central core).

Next, we take into account the radiation caused by the EAS dipole moment [10], i.e., by a shift of the negatively charged part of the EAS disk with respect to the positively charged one by a fixed distance d along one of the transverse axes (e.g., x-axis). Then, the term $\exp\left(-\frac{x^2}{\sigma_\perp}\right)$ should be substituted by

$$0.5\left(1 + \frac{1}{\eta}\right)\exp\left[-\frac{\left(x + \frac{d}{2}\right)^2}{\sigma_\perp}\right]$$

$$+ 0.5\left(1 - \frac{1}{\eta}\right)\exp\left[-\frac{\left(x - \frac{d}{2}\right)^2}{\sigma_\perp}\right].$$

In this case, the following extra multiplier appears in the form factor $F(\mathbf{k})$:

$$F_d(k) = 0.5\left(1 + \frac{1}{\eta}\right)\exp\left(-ik_x\frac{d}{2}\right)$$

$$+ 0.5\left(1 - \frac{1}{\eta}\right)\exp\left(+ik_x\frac{d}{2}\right).$$

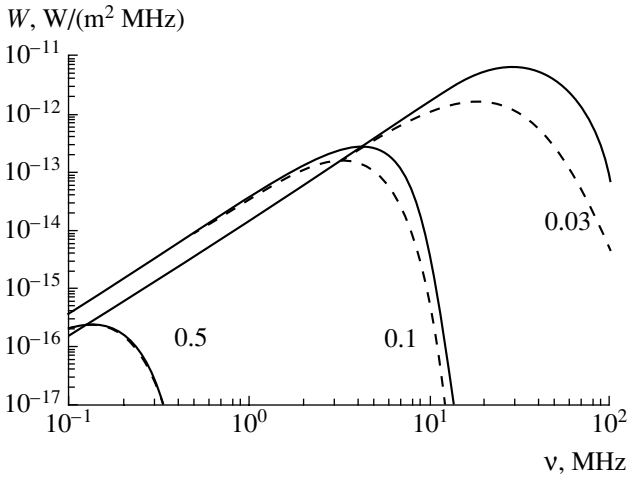


Fig. 2. Spectral density of the radio-emission power flux from an EAS with $E = 10^{20}$ eV at a distance of 1000 km for the same emission angles as in Fig. 1 with (solid line) and without (dashed line) the dipole separation taken into account.

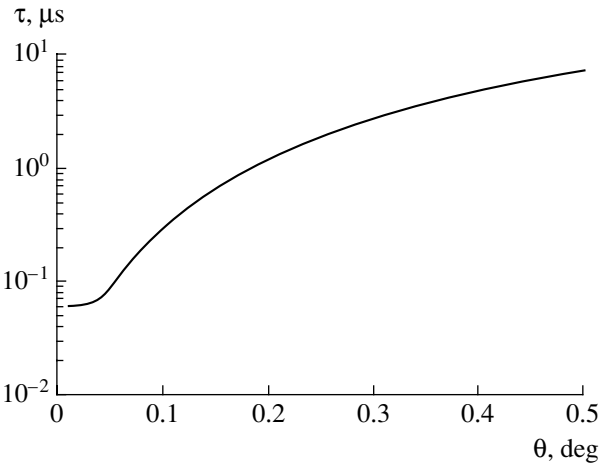


Fig. 3. Angular dependence of the radio-pulse duration for an EAS.

From experimental data, it follows that the charge shift can be as large as the EAS-disk dimension itself [10]. For illustration, the results of the calculations with this separation taken into account are shown in Fig.1 by solid lines. Here, it was assumed that $d = 50$ m, $\eta = 0.2$, and the separation takes place in the observation plane. (It is worth noting that in actual experiments, information on the separation direction can be obtained using measurements of the direction of radiation polarization.) From Fig. 1, it is seen that the contribution of the induced dipole moment is rather noticeable.

The smooth Gaussian shape of the charge distribution, which was chosen above, results in an exponentially rapid drop of the curves in the high-frequency region. At the same time, it is well known that there is a dense core near the EAS axis. Qualitatively, this core can be described (without introducing new parameters) with the help of substituting expression (1) by

$$n(\mathbf{r}) = \frac{\exp\left\{-\left(\frac{8\rho^2}{\sigma_\perp} + \frac{8z^2}{\sigma_\parallel}\right)^{1/2}\right\}}{\sigma_\perp \sigma_\parallel^{1/2} 8^{-1/2}\pi}. \quad (3)$$

Then,

$$F|F_t| = \frac{\exp\left(-\frac{k^2\sigma_t}{4}\right)}{\left(1 + \frac{k^2\sigma_s}{8}\right)^2},$$

where

$$\sigma_s = \sigma_\perp \sin^2\theta + \sigma_\parallel \cos^2\theta.$$

Results of the calculations for the same parameters as above but with relationship (3) taken instead of (1) are shown in Fig. 1 by dashed-dotted and dotted lines.

The duration of the radio pulse from an EAS is on the order of $\frac{1}{\nu_{max}}$. Therefore, the power density is

$$\frac{dW^2}{dSd\nu} \approx \nu_{max}\frac{d^2E}{dSd\nu} = \frac{c}{2\pi}\nu_{max}|\mathbf{E}_\omega|^2.$$

The power-flux spectral density for the same values of the parameters ($E = 10^{20}$ eV, $R = 1000$ km) is shown in Fig. 2. In the small-angle region ($\theta \le 0.1$), within the frequency interval $\nu \approx 30$–50 MHz, the power-flux density is $\frac{dW^2}{dSd\nu} \approx 10^{-11}$–$10^{-13}$ W/(m² MHz). Radio pulses of such magnitude can be easily detected by modern radio receivers. [For comparison, typical magnitudes of signals detected in radio astronomy are as low as 1 Jansky = 10^{-20} W/(m² MHz).]

It is seen from Figs. 1 and 2 that the radiation intensity is strongly dependent on the emission angle θ with respect to the EAS axis. To reconstruct the EAS energy from the measured value of E_ω or W, we need to know the radiation emission angle θ with respect to the EAS axis. This angle can be found by measuring the radio-pulse duration related to θ by the dependence shown in Fig. 3 (or by detection in the stereo mode by a couple of receivers installed aboard satellites). The distance between an EAS and the satellite can be found from the geometry of the horizontal-EAS detection and the orbit altitude with an accuracy of about 10%. It should also be emphasized that direct calibration of the radio-pulse detector installed aboard a satellite can be done using a proton beam from a high-energy proton accelerator,

which is sent into the atmosphere toward the satellite. Bunches of the accelerator beam will produce showers in the atmosphere, whereas their radio pulses will imitate radio pulses from EAS of ultrahigh energies. In this case, $E \approx NE_p$, where N is the number of protons in a bunch and E_p is the energy of an individual accelerated proton. This procedure is similar to that used in the recent SLAC experiment with radio waves generated by electron beams in a solid target [11].

We note that the frequency range $\nu \approx 30$–50 MHz is the most appropriate for EAS-pulse detection by a satellite receiver. This range is determined by the lower boundary of the atmosphere's transparency to radio waves. (This boundary is determined by the reflection of waves with $\lambda \geq 15$–20 m from the ionosphere.) On the other hand, a sharp decrease in the signal amplitude arises at frequencies $\nu \geq 100$ MHz, which is caused by the loss of coherence.

3. The extraction of the desired signal from the background is one of the key problems in EAS detection by the radio method. The background can be of both natural origin (radio emission of the atmosphere and ionosphere, the "radio-sky" radiation) and man-made origin.

Natural terrestrial radio pulses can be generated by both instabilities of the ionosphere and wave disturbances coming from the Earth's surface, which, as a rule, are associated with lightning discharges. It is important that all terrestrial natural types of radiation lie mainly in the low-frequency region (1 Hz–1 kHz) and are inessential within the range of interest $\nu \approx 30$–50 MHz. The Sun is the brightest source in the radio sky. Within the frequency range 30–50 MHz, the quiet Sun produces a background flux on the order of 10^{-16} W/(m^2 MHz), whereas the perturbed Sun can radiate as much as 10^{-12} W/(m^2 MHz). Thus, the expected signal from an ultrahigh-energy EAS will exceed the background from the quiet Sun by a few orders of magnitude. On the other hand, operation in periods of solar perturbations is hampered. The characteristic short duration of the EAS signal (10^{-7}–10^{-6} s) and signal-polarization measurements can be used for background suppression.

The highest background is produced by man-made sources, mainly due to operations of broadcasting stations. To estimate this background, we used the results of an analysis of the "global radio pollution" present in circumterrestrial space [12] and a compilation of radio-noise data measured with satellites [13]. From this comparison, we can conclude that the expected signal for $E \geq 10^{20}$ eV and small emission angles $\theta \leq 0.1$ (even without allowance for the antenna gain) exceeds the background level for radio-loud regions by two or three orders of magnitude. [For example, in the Southern Hemisphere above an ocean area, the background level is ≤10–14 W/(m^2 MHz).]

4. We now estimate the expected detection rate for useful events:

$$\frac{dN}{dt} = JS\Delta\Omega.$$

Here, J is the UHECR flux on the order of 10^{-2} events per km^2 per year for $E \geq 10^{20}$ eV, S is the effective area of the atmospheric surface from which (horizontal) UHECR can be detected at a given satellite position, and $\Delta\Omega$ is an available solid angle of UHECR incidence. We assume that the radio waves are focused onto the antenna by a parabolic reflecting surface [14]. In this case, the signal can be received for all azimuth directions from a circular atmospheric layer with a width $L \approx (2Rh)^{1/2}$ and a radius $R_1 \approx (2RH)^{1/2}$. Here, R is the Earth's radius, H is the satellite-orbit altitude, and h is the effective thickness of the atmosphere. Taking R = 6000 km, H = 500 km, and h = 10 km, we find

$$S = 5 \times 10^6 \text{ km}^2.$$

Assuming that the signal can be extracted from the background for $\theta \leq 0.1$, we find the following estimate for the event-detection rate: $\frac{dN}{dt} \approx 4$ events per day for $E \geq 10^{20}$ eV.

Thus, for an annual operation, more than 1500 UHECR events can be collected (which exceeds the total existing UHECR statistics by about two orders of magnitude) and the measured energy interval can be extended by one or two orders of magnitude.

5. We now summarize the basic conclusions:

Radio signals produced in the atmosphere by UHECR are sufficiently high for their reliable detection at distances on the order of 1000 km and can be detected by a satellite-based detector.

The background hampering this detection is lower than the expected signal in the case of the quiet Sun and for radio-quiet regions and does not prevent signal detection.

Simultaneous measurements of the signal amplitude and pulse duration make it possible to determine the EAS (i.e., UHECR) energy.

The observed area of the atmospheric surface when employing the radio method is expected to be as high as a few million square kilometers. Thus, the corresponding event detection rate could be as high as a few events per day at $E \geq 10^{20}$ eV.

The proposed method provides an opportunity for both considerably increasing the accuracy of UHECR-spectrum measurements and approaching higher energies.

ACKNOWLEDGMENTS

The authors are grateful to E. L. Feĭnberg, N. S. Kardashev, N. G. Polukhina, A. S. Pulinets, and R. L. Sorochenko for fruitful discussions and remarks.

REFERENCES

1. X. Bertou, M. Baratov, and A. Letessier-Selvon, Int. J. Mod. Phys. A **15**, 2182 (2000).
2. G. T. Zatsepin and V. A. Kuz'min, Pis'ma Zh. Éksp. Teor. Fiz. **4**, 114 (1966) [JETP Lett. **4**, 78 (1966)].
3. K. Greisen, Phys. Rev. Lett. **16**, 748 (1966).
4. V. A. Kuz'min and V. A. Rubakov, Yad. Fiz. **61**, 1122 (1998).
5. D. A. Kirzhnits and V. A. Chechin, Pis'ma Zh. Éksp. Teor. Fiz. **14**, 261 (1972) [JETP Lett. **14**, 182 (1972)].
6. K. M. Pichkhadze, V. G. Sysoev, V. A. Tsarev, and V. A. Chechin, Kratk. Soobshch. Fiz. **12**, 9 (2000).
7. V. A. Tsarev and V. A. Chechin, Kratk. Soobshch. Fiz. **4**, 42 (2001).
8. G. A. Askar'yan, Zh. Éksp. Teor. Fiz. **41**, 616 (1961) [Sov. Phys. JETP **14**, 441 (1961)].
9. J. V. Jelly, J. H. Fruin, N. A. Porter, *et al.*, Nature (London) **205**, 327 (1965).
10. H. R. Allan, Progr. Element. Particles Cosmic Ray Phys. **10**, 171 (1971).
11. D. Saltzberger, P. W. Gorham, D. Walz, *et al.*, arXiv:hep-ex/0011001.
12. A. G. Kolesnik and S. A. Kolesnik, Geomagn. Aeron. **36**, 59 (1995).
13. Data base http://www.atnf.csiro.au/SKA/intmit/database.html.
14. N. Z. Gogitidze, V. A. Tsarev, and V. A. Chechin, Nucl. Instrum. Meth. Phys. Res. A **248**, 186 (1986).

Translated by V. Tsarev

Doklady Physics, Vol. 47, No. 4, 2002, pp. 281–285. Translated from Doklady Akademii Nauk, Vol. 383, No. 5, 2002, pp. 617–621.

PHYSICS

Explosive Crystallization Initiated in Nanocrystalline Iron–Carbon Films by an Electron Beam

S. M. Zharkov and L. I. Kveglis

Presented by Academician K.S. Aleksandrov September 17, 2001

Received September 25, 2001

The problem of the thermal stability of a nanocrystalline state in the sense of crystallization upon uniform or local heating is interesting and important for theory and applications. Under the conditions of rather low heat transfer, self-acceleration of the crystallization process is possible. The latent heat of the transition is intensely released at the phase interface and heats a crystallization front, which takes the shape of a thermal domain moving with a velocity of up to several tens of meters per second. In this case, the liquid phase can arise at the crystallization front. This crystallization of a sample is usually called explosive crystallization [1, 2].

The character of the instabilities appearing at the phase interface upon crystallization depends on the process conditions. When the crystallization-front velocity is low and heat transfer is ideal, the crystallization front will be smooth. In the case of a steady-state growth of a crystalline needle in a supercooled melt, any bulge formed at the crystallization front should disappear, which ensures front smoothness. In the case of unsteady-state growth, an increase in the supercooling degree results in the appearance of instabilities at the smooth front of crystallization. These instabilities are characteristic of crystallization from a melt: small sinusoidal disturbances of the Mullins–Sekerka type, dendritic instabilities, cellular structure of the front, and fractal clusters described by the Witten–Sander model. The mechanisms of the formation of various instabilities upon crystal growth were considered in [3]. Explosive crystallization in Dy–Co, Pr–Ni, and Fe amorphous films exposed to an electron beam was reported in [4, 5]. As a result of crystallization, dendritic structures were formed.

In this paper, explosive crystallization initiated in nanocrystalline iron–carbon films by an electron beam from a transmission electron microscope is studied. The parameters of a macrostructure formed upon explosive crystallization are analyzed as functions of the velocity of propagating crystallization. Earlier, dendritic structures formed as a result of the vacuum thermal annealing of films obtained by the same technology were studied [6, 7]. The fractal oxidation of such films exposed to pulsed laser radiation in air was reported [8].

The microstructure and phase composition of the films were studied with a PREM-200 transmission electron microscope. Iron–carbon films 20–50-nm thick with ~20 at. % carbon content were prepared by pulsed plasma sputtering onto various substrates (NaCl, MgO, LiF) in 1.33×10^{-4} Pa vacuum [7]. The films were separated from the substrates in water or in a hydrofluoric acid solution and were placed onto the object-supporting grids for electron microscopy. Crystallization in the films was initiated by an electron beam from the transmission electron microscope under an accelerating voltage of 125 kV and a beam current of 50–75 μA.

In the initial state, the iron–carbon films had a nanocrystalline structure. Diffraction reflexes in the electron diffraction patterns obtained from such films are diffuse halos. Only one strongly broadened reflex was observed in the X-ray diffraction spectra. The dimension size of the particles composing the films in the initial state was determined from the broadening of the X-ray reflex and was 3.5 nm [9]. Explosive crystallization was observed in some nanocrystalline films exposed to an electron beam. It should be noted that this crystallization was observed under the conventional conditions of electron microscopy investigations. The crystallization velocity that was determined visually differed from one sample to another and was up to 1 cm/s. A typical electron microscopy image of a film after such crystallization is presented in Fig. 1. The crystallization process occurred as follows: first, a crystallization center appeared; then, branches with a dendritic structure emerged on different sides from this center. These branches, in turn, initiated the appearance of new crystallization centers. The branch-propagation velocity for the structure shown in Fig. 1 was ~0.25 cm/s. As a result of crystallization, a fraction of the film is covered by dendritic structures. The electron diffraction pattern obtained by the method of microdiffraction from one dendritic branch is given in the insert in Fig. 1. The electron diffraction pattern has a spotted

Kirensky Institute of Physics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia

1028-3358/02/4704-0281$22.00

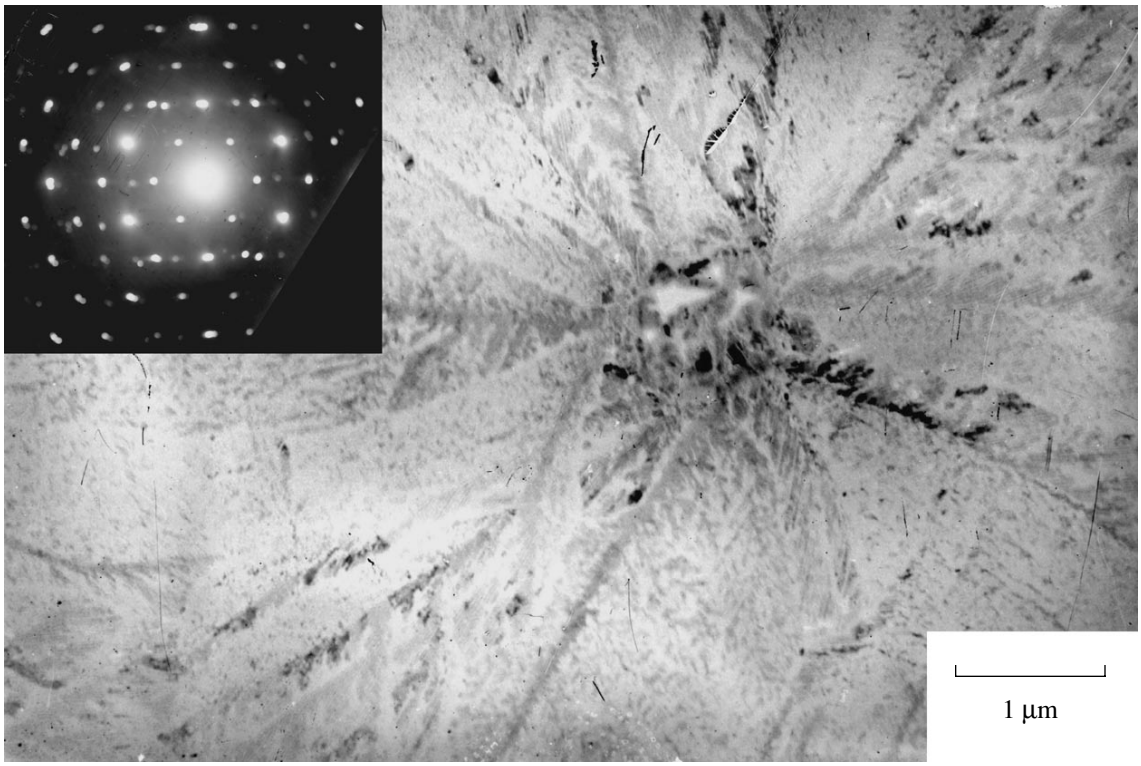


Fig. 1. Electron microscopy image of the dendritic crystal formed in the initial nanocrystalline iron–carbon film exposed to an electron beam. In the insert, the electron diffraction pattern obtained by the method of microdiffraction from a dendritic branch.

form and does not correspond to any known structure of either pure iron or iron–carbon compounds. The reflexes in the electron diffraction patterns obtained from the uncrystallized regions of the film are diffused and have halo shape.

When explosive crystallization was unsuccessful in the films in the initial state, they were annealed in vacuum at T_{ann} = 100–150°C for 30 min. The electron diffraction patterns obtained from the films after annealing did not differ from those for the initial state. Such films were again exposed to an electron beam and, as a rule, crystallization was observed. The electron microscopy image of the films after crystallization is presented in Fig. 2. The crystallization process occurred as follows: first, cells (primary instabilities) grew with a velocity of ~0.01 cm/s in the structureless film (Fig. 2, left part). The electron diffraction pattern obtained by the method of microdiffraction from one cell is also shown in Fig. 2. Then, secondary dendritic-type instabilities appeared and developed. As a result, a fraction of the film was covered by dendritic structures (Fig. 2, right part). The cellular structure was further decomposed due to the growth of secondary dendritic-type instabilities. A similar dendritic structure was observed in the Fe–Ni films upon crystallization from a melt [10].

The form of the dendritic (Fig. 1) and cellular-dendritic (Fig. 2) structures observed upon the crystallization of nanocrystalline iron–carbon films corresponds to the growth of crystals in a supercooled melt [3, 11]. As was mentioned above, a liquid zone moving in front of the crystallization front can arise upon explosive crystallization [2, 4]. The presence of the liquid zone allows the appearance of various instabilities, including dendritic and cellular instabilities, which are characteristic of crystallization from a melt. The existence of a liquid zone moving in front of the crystallization front implies that the crystallization process will be similar to crystallization in a supercooled melt. As was shown in [12], a paraboloid of revolution is the steady-state shape of a crystalline needle growing in a supercooled melt in the case of an isothermal surface. For a nonisothermal needle surface, the paraboloid of revolution is not a steady-state shape. In analyzing this case, Temkin [13] showed that the needle's velocity of motion will be maximum for a certain curvature radius $r = r_0$ of the needle tip. With an increase in supercooling $T_{k,\infty} - T_0$, where $T_{k,\infty}$ is the needle's surface temperature and T_0 is the melt temperature ($T_0 < T_{k,\infty}$), an increase in growth velocity and a decrease in needle thickness is observed. In the sense of shape stability, a needle that has a tip curvature radius r_0 and moves with the maximum possible velocity u_{max} at given supercooling ($T_{k,\infty} - T_0$) will be most stable. Indeed, if the bulge appears at the tip of such a needle, since its growth velocity is always less than u_{max}, it gradually disappears and the needle retains its initial shape. The most probable shape of the needle

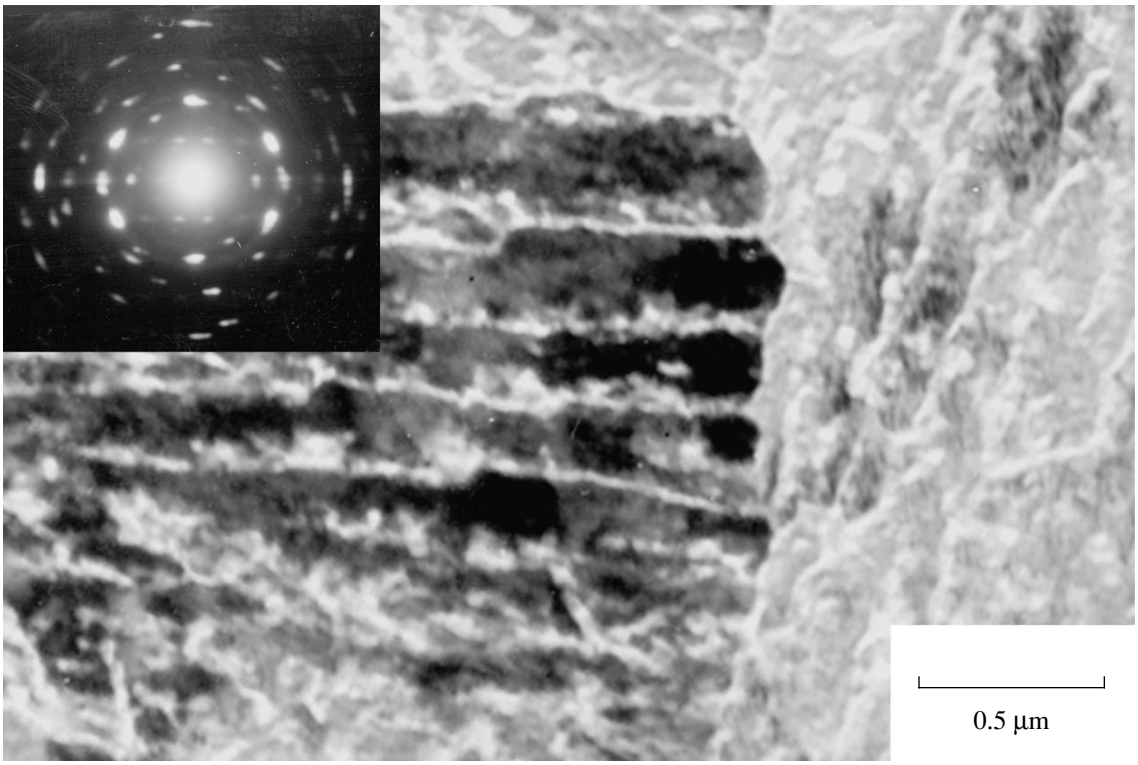


Fig. 2. Electron microscopy image of the cellular–dendritic structure formed in the iron–carbon film that was exposed to an electron beam and annealed (T_{ann} = 100°C) in vacuum. In the insert, the electron diffraction pattern obtained by the method of microdiffraction from a dendritic cell.

is expected to be a shape that ensures the maximum growth velocity [13].

Based on Ivantsov and Temkin's equations, Langer [3] calculated the growth velocity as a function of the tip curvature radius of a needle growing in a supercooled melt. This velocity is shown in Fig. 3. Instead of the term crystalline needle, we will use the terms dendritic branch and cell. From the experimental data obtained in this study of the explosive crystallization of nanocrystalline iron–carbon films, one can estimate the dimensionless velocity V of dendritic or cellular growth and the dimensionless curvature radius R of the dendritic branch or cell tip:

$$V = \frac{vd_0}{2D}, \quad R = \frac{r}{d_0},$$

where v is the growth velocity (cm/s) of the dendritic branch or cell, d_0 is the capillary length (~10^{-8} cm), D is the diffusion coefficient (~10^{-5} cm^2 s^{-1}), and r is the curvature radius (cm) of the dendritic branch or cell tip. The lines shown in Fig. 3 are plotted for the dimensionless supercooling

$$\Delta = (T_{melt} - T)\frac{C_p}{L} = 0.05,$$

where T_{melt} is the melting temperature, T is the temperature of the supercooled melt, C_p is the specific heat, and L is the latent heat. By calculating V and R corresponding to our experiment and indicating the values in Fig. 3, one can verify if the equations describing the

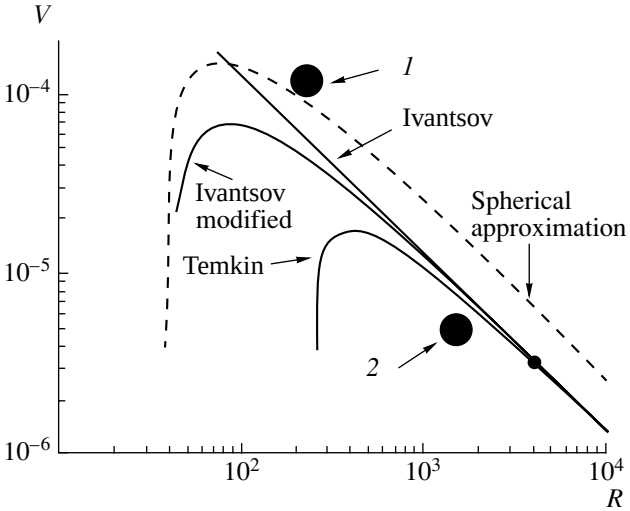


Fig. 3. Dimensionless crystal-growth velocity $\left(V = \frac{vd_0}{2D}\right)$ in a supercooled melt vs. the tip curvature radius of a crystalline needle $\left(R = \frac{r}{d_0}\right)$ [3]. Circles *1* and *2* correspond to the structures that were studied in iron–carbon films and are shown in Figs. 1 and 2, respectively.

crystallization processes in the supercooled melt are adequate for a description of crystallization processes in the films studied.

The dendritic structure shown in Fig. 1 is characterized by a growth velocity of ~0.25 cm/s of the dendritic branches and a curvature radius of 15–25 nm for the dendritic branch tips. The cellular structure shown in Fig. 2 is characterized by a growth velocity of ~0.01 cm/s and curvature radius of 100–160 nm for the cell tips. For the dendritic structure shown in Fig.1, we have $r = 2.5 \times 10^{-6}$ cm, $R = 250$, and $V = 1.25 \times 10^{-4}$ for the maximum curvature radius of the dendritic branch tip (point *1* in Fig. 3). For the cellular structure shown in Fig. 2, we have $r = 1.6 \times 10^{-4}$ cm, $R = 1600$, and $V = 5 \times 10^{-6}$ for the maximum curvature radius of the cell tip (point *2* in Fig. 3).

It is seen that the experimental values reported in this paper (Fig. 3) agree qualitatively both with the Ivantsov curve and with the spherical approximation. Thus, the relation between the growth velocity and curvature radius of the dendritic branch (or cell) tips upon the explosive crystallization of the films under study can be qualitatively described by the relations for crystalline-needle growth in a supercooled melt. In fact, the Ivantsov equation hardly corresponds exactly to the case in hand, because it implies that the crystalline-needle surface is isothermal, which is unlikely to be valid in this case. More likely, the Temkin dependence or the spherical approximation is more adequate for the description of the observed processes. However, in order to plot the dependences really corresponding to our experiment, one has to know the exact values of the quantities entering into the equations (capillary length, diffusion coefficient, supecooling degree, specific heat, latent heat, etc.). The exact determination of the values listed above was not the aim of this study and is a separate difficult problem.

An estimation indicated that the local temperature of the films, which is provided by the electron beam initiating explosive crystallization in nanocrystalline iron–carbon films, was no higher than 200–250°C. From the fact that similar structures were formed upon annealing the films in vacuum at $T_{ann} = 300$°C [6, 7, 9] and under mechanical action, it follows that the energy required for initiating explosive crystallization in the films under study is much less than the energy necessary for film melting. On the other hand, qualitative agreement of the experimental results with those characteristic of crystal growth in a supercooled melt and the shape of structures formed as a result of crystallization indicate that these structures can appear only upon crystallization in a supercooled melt. Consequently, the presence of a liquid zone at the crystallization front is an obligatory condition for the appearance of dendritic and cellular instabilities. Since the liquid zone could not have appeared as a result of an external action, one can assume that it appeared due to the release of energy stored in the initial state of the film. As a minimum, this energy should be sufficient for melting (or quasi-melting) of nanocrystalline particles composing the film. The mechanism of explosive crystallization in the films under study is assumed to be as follows. An electron beam initiates the appearance of the crystallization centers. A liquid zone appears at the crystallization front, and afterwards, self-maintained crystallization propagates along the film. Crystallization-front motion is self-maintained because crystallization at the interface between liquid and solid phase is accompanied by the release of the energy stored in the film. In the process of explosive crystallization, particles melting at the edges are built onto one another and form dendritic and cellular structures. Since these structures are not translationally symmetric, they are scaling-invariant and provide spotted electron diffraction patterns. Only this crystallization model can explain how dendritic (or cellular) structures responsible for spotted electron diffraction patterns similar to those from single crystals are formed from a structurally disordered nanocrystalline state in such a short time interval.

It is shown in [14] that the films are characterized by high oriented and unoriented stresses, which sometimes exceed the ultimate strength of a material in the bulk state by an order of magnitude. The ultimate strength of metal films is usually determined by the presence of numerous defects in the structure and can be several times higher than the ultimate strength of the corresponding bulk materials. It is known [15] that the energetically metastable regions in amorphous and nanocrystalline samples differ in energy from the normal (thermodynamically equilibrium) state by "stress energy" with characteristic values of 5–20 kJ/mol. This energy is sufficient to cause the appearance of the liquid zone in the process of structural rearrangement upon explosive crystallization. It can be expected that approximately the same energy is accumulated in the nanocrystalline iron–carbon films studied in this paper.

It is known that the crystal growth velocity in a supercooled melt depends directly on the degree of supecooling [13]. In turn, the degree of supercooling in this case will be defined by the energy accumulated in the film. In films where high energy was accumulated due to the technology of their preparation, explosive crystallization is obviously accompanied by the formation of dendritic structures, propagates with higher velocity, and requires less energy for initiation. In films accumulating less energy, larger energy is required to initiate crystallization that propagates with lower velocity and is accompanied by the formation of cellular structures. In relatively equilibrium samples, explosive crystallization is not initiated by an electron beam.

In summary, this study demonstrated that an electron beam induces explosive crystallization that propagates with a velocity up to 1 cm/s in nanocrystalline iron–carbon films and is accompanied by the formation of dendritic or cellular–dendritic structures. The electron diffraction patterns obtained after explosive crys-

tallization have a spotted form and are not consistent with any known structure of either pure iron or iron–carbon compounds. The relation between the growth velocity of dendritic branches (or cells) upon explosive crystallization and the curvature radius of dendritic branch (or cell) tips can be qualitatively described by the equations for crystal growth in a supercooled melt. The following dependence characteristic of crystallization in a supercooled melt is observed: the higher the growth velocity, the less the curvature radius of dendritic branch (or cell) tips. It is shown that explosive crystallization in the films under study proceeds with the formation of a liquid zone at the crystallization front. In this case, the liquid zone is formed not due to an external action but to the energy accumulated in the nanocrystalline state of the film.

ACKNOWLEDGMENTS

We are grateful to V.S. Zhigalov, who kindly placed the samples at our disposal and helped us to conduct the experiment, as well as to V.G. Myagkov for discussion of the results. This work was supported by the 6th Competition (1999) for Scientific Projects of Young Scientists of RAS (project no. 56), the Russian Foundation for Basic Research (project no. 00-02-17358a), and INTAS (grant no. 00-100).

REFERENCES

1. T. Takamory, R. Messier, and R. Roy, Appl. Phys. Lett. **20**, 201 (1972).
2. V. A. Shklovskiĭ and V. M. Kuz'menko, Usp. Fiz. Nauk **157**, 311 (1989) [Sov. Phys. Usp. **32**, 163 (1989)].
3. J. S. Langer, Rev. Mod. Phys. **52**, 1 (1980).
4. V. G. Myagkov, L. I. Kveglis, and G. I. Frolov, Poverkhnost, No. 9, 131 (1992).
5. V. G. Myagkov, L. I. Kveglis, V. S. Zhigalov, and G. I. Frolov, Poverkhnost, No. 1, 105 (1994).
6. S. M. Zharkov, V. S. Zhigalov, L. I. Kveglis, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 872 (1997) [JETP Lett. **65**, 915 (1997)].
7. G. I. Frolov, V. S. Zhigalov, L. I. Kveglis, S. M. Zharkov, *et al.*, Fiz. Met. Metalloved. **88** (2), 85 (1999) [Phys. Met. Metallorg. **88**, 183 (1999)].
8. V. G. Myagkov, V. S. Zhigalov, and S. M. Zharkov, Dokl. Akad. Nauk **346**, 612 (1996) [Phys. Dokl. **41**, 55 (1996)].
9. S. M. Zharkov, Candidate's Dissertation in Mathematical Physics (Krasnoyarsk, 1999).
10. W. J. Boettinger, S. R. Coriell, A. L. Greer, *et al.*, Acta Mater. **48**, 43 (2000).
11. W. Kurz and D. J. Fisher, *Fundamentals of Solidification* (Switzerland. Trans. Tech. Publ., 1986).
12. G. P. Ivantsov, Dokl. Akad. Nauk SSSR **58**, 567 (1947).
13. D. E. Temkin, Dokl. Akad. Nauk SSSR **132**, 1307 (1960).
14. V. A. Buravikhin, *Effect of Mechanical Stresses on the Magnetic Properties of Films* (Vost.-Sib. Knizhn. Izd., Irkutsk, 1968).
15. S. F. Timashev and L. I. Trakhtenberg, Zh. Fiz. Khim. **67**, 448 (1993).

Translated by T. Galkina

Doklady Physics, Vol. 47, No. 4, 2002, pp. 286–288. Translated from Doklady Akademii Nauk, Vol. 383, No. 5, 2002, pp. 622–624.

PHYSICS

Relativistic Quantum Rotator in a Periodic Field

D. U. Matrasulov, G. M. Milibaeva, and Corresponding Member of RAS P. K. Khabibullaev

Received September 12, 2001

The problem of quantum chaos in relativistic dynamical systems is investigated by solving the Dirac equation for a periodically disturbed rotator, which is the simplest model for studying quantum chaos in dynamical systems. To analyze rotator evolution in classical mechanics, one usually applies the equations in moment coordinates or in the action–angle phase space. These equations provide the mappings of equations that coincide with so-called standard mappings [1].

The evolution of a disturbed quantum rotator can also be described by (quantum) mappings that are obtained from the wave equation and describe the evolution of a wave function. In this paper, we analyze the quantum dynamics of a periodically disturbed rotator of spin 1/2. Although quantum chaos dynamics has been successfully developed over the last three decades, quantum chaos has primarily been studied for nonrelativistic systems. Nevertheless, many actual physical systems, such as a relativistic atom in a monochromatic field, the so-called quark–antiquark system (quasiatomic meson) that are disturbed by a periodic field, and some other systems, can be simulated by a relativistic quantum rotator that is periodically disturbed.

To solve the problem of the relativistic disturbed rotator, we should know the eigenvalues and eigenfunctions of the free rotator.

Many rotating systems, such as molecules, deformed nuclei, and some cosmological objects, can be simulated by the free rotator, whose properties are well studied in both classical and quantum cases and are available in any textbook of classical and quantum mechanics (e.g., [4–6, 8]). For this reason, the rotator is a convenient model for studying dynamical chaos in periodically disturbed systems [1, 9–11].

First, we solve the following Dirac equation for the free rotator:

$$(\boldsymbol{\alpha}\mathbf{p} + \beta)\psi = E\psi, \tag{1}$$

where $\alpha_\theta = \begin{pmatrix} 0 & \boldsymbol{\sigma}_\theta \\ \boldsymbol{\sigma}_\theta & 0 \end{pmatrix}$ and $\beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$ are the ordinary Dirac matrices, $p = i\dfrac{\partial}{\partial\theta}$ (hereafter, we set $m = \hbar = c = 1$), and $\psi = \begin{pmatrix} \varphi \\ \chi \end{pmatrix}$. The components φ and χ satisfy the equations

$$\begin{aligned} \boldsymbol{\sigma}\mathbf{p}\chi + \varphi &= E\varphi, \\ \boldsymbol{\sigma}\mathbf{p}\varphi - \chi &= E\chi. \end{aligned} \tag{2}$$

Expressing χ in terms of φ, we obtain

$$\chi = \frac{\boldsymbol{\sigma}\mathbf{p}}{1+E}\varphi. \tag{3}$$

The substitution of Eq. (3) into the first of Eqs. (2) gives

$$\frac{(\boldsymbol{\sigma}\mathbf{p})^2}{1+E}\varphi + \varphi = E\varphi \tag{4}$$

or

$$(\boldsymbol{\sigma}\mathbf{p})^2\varphi = (E^2 - 1)\varphi. \tag{5}$$

Using the relationship [5, 8]

$$(\boldsymbol{\sigma}\mathbf{p})^2 = p^2 + i(\boldsymbol{\sigma}[\mathbf{pp}]) = p^2, \tag{6}$$

we arrive at the equation

$$p^2\varphi = (E^2 - 1)\varphi \tag{7}$$

or

$$-\frac{\partial^2\varphi}{\partial\theta^2} = (E^2 - 1)\varphi = \varepsilon\varphi, \tag{8}$$

where

$$\varepsilon = E^2 - 1. \tag{9}$$

This equation coincides formally with the Schrödinger equation for the free rotator [5–8]

$$\varphi'' = -\varepsilon\varphi \tag{10}$$

and has the solution

$$\varphi_n = \frac{1}{\sqrt{2\pi}}e^{in\theta}. \tag{11}$$

Heat Physics Department, Academy of Sciences of Uzbekistan,
ul. Katartal 28, Tashkent, 700135 Uzbekistan

1028-3358/02/4704-0286$22.00

Therefore, the energy eigenvalues of the relativistic quantum rotator have the form

$$E_n = \sqrt{1+n^2}. \tag{12}$$

Taking into account that

$$\varphi = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}, \tag{13}$$

we obtain the small component in the form

$$\chi = \frac{\boldsymbol{\sigma}\mathbf{p}}{1+E}\varphi = \frac{1}{1+E}\begin{pmatrix} 0 & e^{-i\theta}\dfrac{\partial}{\partial\theta} \\ -e^{-i\theta}\dfrac{\partial}{\partial\theta} & 0 \end{pmatrix}\begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = \frac{in}{\sqrt{2\pi}(1+E)}\begin{pmatrix} e^{i(n-1)\theta} \\ -e^{i(n+1)\theta} \end{pmatrix}. \tag{14}$$

Thus, the wave functions of a free relativistic quantum rotator of spin 1/2 have the form

$$\psi_n = N\begin{pmatrix} \varphi \\ \chi \end{pmatrix} = N f_n e^{in\theta} = \frac{N}{\sqrt{2\pi}}\begin{pmatrix} 1 \\ 1 \\ \dfrac{in}{1+E_n}e^{-i\theta} \\ -\dfrac{in}{1+E_n}e^{i\theta} \end{pmatrix} e^{in\theta}. \tag{15}$$

As is seen from Eqs. (13) and (15), the energy and the wave function in the nonrelativistic limit ($n \ll 1$) coincide with the respective nonrelativistic expressions.

From the normalization condition

$$\int_0^{2\pi} \psi_m^* \psi_n d\theta = \delta_{mn}, \tag{16}$$

we obtain the normalization constant

$$N = \frac{1}{\sqrt{4\pi\left(1+\dfrac{n^2}{(1+E_n)^2}\right)}}. \tag{17}$$

Thus, we solve the Dirac equation for the free rotator. Then, we use the above results to solve the problem of the relativistic quantum disturbed rotator, for which the Dirac equation has the form

$$i\frac{\partial\Phi}{\partial t} = \hat{H}\Phi, \tag{18}$$

where

$$\hat{H} = \alpha_\theta p_\theta + \beta + \delta\left(\frac{t}{T}\right)\cos\theta. \tag{19}$$

We solve this equation by the method used in [13]. For this aim, we decompose the wave functions $\Phi(\theta, t)$ of the disturbed rotator into the eigenfunctions of the free relativistic rotator [1, 13]; i.e.,

$$\Phi(\theta, t) = \sum_n A_n(t)\psi_n. \tag{20}$$

For any time T different from the points where the Dirac delta function is nonzero, the coefficients $A_n(t)$ evolve as [13]

$$A_n(t+T) = A_n(t)e^{-iE_n t} = A_n(t)e^{-i\sqrt{1+n^2}\theta}. \tag{21}$$

In an infinitesimal time interval around a point where the Dirac delta function is nonzero, Eq. (18) takes the form

$$i\frac{\partial\Phi}{\partial t} = -\cos\theta\,\delta\left(\frac{t}{T}\right)\Phi. \tag{22}$$

Integration of this equation over the infinitesimal interval (before and after the point where the Dirac delta function is nonzero) from $(t+T)$ to $(t+T^+)$ gives

$$\Phi(\theta, t+T^+) = \Phi(\theta, t+T)e^{iT\cos\theta}. \tag{23}$$

Decomposing both sides of this equation into the free-rotator eigenfunctions, we obtain

$$\sum_n A_n(t+T^+)\psi_n(\theta) = \sum_{r,s} A_r(t+T)\psi_r(\theta)b_s(T)e^{is\theta}. \tag{24}$$

Here, we used the expression

$$e^{iT\cos\theta} = \sum_s b_s(T)e^{is\theta}, \tag{25}$$

where

$$b_s(T) = i^s J_s(T) = b_{-s}(T), \tag{26}$$

with the Bessel functions J_s.

Multiplying both sides of Eq. (24) by ψ^* and taking into account that

$$\int_0^{2\pi} d\theta\psi_n^*\psi_r e^{is\theta} = 2\left(1+\frac{nr}{(1+E_n)(1+E_r)}\right)\delta_{s,n-r}, \tag{27}$$

we arrive at the relativistic quantum mapping

$$A_n(t + T^+) = 2\sum_r A_r(t) b_{n-r}\left(1 + \frac{nr}{(1 + E_n)(1 + E_r)}\right)e^{-iE_nT}, \quad (28)$$

which coincides, in the nonrelativistic limit ($n, r \ll 1$), with the nonrelativistic mapping from [13].

Thus, by solving the Dirac equation for a periodically excited rotator, we derive the relativistic quantum mapping for spin 1/2. This mapping is one of the first steps in the investigation of quantum chaos in relativistic systems. The above results can be used to describe various periodically excited relativistic systems, such as heavy relativistic atoms or relativistic molecules in a laser field, so-called quasiatomic mesons in a monochromatic field, etc.

REFERENCES

1. B. V. Chirikov, Phys. Rep. **52**, 265 (1979).
2. D. U. Matrasulov, V. I. Matveev, and P. K. Khabibullaev, Dokl. Akad. Nauk **367**, 321 (1999) [Dokl. Phys. **44**, 420 (1999)].
3. D. U. Matrasulov, Phys. Rev. A **60**, 700 (1999).
4. L. D. Landau and E. M. Lifshitz, *Mechanics* (Nauka, Moscow, 1988; Pergamon Press, Oxford, 1976).
5. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989; Pergamon Press, Oxford, 1976).
6. H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading Mass., 1953; Nauka, Moscow, 1975).
7. A. A. Sokolov and I. M. Ternov, *Quantum Mechanics and Atomic Physics* (Prosveshchenie, Moscow, 1970).
8. A. S. Davydov, *Quantum Mechanics* (Nauka, Moscow, 1973; Pergamon Press, Oxford, 1976).
9. G. M. Zaslavskiĭ, *Stochasticity of Dynamical Systems* (Nauka, Moscow, 1984).
10. G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to Nonlinear Physics* (Nauka, Moscow, 1988).
11. F. M. Izraelev, Phys. Rep. **196**, 299 (1990).
12. A. I. Akhiezer and V. B. Berestetskiĭ, *Quantum Electrodynamics* (Nauka, Moscow, 1981).
13. G. Casati, B. V. Chirikov, F. M. Izraelev, *et al.*, Lett. Note. Phys. **93**, 334 (1979).

Translated by R. Tyapaev

Doklady Physics, Vol. 47, No. 4, 2002, pp. 289–293. Translated from Doklady Akademii Nauk, Vol. 383, No. 5, 2002, pp. 625–629.

PHYSICS

Diffraction of Extragalactic Radio Waves on a Focusing Spherical Layer Formed by the Earth's Atmosphere

G. M. Strelkov and V. I. Naryshkin

Presented by Academician V.A. Kotel'nikov October 8, 2001

Received October 23, 2001

Observational data concerning radio-wave emission from extragalactic sources form, in large part, modern concepts of physical processes proceeding in certain regions of the Universe, as well as of its geometric and evolutionary characteristics [1–5]. A radio-wave radiation incident on the Earth partially interacts with its atmosphere. The basic characteristics of the atmosphere related to the transmission of radio waves through it, namely, the refractive index and the absorption coefficient, have been well studied. The refractive index decreases with distance from the Earth's surface following a nearly exponential law [6]. Therefore, the atmosphere represents a focusing spherical layer. Thus, it should be expected that radio waves from an extragalactic source, after passing through this layer, would be focused in the Earth's geometric shadow.

In this paper, we estimate the position, altitude, and structure of the field-intensity maxima formed in this case as applied to the millimeter-wave region.

The geometry of the problem is illustrated in Fig. 1. Let a plane radio wave from an extragalactic source be incident on the Earth surrounded by the atmosphere with the characteristic thickness Δ_a. The incident direction is determined by the wave vector $\mathbf{k}_0$. The ring-shaped part of the wave front, with inner radius R_E and width Δ_a, passes through the atmosphere, undergoes refraction and partial absorption in it, and forms a field in the Earth's geometric shadow. In order to describe radio-wave propagation in the atmosphere, we introduce the cylindrical coordinate system with the z-axis directed along the vector $\mathbf{k}_0$ and passing through the Earth's center (the point O). The coordinate z is counted off from the point O; therefore, the field intensity for an incident plane wave at its entry into the atmosphere is specified by the condition

$$E(\rho; \varphi; -(R_E + \Delta_a)) = 1. \quad (1)$$

Institute of Radio Engineering and Electronics, Russian Academy of Sciences (Fryazino Branch), pl. Vvedenskogo 1, Fryazino, Moscow oblast, 141120 Russia

Wave propagation in the atmosphere is described by the Helmholtz equation for the complex-valued field amplitude E:

$$\nabla^2 E + k^2(\mathbf{r})E = 0. \quad (2)$$

Here, $\mathbf{r}$ is the radius vector; $k = k_0(n + in_i)$ is the wave number; and n and $n_i \ll n$ are the real and imaginary parts of the complex refractive index for the medium, respectively. The dependence of the quantity n on the altitude h with respect to the Earth's surface is described by the formula corresponding to the conventional radio model of the atmosphere [6]:

$$n(h) = \begin{cases} 1 + 375 \times 10^{-6}\exp(-0.1393h), & h < 10 \text{ km} \\ 1 + 93.1 \times 10^{-6}\exp(-0.1493(h - 10)), & h \geq 10 \text{ km}. \end{cases} \quad (3)$$

Relation (3) allows us to consider the problem of evaluating the amplitude E as an axially symmetric one. We write out the amplitude $E(\mathbf{r})$ in the form

$$E(\mathbf{r}) = A(\mathbf{r})\exp(ik_0\psi(\mathbf{r})), \quad (4)$$

where $A(\mathbf{r})$ is a complex-valued (in contrast to, for example, the suggestion of [7, 8]) amplitude function and $\psi(\mathbf{r})$ is a real-valued function. Taking into account that the equality $n^2(\mathbf{r}) - n_i^2(\mathbf{r}) = n^2(\mathbf{r})$ is satisfied with high accuracy, we obtain, instead of Eq. (2), the relationship

$$\nabla^2 A + 2ik_0 \times \nabla\psi \times \nabla A + ik_0 A \times \nabla^2\psi + 2ik_0^2 nn_i A - k_0^2[(\nabla\psi)^2 - n^2(\mathbf{r})]A = 0. \quad (5)$$

We choose the function $\psi(\mathbf{r})$ such that it satisfies the equation

$$(\nabla\psi)^2 - n^2(\mathbf{r}) = 0. \quad (6)$$

In this case, the quantity $\psi(\mathbf{r})$ is defined as an eikonal and is expressed in terms of a curvilinear integral along

1028-3358/02/4704-0289$22.00

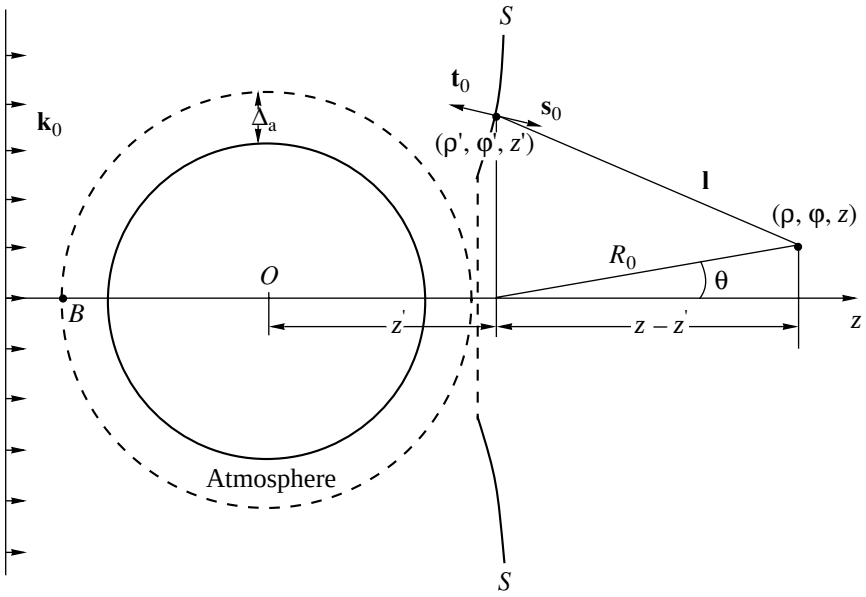


Fig. 1. Geometry of the problem.

a ray $\mathbf{r}(s)$ (s is the ray length). The trajectory of this ray is described by the equation [7]

$$\frac{d(n\mathbf{s}_0)}{ds} = \nabla n(\mathbf{r}), \tag{7}$$

where $\mathbf{s}_0$ is the unit vector tangent to the curve $\mathbf{r}(s)$.

Together with the coordinate system (ρ, φ, z), we introduce the orthogonal, axially symmetric, curvilinear coordinate system (Φ, φ, ψ) with its origin at the point B (see Fig. 1). In this case, the coordinate lines are the rays $\mathbf{r}(s)$, along which Φ = const and φ = const, and the lines orthogonal to them, along which φ = const and ψ = const. These lines are referred to as front lines. The coordinate φ coincides with the cylindrical coordinates introduced above and therefore needs no comment. In the newly introduced coordinate system, the vector $\mathbf{s}_0$ is a unit vector along the rays, while the unit vector directed along the front lines is denoted as $\boldsymbol{\phi}_0$. Therefore, the coordinates ψ and Φ of an arbitrary point $M(\Phi, \varphi, \psi)$ in the plane φ = const are the value of the eikonal at this point and the value of Φ for the ray passing through it, respectively. For $z = z_B$, a plane wave incident on the Earth coincides with the coordinate surface $\psi = 0$. In the coordinates (Φ, φ, ψ), the ψ-dependent terms entering into Eq. (5) take the form

$$\nabla\psi = \frac{1}{\sqrt{g_{33}}}\mathbf{s}_0, \quad 2ik_0\nabla\psi\times\nabla A = 2ik_0 n\frac{1}{\sqrt{g_{33}}}\frac{\partial A}{\partial\psi}; \tag{8}$$

$$\nabla^2\psi = \frac{n}{R} + \frac{n}{\rho}\mathbf{s}_0\operatorname{grad}\rho + \mathbf{s}_0\operatorname{grad} n. \tag{9}$$

In these equations, g_{11} and g_{33} are the metric tensor components being determined in orthogonal coordinates by the relation

$$\begin{aligned}(\delta l)^2 &= (\delta\phi)^2 + \rho^2(\delta\varphi)^2 + (\delta s)^2 \\ &= g_{11}(\delta\Phi)^2 + \rho^2(\delta\varphi)^2 + g_{33}(\delta\psi)^2.\end{aligned} \tag{10}$$

Here, δl is the distance (in the cylindrical coordinates) between two close spatial points; $\delta\Phi$, $\delta\varphi$, and $\delta\psi$ are the coordinate differences for these points; $\delta\phi = \delta l|_{\varphi,\,\psi=\text{const}}$; and $\delta s = \delta l|_{\Phi,\,\varphi=\text{const}}$. According to Eq. (6), $g_{33} = \frac{1}{n^2}$. In order to find the quantity g_{11} at a point (Φ, φ, ψ) on the wave front, the distance δl from this point to a close point $(\Phi + \delta\Phi, \varphi, \psi)$ should be evaluated in the cylindrical coordinates with $g_{11} = \left(\frac{\delta l}{\delta\Phi}\right)^2$. In Eq. (9), the radius of curvature for an element $\delta\phi(\Phi, \varphi, \psi)$ in the plane φ = const is denoted by R. This radius is determined by the relationship $R = 2g_{11}\sqrt{g_{33}}\left(\frac{\partial g_{11}}{\partial\psi}\right)^{-1}$.

Substituting Eqs. (8) and (9) into (5) and taking into account that the problem is axially symmetric and that the angle of refraction in the Earth's atmosphere does not exceed $\sim 2.4\times10^{-2}$ [6] and $\Delta_a \ll 2R_E$, we arrive at the parabolic equation for the complex amplitude function in the coordinates (Φ, φ, ψ):

$$\begin{aligned}2ik_0 n\rho\sqrt{g_{11}}\frac{\partial A}{\partial\psi} &= \frac{\partial}{\partial\Phi}\left(\frac{\rho}{n\sqrt{g_{11}}}\frac{\partial A}{\partial\Phi}\right) - ik_0\rho\sqrt{g_{11}} \\ &\times\left[\frac{n}{R} + \frac{n}{\rho}\mathbf{s}_0\operatorname{grad}\rho + \mathbf{s}_0\operatorname{grad} n + \gamma\right]A,\end{aligned} \tag{11}$$

where $\gamma = 2k_0 n n_i$ is the absorption coefficient. According to Eq. (1), the amplitude function for the plane incident wave front (which is perpendicular to the z-axis and passes through the point B) is given by the relation $A(\Phi, \varphi, 0) = 1$. The upper $\Phi = \Phi_{max}$ and lower $\Phi = \Phi_{min}$ limits of integration over Φ are determined by the ray trajectories originating at points of the initial front, which bound the interval of values of Φ for $\psi = 0$. In this case, we choose the boundaries such that, at $z = 0$, the upper boundary passes through the mesopause and the lower one is tangent to the Earth's surface. The following boundary conditions are imposed:

$$A(\Phi_{max}; \varphi; \psi) = 1; \tag{12}$$

$$A(\Phi_{min}; \varphi; \psi) = X(\Phi_{min}; \varphi; \psi)\exp\left(\frac{-\tau(\Phi_{min}; \varphi; \psi)}{2}\right). \tag{13}$$

In condition (13), we denote the attenuation factor due to wave refraction [6] and the optical depth of the atmosphere for the lower ray as X and τ, respectively, both quantities being functions of the ray length. For rays passing through the atmosphere close to the Earth's surface, the values of τ for millimeter waves amount to several tens or even hundreds depending on the particular value of λ.

Each evaluation step over ψ involved three stages. Initially, a successive front line was constructed using Eqs. (6) and (7). Then, the amplitude A was evaluated from this front and Eq. (11). Finally, the front was corrected so that the equiphase condition was satisfied on it. The solution to Eq. (11) was found by the Crank–Nicholson scheme [9].

In the region behind the Earth, i.e., for $z \geq R_E + \Delta_a$, where $n = 1$, $k = k_0$, and $\gamma = 0$, the following integral formula describes the solution to Eq. (2):

$$E(\rho; \varphi; z) = \frac{1}{4\pi}\int_S \left(G\frac{\partial E}{\partial t} - E\frac{\partial G}{\partial t}\right)dS. \tag{14}$$

Here, S is the integration surface, $\mathbf{t}_0$ is the outer normal to S, $G = \dfrac{e^{ik_0 l}}{l}$ is the Green's function for Eq. (2) at $n = 1$, $l = |\mathbf{l}|$, and $\mathbf{l}$ is the vector going out from the point (ρ', φ', z') on the surface S to the observation point (ρ, φ, z). It is natural to take the equiphase wave front as the surface of integration in Eq. (14). This surface is found by solving the problem of the passage of a plane wave through the atmosphere. Therefore, the field intensity and its normal derivatives are known for this surface, with $\mathbf{t}_0 = -\mathbf{s}_0$.

After expanding the derivatives entering into the integrand in Eq. (14), we reduce it to the form

$$E(\rho; \varphi; z) = -\frac{e^{ik_0\psi'}}{4\pi}\int_{\Phi_{min}}^{\Phi_{max}} d\Phi' \sqrt{g_{11}}\rho' \int_0^{2\pi} e^{ik_0 l}\left(\frac{1}{l}\frac{\partial A}{\partial \psi} + A\left(\frac{ik_0}{l} + \frac{ik_0 l - 1}{l^3}(\mathbf{s}_0 \cdot \mathbf{l})\right)\right)d\varphi'. \tag{15}$$

When integrating over φ', we represent the modulus of the vector $\mathbf{l}$ in the form [10]

$$l = R_0\left(1 + \left(\frac{\rho'}{R_0}\right)^2 - 2\frac{\rho'}{R_0}\sin\theta\cos\varphi'\right)^{1/2}, \tag{16}$$

where $R_0 = \sqrt{(z - z')^2 + \rho^2}$ and $\sin\theta = \dfrac{\rho}{R_0}$. Under the condition

$$\frac{2\rho'}{R_0}\sin\theta \ll 1, \tag{17}$$

the following expansions are valid [$r_0 = \sqrt{R_0^2 + (\rho')^2}$]:

$$l = r_0 - \frac{\rho\rho'\cos\varphi'}{r_0} + \ldots, \quad \frac{1}{l} = \frac{1}{r_0} + \frac{\rho\rho'\cos\varphi'}{r_0^3} - \ldots; \tag{18}$$

$$\frac{ik_0 l - 1}{l^3} = \frac{ik_0 r_0 - 1}{r_0^3} - \frac{3 - 2ik_0 r_0}{r_0^5}\rho\rho'\cos\varphi' + \ldots. \tag{19}$$

In addition,

$$\mathbf{s}_0\mathbf{l} = \Delta z'\cos\alpha - \rho'\sin\alpha + \rho\sin\alpha\cos\varphi', \tag{20}$$

where $\Delta z' = z - z'$ and α is the angle between the unit vector $\mathbf{s}_0$ and the z-axis. Substituting relationships (18)–(20) into Eq. (15) and integrating it over φ', we arrive at the final expression for the field intensity in the Earth's geometric shadow:

$$E(\rho; \varphi; z) = \frac{e^{ik_0\psi'}}{2}\int_{\Phi_{min}}^{\Phi_{max}} d\Phi' \sqrt{g_{11}}\frac{e^{ik_0 r_0}}{r_0}\rho' \times \left[\frac{\partial A}{\partial \psi}\left[J_0\left(\frac{k_0\rho\rho'}{r_0}\right) - i\frac{k_0\rho\rho'}{r_0}J_1\left(\frac{k_0\rho\rho'}{r_0}\right)\right] + A\left[\left(\frac{ik_0 r_0 - 1}{r_0^2}(\Delta z'\cos\alpha - \rho'\sin\alpha) + ik_0\right)J_0\left(\frac{k_0\rho\rho'}{r_0}\right) + i\left(\frac{ik_0 r_0 - 1}{r_0^2}\rho\sin\alpha + i\frac{k_0\rho\rho'}{r_0^2} - \frac{3 - 2ik_0 r_0}{r_0^4} \times (\Delta z'\cos\alpha - \rho'\sin\alpha)\rho'\rho\right) \times J_1\left(\frac{k_0\rho\rho'}{r_0}\right) - \frac{3 - 2ik_0 r_0}{r_0^4}\rho'\rho^2\sin\alpha\right]\right] \tag{21}$$

$$\times\left[J_0\left(\frac{k_0\rho\rho'}{r_0}\right)-J_2\left(\frac{k_0\rho\rho'}{r_0}\right)\right]\Bigg]\Bigg].$$

The diffraction pattern of the field in the shadow region is basically determined by the radiation wavelength λ and the altitude dependences for both the refractive index and the atmospheric absorption coefficient. The effect of the latter on the focusing phenomenon is briefly considered below for the wavelength λ = 2.109 mm, which corresponds to the absorption resonance of ozone. The altitude dependence of γ is determined by the absorption in oxygen, water vapor, and ozone, with ozone absorption prevailing for $h \geq \sim 15$ km [11]. A shift of only ±0.005 mm in the wavelength λ does not significantly affect the absorption coefficients for O_2 and H_2O but results in a frequency shift out of the resonance region occupied by the O_3 line. Therefore, the altitude dependence of γ varies significantly [11]. The densities $|A|^2$ of the relative radiation-energy flux, which were evaluated by means of Eq. (21), are shown in Fig. 2 for a spatial region in the neighborhood of the *z*-axis. The quantity $|A|^2$ attains a maximum $|A_{max}|^2$ at the point z_{max} marked by the cross. In the central darkest region, the values of $|A|^2$ range from 0.1 to 0.25 of $|A_{max}|^2$. The regions corresponding to decreasing degrees of darkness correspond to ratios of $\frac{|A|^2}{|A_{max}|^2}$ within the ranges 0.25–0.01, 0.01–0.0025, 0.0025–0.0001, and 0.0001–0. A typical interference pattern is observed in the vicinity of the symmetry axis. Indeed, spatial regions in which the values of $|A|^2$ are several tens (or even hundreds) of thousands of times larger than those in the incident wave alternate with regions in which $|A|^2$ is close to zero. The region in which the quantity $|A|^2$ attains its maximum is similar to a certain cord coinciding with the *z*-axis. The diameter of the cord is only about 10–30 cm, while its length amounts to several million kilometers. The focusing effect becomes pronounced at a certain distance from the Earth (see Fig. 3). The lack of atmospheric absorption at altitudes exceeding about 15 km most significantly affects the field characteristics in the vicinity of the *z*-axis. Indeed, the values of $|A_{max}|^2$ corresponding to Fig. 2a and Fig. 2b, respectively, are equal to 9.52×10^4 ($z_{max} = 7.3 \times 10^5$ km) and 4.94×10^5 ($z_{max} = 8.0 \times 10^5$ km). In the latter case, both the transverse and longitudinal cord sizes are doubled.

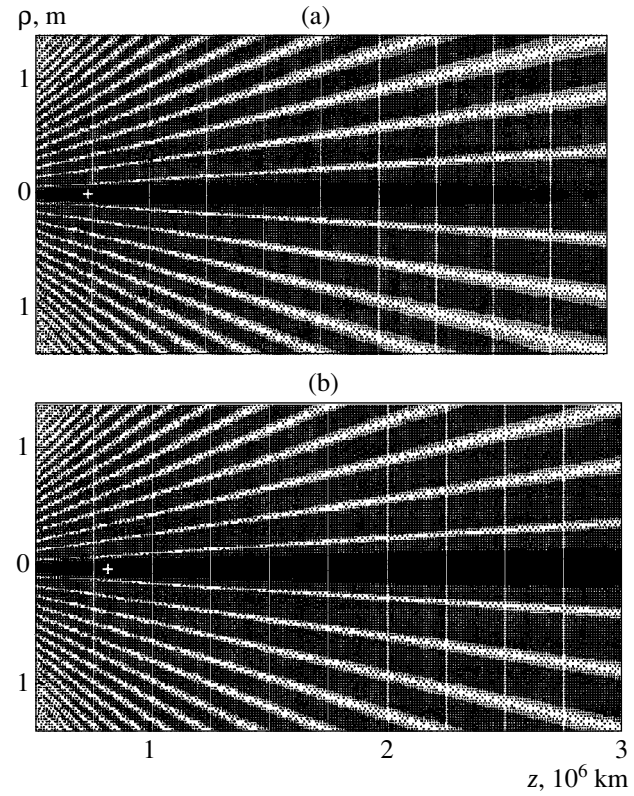


Fig. 2. Relative radiation energy-flux density in the Earth's geometric shadow (a) with and (b) without allowance for the absorption in ozone (λ = 2.109 mm and Δ_a = 100 km).

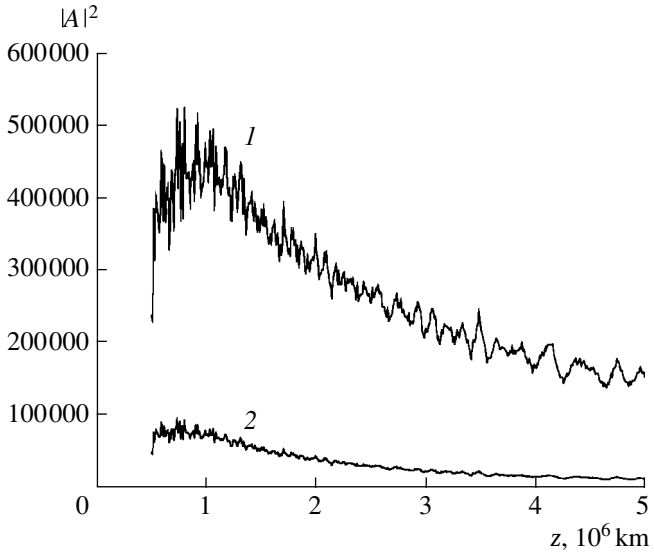


Fig. 3. Relative radiation energy-flux density along the *z*-axis (*2*) with and (*1*) without allowance for ozone absorption (λ = 2.109 mm).

Thus, the estimates obtained substantiate the hypothesis on strong focusing of millimeter-wavelength extragalactic radiation in the Earth's geometric shadow.

REFERENCES

1. M. G. Larionov, Pis'ma Astron. Zh. **24**, 3 (1998) [Astron. Lett. **24**, 1 (1998)].

2. M. G. Larionov, Pis'ma Astron. Zh. **22**, 434 (1996) [Astron. Lett. **22**, 390 (1996)].

3. G. A. Moellenbrock, K. Fujisawa, R. A. Preston, *et al.*, Astron. J. **111**, 2174 (1996).

4. T. Jamada, K. Ohta, A. Tomita, and T. Takata, Astron. J. **110**, 1564 (1995).
5. K. S. Stankevich, V. P. Ivanov, and S. P. Stolyarov, Pis'ma Astron. Zh. **25**, 584 (1999) [Astron. Lett. **25**, 501 (1999)].
6. M. A. Kolosov and A. V. Shabel'nikov, *Refraction of Electromagnetic Waves in the Earth's, Venusian, and Martian Atmospheres* (Sov. Radio, Moscow, 1976).
7. Yu. A. Kravtsov, S. M. Rytov, and V. I. Tatarskiĭ, *Introduction to Statistical Radio Physics* (Nauka, Moscow, 1983), Vol. 2.
8. M. B. Vinogradova and A. A. Semenov, *Basic Principles of the Propagation Theory for Ultra-Short Waves in the Troposphere* (Academy of Sciences of the USSR, Moscow, 1963).
9. G. I. Marchuk, *Methods of Computational Mathematics* (Springer-Verlag, New York, 1975; Nauka, Moscow, 1980).
10. R. King and G. Smith, *Antennas in Matter* (MIT Press, Cambridge (USA), 1981; Mir, Moscow, 1984), Vol. 2.
11. G. M. Strelkov, Issledovaniya Zemli iz Kosmosa, No. 1, 25 (1995).

Translated by V. Chechin

Doklady Physics, Vol. 47, No. 4, 2002, pp. 294–296. Translated from Doklady Akademii Nauk, Vol. 383, No. 6, 2002, pp. 758–760.

PHYSICS

Stability Criteria for Sound Modes and Alfvén Modes of Plasma in Toroidal Systems

A. B. Mikhailovskiĭ and A. A. Skovoroda

Presented by Academician A.M. Fridman December 3, 2001

Received December 7, 2001

1. Recently, a series of so-called alternative toroidal systems has been suggested [1–3]. In contrast to classical toroidal systems, such as tokamaks and stellarators, whose magnetohydrodynamic (MHD) stability is calculated on the basis of the well-known Mercier stability criterion (see, e.g., [4, chapters 7–9]), the alternative systems are assumed to be stable if a stability criterion of the type [5]

$$-\nabla p \times \nabla U + \frac{\gamma_0 p (\nabla U)^2}{|U|} \geq 0 \qquad (1)$$

is satisfied. Here, p is the equilibrium plasma pressure, γ_0 is the adiabatic exponent, $U = -\oint \frac{dl}{B}$, B is the equilibrium magnetic field, and dl is the length element of a magnetic-field line. The magnetic-field lines are assumed to be closed so that $\oint$ implies the integral along the entire field line.

The assumption of closure of the field lines for a toroidal system is an evident idealization associated with ignoring the magnetic-field shear. A rejection of this idealization is, to a certain extent, a rejection of the idea of alternativeness of the corresponding systems. Then, generally speaking, such systems turn out to be nothing but some varieties of stellarators. In this case, it seems to be of practical importance to know whether toroidal systems satisfying a condition of type (1) have any additional advantages over systems that satisfy the Mercier stability criterion but do not satisfy a condition of type (1). Thus, in connection with the development of a program of studies of alternative toroidal systems, it is necessary to clarify the question of the most important stability criterion. Is it the Mercier stability criterion or a criterion of type (1)? The goal of the present paper is to find an approach to solving this dilemma.

2. At first glance, we may think that this dilemma is being solved trivially and the answer is that the Mercier stability criterion is more important. Such a standpoint has a rather reasonable foundation, the essence of which is the following. According to the well-known results of the theory of MHD plasma instabilities (see, e.g., [6, chapter 12] and references therein), the Mercier stability criterion characterizes the so-called most dangerous modes. In other words, it predicts the lowest thresholds for exciting instabilities. It follows from this fact that, if the Mercier stability criterion is violated, the plasma is unstable regardless of whether a stability criterion of type (1) is satisfied or violated.

Meanwhile, the importance of the instabilities is determined not only by their excitation thresholds, but also by their growth rates. It is evident that, all other conditions being equal, the most important instabilities are those with the largest growth rates. In addition, we should keep in mind that instabilities with sufficiently small growth rates, predicted by the ideal magnetohydrodynamics, may be absent in reality. This is a result of their suppression by effects lying beyond the scope of ideal magnetohydrodynamics, such as drift effects or effects of finite plasma resistivity (cf. [4, chapters 21, 22] for details). For this reason, the Mercier stability criterion cannot be *a priori* considered as the most important one.

Consequently, in order to solve the above-formulated dilemma, analysis of the growth rates of the instabilities is necessary. From the mathematical standpoint, such an analysis is a much more complicated problem than deriving the stability criteria. This is clear from the fact that when deriving a stability criterion, we are dealing with mode equations which are simplified by ignoring terms with the mode growth rates. At the same time, analysis of the growth rates implies extracting the necessary information from nonsimplified equations. Development of an approach to studying such equations is one more goal of the present paper.

3. Initially, we consider the problem of interest using the plasma cylinder model, and then we explain how this problem should be solved in the general case of toroidal systems.

We now make use of the explanations given in [4] (see chapter 4 of this book) with allowance for effects important for deriving both the Mercier stability crite-

Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia

1028-3358/02/4704-0294$22.00

rion and the criterion of type (1). With their help, we find that, for finite growth rates, the MHD modes are described by the equation for small oscillations of the form

$$\frac{d}{dx}x^2\frac{d\xi}{dx} - x^2\xi + x_A^2\frac{d^2\xi}{dx^2} - \left(U_0 + \frac{U_c}{1 + x^2/x_s^2} + x_A^2\right)\xi = 0. \quad (2)$$

Here, ξ is the perturbed radial displacement of plasma and x is a dimensionless variable that characterizes the distance from the "rational" magnetic surface (in which the magnetic-field lines are closed). Furthermore, $x_A = \frac{\gamma L_s}{c_A}$ and $x_s = \frac{\gamma L_s}{c_s}$ are dimensionless quantities characterizing the so-called Alfvén and sound length scales, respectively; γ is the growth rate; L_s is the so-called shear length, i.e., a distance characterizing the shear of the magnetic field; and c_A and c_s are the Alfvén and sound speeds, respectively. The parameters U_0 and U_c are defined by formulas

$$U_0 = \frac{2L_s^2 p'}{\gamma_0 R p}\frac{c_s^2}{c_A^2}, \quad (3)$$

$$U_c = \frac{4L_s^2}{R^2}\frac{c_s^2}{c_A^2}. \quad (4)$$

Here, R is the radius of curvature for the field lines; the prime corresponds to the radial derivative.

The parameter U_0 characterizes the magnetic well (for $U_0 > 0$) or the magnetic hill (for $U_0 < 0$). Physically, it means the same as the first term on the left-hand side of inequality (1). In terms of U_0, the Mercier stability criterion is written as

$$\frac{1}{4} + U_0 > 0. \quad (5)$$

The term $\frac{1}{4}$ on the left-hand side of this inequality characterizes the stabilizing effect of the shear.

The parameter U_c describes the plasma-compressibility effect characterized by the second term on the left-hand side of inequality (1). In terms of U_0 and U_c, this inequality implies that

$$U_0 + U_c > 0. \quad (6)$$

The essence of the approach to solving Eq. (2), which is developed in the present paper, is the following. The space x is divided into three regions: $x \gg x_s$, $x \approx x_s$, $x \approx x_A$. In each of these regions, Eq. (2) has an exact analytical solution. In the region $x \gg x_s$ and in the regions $x \approx x_s$, $x \approx x_A$, such a solution is expressed in terms of Bessel functions and hypergeometric functions, respectively. The solution found for the region $x \approx x_s$ is sewn at large x with the solution found for the region $x \gg x_s$. For small x, it is sewn with the solution found for the region $x \approx x_A$. Details of such a sewing procedure can be found by turning to [4, Section 5.1.1]. As a result of the above procedure, we find a dispersion relation containing all desired information on the mode growth rates.

We have analyzed this information with an accent on solving the dilemma formulated above. Furthermore, it has been determined that if, on the one hand, the Mercier stability criterion (5) is violated, i.e.,

$$\frac{1}{4} + U_0 < 0, \quad (7)$$

and, on the other hand, an inequality of type (6), which is complemented by the term with the shear, is satisfied, i.e.,

$$\frac{1}{4} + U_0 + U_c > 0, \quad (8)$$

then only the sound modes with the growth rate

$$\gamma \approx \frac{c_s}{L_s} \quad (9)$$

are unstable. Under above conditions, Alfvén modes are not excited. At the same time, if the stability criterion (8) is violated, these modes will be excited with growth rates on the order of $\gamma \approx \frac{c_A}{L_s}$, which are considerably larger than those for sound modes (9) $\left(\text{by a factor of } \frac{c_A}{c_s}\right)$.

Thus, our analysis shows that the stability criterion of type (1) [to be more precise, of type (8)], but not the Mercier stability criterion, is more important. Hence, it follows that the above-mentioned alternative toroidal systems have certain advantages over the traditional stellarators.

We note that, from heuristic positions, stability criterion (8) was initially proposed in [7]. Therefore, it can be called the Trubnikov stability criterion. Stability criterion (1) goes back to [8]. The importance of the stability criterion of type (1) was initially emphasized in [9], then in [5], and in a series of other papers by the same author. In this connection, the stability criterion of type (1) can be called the Bernstein–Kadomtsev stability criterion.

4. According to [10], MHD stability in arbitrary toroidal systems is determined by the parameters E, F, G, H, K introduced therein. In addition, an approach to calculating these parameters was given in [10]. In this connection, they can be called Glasser parameters.

Turning to [10], we can find that, in terms of the Glasser parameters, the equation for the MHD modes of an ideal plasma can be represented in the form

$$\frac{d}{dx}\left[\left(1+\frac{KH^2}{G+KF}\frac{1}{1+x^2/x_s^2}\right)x^2\frac{d\xi}{dx}\right]+x_A^2\frac{d^2\xi}{dx^2}$$

$$+(E+F+H)\left[1-\frac{K(E+F+H)}{G+KF}\frac{1}{1+x^2/x_s^2}\right. \qquad (10)$$

$$\left.+\frac{KH}{G+KF}\frac{d}{dx}\left(\frac{x}{1+x^2/x_s^2}\right)\right]\xi = 0.$$

The expressions for x_A and x_s entering into this equation can be restored by turning to [10]. It follows from (10) that the parameters U_0 and U_c entering into the Mercier and BKT stability criteria, i.e., into (5) and (8), are related to the Glasser parameters by the relationships

$$U_0 = -(E+F+H), \qquad (11)$$

$$U_c = \frac{K(E+F+H)(E+F+H^2)}{G+K(F+H^2)}. \qquad (12)$$

Using (11), (12), we can determine the importance of these criteria for a certain specified toroidal system in the same manner as for a plasma cylinder. In addition, to identify the compression trend as a whole, stability criterion (8) can be called the Bernstein–Kadomtsev–Trubnikov (BKT) stability criterion.

DEDICATION

This work is dedicated to the blessed memory of our colleague Dmitriĭ Aleksandrovich Panov.

ACKNOWLEDGMENTS

The authors are grateful to Drs. V.V. Arsenin, S.V. Konovalov, M.I. Mikhailov, S.E. Sharapov, A.V. Timofeev, and V.S. Tsypin for fruitful discussions that stimulated this work.

This work was supported in part by the Council for Grants and State Support of Leading Scientific Schools of the President of the Russian Federation, project no. 00-15-96526, and the Russian Foundation for Basic Research, project no. 00-02-17105.

REFERENCES

1. A. I. Morozov, V. P. Pastukhov, and A. Yu. Sokolov, in *Proc. Workshop on* D–^{3}He *Based Reactor Studies, Moscow, Russia, 1991* (Kurchatov Inst. Atomic Energy, Moscow, 1991), p. 1.
2. J. Kesner, L. Bromberg, D. Garnier, and M. Manuel, in *Proc. 17th Int. Atomic-Energy Agency Fusion Energy Conf., Yokohama, Japan, 1998* (Yokohama, 1998), Vol. 3, p. 1165.
3. V. V. Arsenin, E. D. Dlougach, V. M. Kulygin, *et al.*, Nuclear Fusion **41**, 945 (2001).
4. A. B. Mikhailovskiĭ, *Instabilities in a Confined Plasma* (Inst. Phys. Publ., Bristol, 1998).
5. B. B. Kadomtsev, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Gosatomizdat, Moscow, 1963; Consultants Bureau, New York, 1966), Vol. 2, p. 153.
6. A. B. Mikhailovskiĭ, *Theory of Plasma Instabilities* (Atomizdat, Moscow, 1971; Consultants Bureau, New York, 1974), Vol. 2.
7. B. A. Trubnikov, *Theory of Plasma* (Énergoatomizdat, Moscow, 1996).
8. I. B. Bernstein *et al.*, Proc. Roy. Soc. A **224**, 17 (1958).
9. B. B. Kadomtsev, *Plasma Physics and the Problems of Controlled Thermonuclear Reactions* (Akad. Nauk SSSR, Moscow, 1958; Pergamon, London, 1960), Vol. 4, p. 450.
10. A. H. Glasser, J. M. Greene, and J. L. Johnson, Phys. Fluids **18**, 875 (1975).

Translated by authors

Doklady Physics, Vol. 47, No. 4, 2002, pp. 297–301. Translated from Doklady Akademii Nauk, Vol. 383, No. 5, 2002, pp. 630–634.

TECHNICAL PHYSICS

Highest Wave Types in Transmission Lines for Three-Dimensional Microwave and Extremely High Frequency Integrated Circuits

S. B. Klyuev, E. V. Mezhekova, E. I. Nefedov, and R. S. Popov

Presented by Academician O.M. Belotserkovskiĭ February 28, 2001

Received June 21, 2001

1. SYSTEMS FOR SUPERHIGH-SPEED INFORMATION PROCESSING IN MICROWAVE AND EXTREMELY HIGH FREQUENCY RANGES

One of the fundamental problems at the modern stage of the development of human society is the necessity of systems for superhigh-speed information processing (SSIP) that are highly efficient, technologically effective, and relatively inexpensive in mass production [1]. It is also required that these SSIP have a high operating speed; be reliable while receiving, storing, and transmitting information; and have low mass and small size, high technological effectiveness, low cost, etc. Personal computers, which have become very popular throughout the world, are now produced on a large scale.

Modern personal computers are based on large-scale integrated (LSI) and superlarge-scale integrated (SLSI) circuits with a clock frequency $f_m \sim 50$–300 MHz. However, an increase in the clock frequency even by a factor of 10, and especially of 10^2 to 10^3, meets certain fundamental difficulties [2, 3].

The alternative to this direction are SSIP in which information processing proceeds immediately within microwave and extremely high (EHF) frequency ranges at clock frequencies of 10 to 100–500 GHz. The realization of this method of information processing is possible on the basis of three-dimensional integrated circuits (3D IC), which were proposed and partly mastered in Russia more than 20 years ago [1, 4–6]. These IC have many remarkable properties. For example, their mass and size parameters are better by one to three orders of magnitude. The operating speed of microwave and/or EHF SSIP modules is also improved by two to four orders of magnitude, etc. [7–10].

In this paper, we consider possibilities for a regime of highest wave types in certain kinds of transmission lines used in microwave 3D IC and EHF 3D IC.

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, pl. Vvedenskogo 1, Fryazino, Moscow oblast, 141190 Russia

2. A GENERALIZED DOUBLE-SLOTTED WAVEGUIDE: SETTING THE PROBLEM AND METHOD OF SOLVING

One of the basic types of transmission lines for microwave and EHF 3D IC is a generalized double-slotted waveguide (DSW) whose cross section is shown in Fig. 1. The electrodynamic theory of a generalized DSW was developed in [13], where attention was mainly attracted to the single-wave regime. Here we consider as a basic generalized-DSW model its closed version with imaginary (virtual) electric and/or magnetic walls spaced at a distance b (Fig. 2).

To analyze the structure presented in Fig. 2, we employ one of the most efficient methods of computational electrodynamics, namely, the method of partial domains [6, 11–14]. In this method, the electric and magnetic fields in the domains *1–3* (Fig. 2) are written in the form of infinite series for the eigenfunctions of the Helmholtz equation. The application of boundary conditions for tangents composing the electromagnetic field leads to a system of equations that link Fourier transforms of the electric-field tangent components in

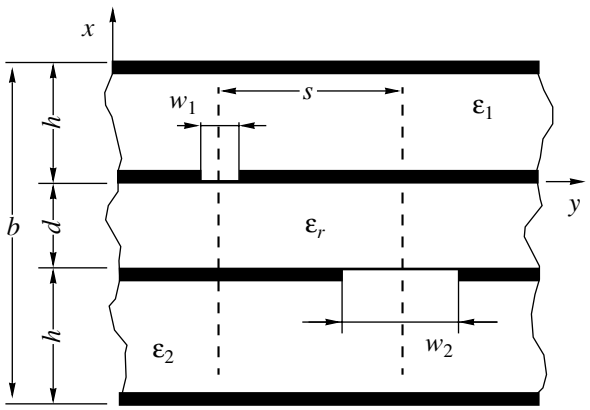


Fig. 1. Cross section of a generalized double-slotted waveguide (almost open transmission line).

1028-3358/02/4704-0297$22.00 © 2002 MAIK “Nauka/Interperiodica”

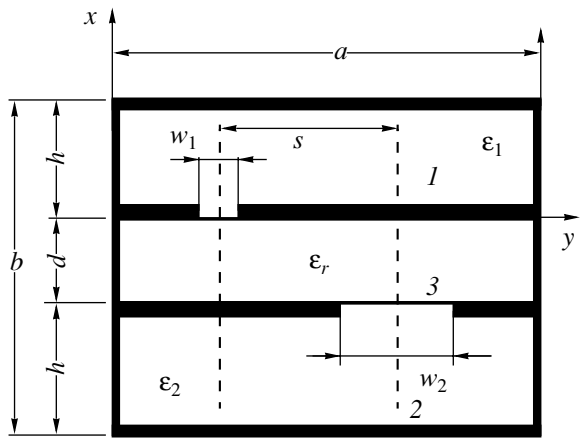


Fig. 2. Closed model of a generalized double-slotted waveguide.

slot apertures with Fourier transforms of the electric-current density in the metal that forms the slots:

$$\begin{bmatrix} \tilde{J}_y \\ \tilde{J}_z \end{bmatrix} = \begin{bmatrix} \tilde{G}_{yy} & \tilde{G}_{yz} \\ \tilde{G}_{zy} & \tilde{G}_{zz} \end{bmatrix} \begin{bmatrix} \tilde{E}_y \\ \tilde{E}_z \end{bmatrix}. \quad (1)$$

Here, $\tilde{G}_{ij}$ are the Green's function elements of the waveguide structure (Fig. 2) [13].

We use the Galerkin method to algebraize relationship (1). In this method, the aperture distribution of the tangential components of the electric field is represented in the form of an expansion in terms of the first- and second-kind Chebyshev polynomials with allowance for the behavior of the field components near the edge:

$$E_y^{(i)} = \sum_{k=0}^{\infty} a_k^i f_k^{(i)}(y) = \sum_{k=0}^{\infty} \frac{a_k^i T_k(y)}{\sqrt{1-y^2}},$$
$$E_z^{(i)} = \sum_{k=0}^{\infty} b_n^i g_n^i(y) = \sum_{k=0}^{\infty} b_n^i U_n(y)\sqrt{1-y^2}. \quad (2)$$

Here, $T_k(\cdot)$ and $U_n(\cdot)$ are the first- and second-kind Chebyshev polynomials of the *k*th and *n*th orders, respectively, and *i* is the slot ordering number.

As a result of employing the Galerkin method to relationship (1) and after the truncation (reduction) of series (2), we arrive at a homogeneous system of linear algebraic equations (SLAE) with respect to unknown coefficients in the expansion of aperture fields [13]:

$$\begin{bmatrix} P & Q \\ R & S \end{bmatrix} \begin{bmatrix} a_i^1 \\ b_i^1 \\ a_j^2 \\ b_j^2 \end{bmatrix} = \vec{0}. \quad (3)$$

The condition for the solution of homogeneous SLAE (3) to be nontrivial results in the dispersion equation for the waveguide structure:

$$\det||PQRS|| = 0. \quad (4)$$

This equation can be solved with respect to the longitudinal wave number on the basis of a certain numerical method (e.g., Newton's method) using a computer.

3. RESULTS

In accordance with the method developed and described in the preceding section (see also [13, 14]), thorough numerical and numerical-analytical investigations were carried out. In them, wave characteristics for several types of transmission lines that are most interesting for SSIP based on 3D IC, preferably in the EHF range, were determined. We begin our analysis with a DSW that presents a baseline in our further consideration.

3.1. Double-slotted waveguide ($w_1 = w_2$, $s = 0$, see Fig. 2). This type of DSW is widely used in 3D IC of the EHF range. In our study, basic attention was attached to the highest fast-wave types. Previously, the behavior and properties of the lowest wave types, namely, H_1 quasi-waves, were thoroughly investigated in [5, 15]. In particular, it was shown that in order to realize the single-wave regime intrinsic to planar IC, the conditions $d < \frac{\lambda}{\sqrt{\varepsilon}}$ and $h_1 = h_2 = d\sqrt{\varepsilon}$ had to be fulfilled. We are interested in a more general case of the multiwave regime where several half-waves λ_0 can be contained within the length $d\sqrt{\varepsilon}$. The relationship

$$k\sqrt{\varepsilon}d = 2\pi q + p, \quad q = 1, 2, \ldots \quad (5)$$

(where p is a complex-valued parameter such that $|p| \ll 1$) is the condition for the existence of a resonance between the plates composing a DSW. This condition is well known in the theory of open cavities, as well as in the theory of microstrip lines and resonators based on them.

Condition (5) implies that an almost integer number of half-waves lie within the gap *d* between the metallic plates.

The concept of the behavior of slowing-down $\eta = \frac{\lambda_0}{\lambda'}$ for several wave types in a DSW as a function of the normalized distance $\tilde{d} = \frac{d\sqrt{\varepsilon_r}}{\lambda_0}$ between the metallic plates at a constant distance between screening metal layers makes it possible to obtain the numerical data given in Fig. 3. In Fig. 3 (as in Fig. 4), curves related to a DSW are plotted by solid lines. Curves (dashed lines)

for a planar waveguide $\left(w_1 = w_2 = \infty, \frac{\partial}{\partial x} \equiv 0\right)$ with a dielectric plate of the same thickness d are also shown in the same figures. The regions $\eta > 1$ and $\eta < 1$ correspond to the slow and fast waveguide waves, respectively. In accordance with relationship (5), there are characteristic resonance points $\tilde{d}_{\text{res}} = \frac{q}{2}$, $q = 1, 2, \ldots$ on the abscissa axis. The resonance points are characterized by the existence of certain minima on the curves for DSW waves. These minima coincide with flattened segments for slowing-down curves corresponding to a planar waveguide containing a dielectric plate. Thus, in the case of a resonance between the plates forming the DSW, slowing down of fast waves is basically determined by the parameters of the dielectric layer. This corresponds to the concepts developed for waves of the highest types (see, e.g., [5]) in microstrip lines when resonance-type impedance boundary conditions can be put at the boundaries for $x = w$ and $x = \frac{a + w}{2}$, and $y \in (h, h + d)$. In this case, the field of the wave propagating along the transmission line barely penetrates into the space between the metallic DSW plates.

The physical mechanism underlying the appearance of local minima in the dependence of the slowing-down coefficients for DSW waves on the thickness of the dielectric plate was very attentively studied in the course of analysis of the electromagnetic field distribution for DSW proper waves in the transmission-line cross section.

In the case of a thin dielectric substrate (Fig. 3), the propagation of four types of waves is possible. By analogy with waves corresponding to the dielectric plate, these waves can be classified as LM_1, LM_2, LM_3, and LM_4 quasi-waves. The last two waves are conditionally related to LM_3, and LM_4 quasi-waves, since, as the corresponding waves of a planar dielectric waveguide, they have the predominant component of the electric field, which is parallel to the plate plane. In addition, they have variations of the field direction along the normal to the dielectric plane, where the number of these variations corresponds to the accepted classification. However, this number of field-direction variations is by no means associated with the maintenance of the wave by the dielectric layer, especially for the LM_4 quasi-wave, since for a given geometry of the directing structure, the corresponding wave of the planar dielectric waveguide is cut off. In fact, these variations are explained by the possibility of propagating the H_1 wave of a planar waveguide, which is composed of a screen and a metal layer forming the slot aperture (regions *1* and *3* in Fig. 2). The main part of the energy transmitted by these waves is concentrated in hollow rectangular $b \times h_1$ and $b \times h_2$ waveguides (see insert in Fig. 3).

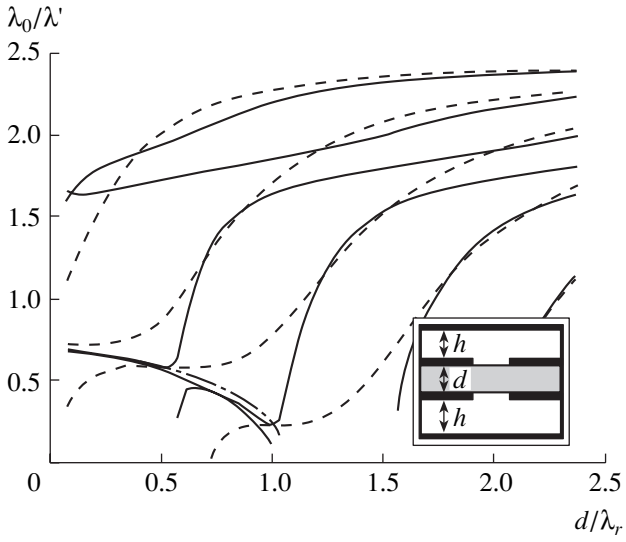


Fig. 3. Slowing down of DSW proper waves (solid lines) for a constant distance between screening metal layers and of LM_n waves for a dielectric plate (dotted lines) as a function of the dielectric-substrate thickness. The geometric parameters are the same in both cases: $f = 9.375$ GHz, $h_1 + d + h_2 = 4.6$ cm, $\varepsilon_r = 6.0$, $b = 1$ cm, and $w_1 = w_2 = 1$ mm.

In the connection region for the H_1 wave of the planar waveguide and the LM_3 wave of the screened dielectric layer, the slowing-down curve of the LM_3 wave has a local minimum. Here, the transformation of DSW-wave types occurs. Immediately at the critical point $d = 0.5$, the H_1 wave of a planar waveguide filled with a dielectric is disconnected from the LM_3 wave of this waveguide, whereas the effect of the H_1 wave of the planar dielectric waveguide on the LM_3 wave is already weakened. Therefore, at this point, the energy loss of the LM_3 quasi-wave, which is caused by the transformation of the wave energy into other wave types, exhibits a minimum. After the value of $\tilde{d}$ has exceeded its critical value, the mechanism of wave-energy maintenance is basically determined by the dielectric layer. Thus, in this case, the diffraction loss increases due to the connection of the LM_3 wave of the planar dielectric waveguide and the H_1 wave of the hollow planar waveguide.

Near the critical thickness, for the LM_2 wave of the screened planar dielectric waveguide, the possibility of the propagation of an actual LM_4 quasi-wave appears in the DSW. With increasing thickness of the dielectric plate, the slowing-down curves for the wave pass the coupling region for the LM_4 wave of the planar dielectric waveguide and the H_1 wave of the hollow planar waveguide.

The propagation mechanism for the LM_4 wave in a DSW immediately at the critical point $\tilde{d} = 1.0$ is similar to the above-described behavior of the LM_3 quasi-

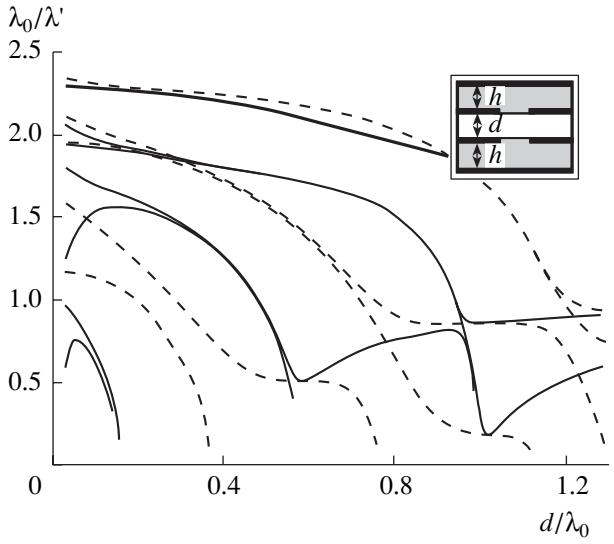


Fig. 4. Slowing down of proper waves for an inverted DSW (solid lines) as a function of the thickness of the air layer for a constant distance between screening metal layers and of LM_n waves (dotted lines) in a planar waveguide filled with a dielectric with a slot. The geometry parameters are the same in both cases: f = 9.375 GHz, $h_1 + d + h_2$ = 4.6 cm, ε_r = 6.0, b = 1 cm, $w_1 = w_2$ = 1 mm.

wave at the point $\tilde{d} = 0.5$ with an accuracy to the formal replacements $LM_3 \leftrightarrow LM_4$ and $H_1 \leftrightarrow H_2$. However, there exists one important distinction explaining the greater depth of the local minimum for slowing down of the LM_4 quasi-wave compared to the LM_3 quasi-wave. This distinction consists in the fact that for the LM_4 quasi-wave, two significant phenomena have approximately coincided at the same point. We imply both the end and the onset of the cutoff for the H_2 wave of the planar waveguide filled with a dielectric and for the H_1 wave of the planar hollow waveguide, respectively. This results not only in minimizing the coefficients for the transformation of the LM_4 quasi-wave into other waves of the planar dielectric waveguide and planar waveguide filled with a dielectric, but also in the disappearance of the energy loss caused by the transformation of the LM_4 quasi-wave into the H_1 wave of the shallow planar waveguide. In other words, in the case under consideration, the quality factor (Q-factor) for oscillations in the regions between DSW metallic planes increases and the impedance boundaries become less transparent [5]. After the quantity $\tilde{d}$ has exceeded its critical value, the diffraction energy loss will increase due to the connection of the LM_4 quasi-wave of the planar dielectric waveguide and H_2 quasi-wave of the planar waveguide.

3.2. Symmetric double slotted waveguide with different slots (Fig. 2, $s = 0$, $w_1 \neq w_2$). As a detailed analysis has shown in the resonance case, the slowing down of DSW proper waves of this type depends relatively weakly on the width of one of the slots, provided that the width of the other slot remains constant. For example, the variation of slowing-down reaches several percent when varying w_2 from w_1 to $w_2 = 2w_1$. The physical mechanism ensuring the propagation in this structure of wave types that exhibit a high value of the Q-factor is similar to that described in item 3.1. However, we should mention that when interpreting the results of numerical experiments, it is necessary to lead the discussion in terms of LM waves for a metal layer covered by a dielectric.

3.3. Inverted symmetric equally slotted double-slotted waveguide (Fig. 2, $s = 0$, $w_1 = w_2$, $\square_1 > \square_r$, $\square_2 > \square_r$). During the development of various basis elements for microwave and EHF SSIP modules based on 3D IC, inverted symmetric equally slotted ($w_1 = w_2$) and unequally slotted ($w_1 \neq w_2$) DSW are of interest in many cases and find wide application.

One example of the behavior of fast and slow waves for a symmetric inverted DSW is presented in Fig. 4. The slowing down of waves in a DSW and waves of a planar $\left(\frac{\partial}{\partial x} \equiv 0\right)$ waveguide (filled with a dielectric with a slot in it of the same width d as in the DSW) are shown by solid and dotted lines, respectively. The resonant behavior of the dependences $\eta = \eta\left(\frac{d\sqrt{\varepsilon}}{\lambda_0}\right)$ is clearly seen in Fig. 4. However, in contrast to a standard DSW (Fig. 3), the resonance points are slightly shifted from the points $\frac{nd}{\lambda_0}$, n = 1, 2, …. As is seen, with increasing n, the resonance points approach their standard values.

In this case, as in the case of a classical DSW (see Section 3.1), two key structures play the principal role in the formation of the repeated reflection of Brillouin waves. These are a hollow waveguide composed of metal layers that form slot apertures and a planar waveguide composed of a metal layer that forms a slot and the metal of a screen filled with a dielectric with relative permittivity ε_r.

A local minimum of the LM_5 quasi-wave near the value $\frac{d}{\lambda_r} = 0.5$ of the normalized thickness of the air layer is caused by the onset of cutoff for the H_2 wave of the planar waveguide. This waveguide is composed of a metal screen and metal forming a slot and filled with a dielectric. The dielectric layers play the main role in the process of maintaining the wave at this point of the dependence under consideration. The basic energy transferred by the electromagnetic wave is concentrated in regions *1* and *3* (see Fig. 2), which are filled with a dielectric with a large relative permittivity. Therefore, the passage of the air-layer thickness through the criti-

cal value for the H_1 wave of the planar waveguide seems to escape notice.

In the case presented in Fig. 4, the best quality resonance exists for the LM_4 quasi-wave in the vicinity of a point corresponding to the critical thickness $\frac{d}{\lambda_0} \approx 1.0$. This fact is explained by the approximate coincidence of the critical thickness of the air layer for the H_2 wave of a planar waveguide filled with air and the H_1 wave of a planar waveguide filled with a dielectric.

It is worth noting that, as in the case of a standard DSW (Fig. 3), resonance points for waves in this type of DSW coincide with plane segments of dependences for the corresponding waves of a planar waveguide with a slot in the dielectric (Fig. 4).

One more feature of the proper-wave spectrum confirms the fact that the inverted variant of the DSW is a rather interesting waveguide structure for 3D IC. We imply the proximity in values of the propagation constants for even and odd waves. It is well known that in order to develop strongly connected and oppositely directed couplers possessing a high directivity, it is necessary to equalize the propagation constants for even and odd waves. To this end, multilayer and anisotropic substrates are applied in transmission lines with front communication. The inverted DSW variant makes it possible to attain the desired result without special technological effort. This is due to the fact that quasi-even and quasi-odd waves of the inverted DSW are almost degenerate in the wide variation range of the air-layer thickness. The degeneration ceases in the case of a thin gap and in the case of a thick air layer.

ACKNOWLEDGMENTS

The authors are grateful to Yu.V. Gulyaev and M. Davidovich for their support of the present work.

REFERENCES

1. E. I. Nefedov, *The Modern Radio Electronics* (Nauka, Moscow, 1986).
2. *Basic Limits in Radio Electronics and Adjacent Fields.* Proceedings of the Institute of Electronic and Radio Engineers **69** (2), 3 (1981).
3. Yu. V. Gulyaev, B. V. Sandomirskiĭ, A. A. Sukhanov, *et al.*, Usp. Fiz. Nauk **114**, 475 (1984) [Sov. Phys. Usp. **114** (1984)].
4. Yu. V. Gulyaev, V. I. Gvozdev, and E. I. Nefedov, Dokl. Akad. Nauk SSSR **290**, 343 (1986) [Sov. Phys. Dokl. **31**, 760 (1986)].
5. E. I. Nefedov and A. T. Fialkovskiĭ, *Strip Transmission Lines: Electrodynamic Principles for Automated Design of Microwave-Range Integrated Circuits* (Nauka, Moscow, 1980).
6. V. I. Gvozdev and E. I. Nefedov, *3-dimensional Integrated Microwave Circuits* (Nauka, Moscow, 1985).
7. V. I. Nefedov, Dokl. Akad. Nauk **339**, 199 (1994) [Phys.–Dokl. **39**, 794 (1994)].
8. V. I. Nefedov, Radiotekh. Elektron. **38**, 593 (1993).
9. V. A. Neganov, E. I. Nefedov, and G. P. Yarovoĭ, *Strip–Slotted Lines for Microwave and Extremely High Frequencies* (Nauka, Moscow, 1996).
10. E. I. Nefedov, A. A. Protopopov, A. N. Sementsov, and A. A. Yashin, *Interaction of Physical Fields with Living Substances* (Izd. Tul'skogo Univ., Tula, 1995).
11. E. I. Nefedov, V. V. Kozlovskiĭ, and A. V. Zgurskiĭ, *Microstrip Radiating and Resonant Devices* (Tekhnika, Kiev, 1990).
12. E. I. Nefedov, A. S. Saidov, and A. R. Tagilaev, *Broadband Microstrip Control Microwave Devices* (Radio i Svyaz', Moscow, 1994).
13. E. I. Nefyodov and R. S. Popov, *Non-Symmetric Double-Sided Slotted Waveguides as a Basis Element of Three-Dimensional Integrated Circuits for Microwaves and EHF*, Proc. Int. URSI Symp. (Montreal, Canada, 1997), Vol. 1, p. 210.
14. B. Bhat and S. K. Koul, *Analysis, Design, and Application of Fin Lines* (Artech House, Norwood, 1987).
15. S. E. Bankov, V. F. Vzyatyshev, and E. V. Rodionova, Radiotekh. Electron. **33**, 2373 (1988).

Translated by G. Merzon

Doklady Physics, Vol. 47, No. 4, 2002, pp. 302–304. Translated from Doklady Akademii Nauk, Vol. 383, No. 6, 2002, pp. 761–763.

TECHNICAL PHYSICS

Structural Inhomogeneity and Magnetic Properties of Soft Magnetic Materials

Yu. N. Dragoshanskii*, V. V. Gubernatorov*, B. K. Sokolov*, and V. V. Ovchinnikov**

Presented by Academician V.P. Skripov, October 2, 2001

Received October 22, 2001

Soft magnetic materials are characterized by the following features: a small coercive force H_c; a large magnetic permeability μ and induction B; and a low total energy loss P_t for the magnetization reversal, which is composed of the hysteresis P_h and eddy-current P_e components. These magnetic properties of a material are determined by the type and sharpness of its crystal texture; the presence of contaminations, inclusions, and residual internal stresses in the material; and its chemical and structural heterogeneity.

In this paper, we studied the effect of structural inhomogeneity on the energy loss P_t for anisotropic electrical-sheet steel [Fe–Si(3 wt %) alloy with the rolling edging texture (110)[001]] and the finemet (Fe73,5Cu1Nb3Si13,5B9 alloy) as examples. In this case, we paid most attention to the effect on P_t of the magnetic-domain structural inhomogeneity caused by a change in the crystal structure of the material.

Many researchers share the opinion, expressed as far back as 1947, that the structural inhomogeneity of a ferromagnet causes extra magnetic loss [2]. Indeed, with increasing structural homogeneity of a material, the quantities μ and B increase, while H_c and P_h decrease [3]. However, in this case, magnetic domains with an anomalously large width are formed in the material, which results in an increase in P_e [4]. According to [5], the quantity P_e in a ferromagnet exceeds the conventional eddy-current loss P_0 evaluated under the conditions of uniform μ by several times: $P_e = P_0 \times 1.63D/d$, where d is the thickness of the ferromagnetic sample and D is the domain width. For present-day high-texture anisotropic electrical-sheet steels, $D \gg d$. Therefore, the contribution of P_e to P_t may be as much as 80% or even larger [4]. As the homogeneity of anisotropic electrical-sheet steel increases, the increase in P_e is much greater than the decrease in P_h, which leads to the increase in P_t. Because of this, reducing the magnetic-domain width (fragmentation) is an important problem in improving the properties of soft magnetic materials.

Two methods for lowering P_t by decreasing P_e have been developed at the Institute of Metal Physics, Ural Division, Russian Academy of Sciences. One of them is based on intentionally introducing local irregularities into the magnetic-domain structure by a mechanical method (e.g., by making scratches [6] and by local bending at a small angle [4]) or by short-time heating (e.g., by local laser irradiation [7]). The presence of narrow zones of these irregularities, which are spaced 4–5 mm apart and are located across an anisotropic electrical-sheet steel tape, can result in a decrease in P_t of 5–10%. This effect is associated with both a decrease in the domain width by a factor of 1.5–2 and the origination of extra wedge-shaped magnetic-reversal centers (see Fig. 1). In this case, μ and B slightly decrease, whereas H_c and P_h increase in the material.

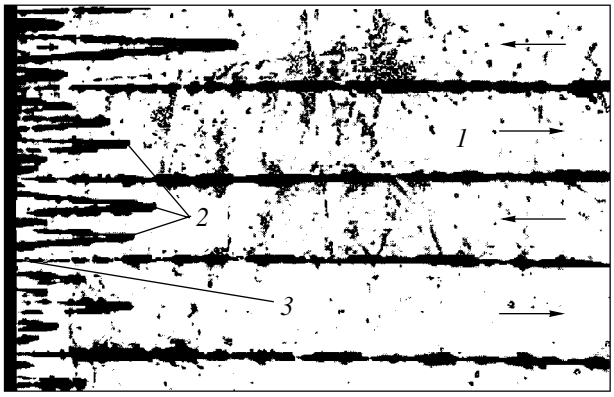


Fig. 1. Magnetic domains in anisotropic electrical-sheet steel [6]: (*1*) basic strip domains and (*2*) tapered closure domains, i.e., magnetic-reversal centers near (*3*) a scratch-like transverse irregularity (on the left).

* *Institute of Metal Physics, Ural Division, Russian Academy of Sciences, ul. S. Kovalevskoĭ 18, Yekaterinburg, 620219 Russia*

** *Institute of Electrophysics, Ural Division, Russian Academy of Sciences, ul. Komsomol'skaya 34, Yekaterinburg, 620219 Russia*

1028-3358/02/4704-0302$22.00

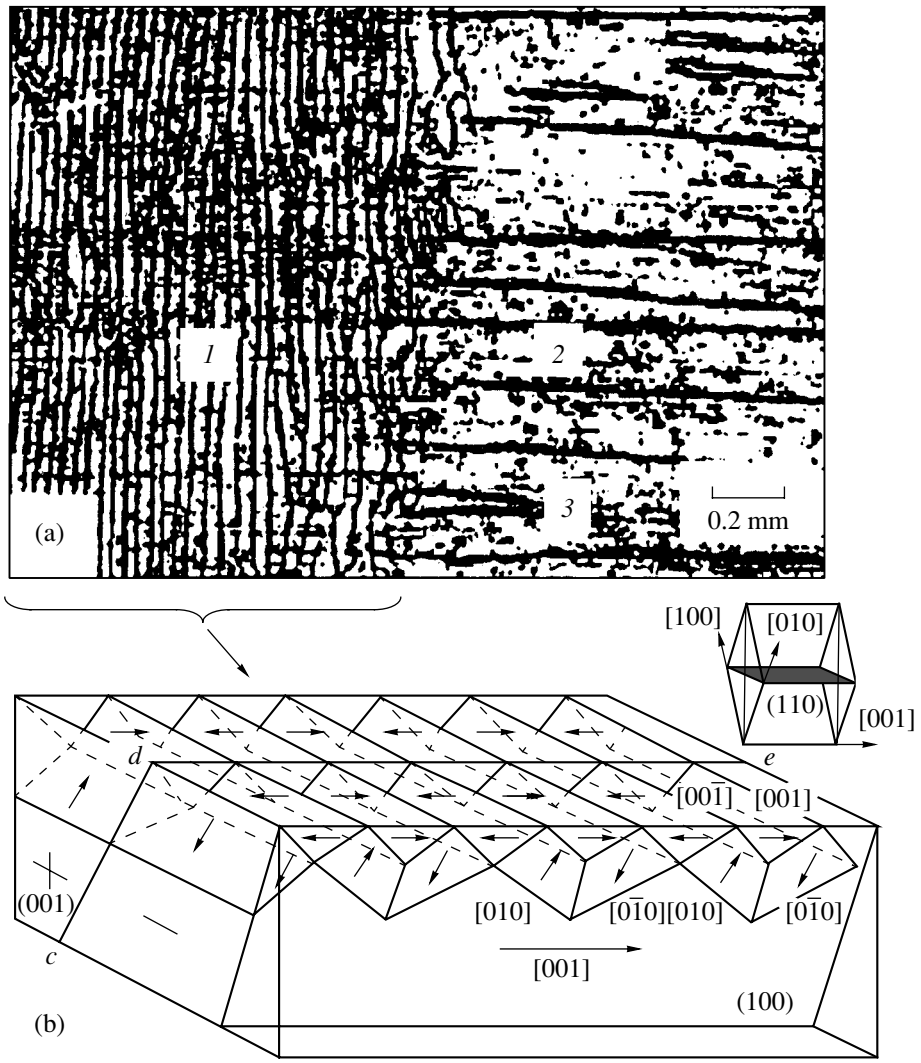


Fig. 2. Magnetic domain structure (a) on the surface and (b) in the bulk of a FeSi crystal after irradiation $(\bar{1}10)$ by argon ions [9]. The right-hand part of the surface was screened from the irradiation. Arrows indicate the magnetization directions in the domains: (*1*) labyrinth domains of the *B*-type, (*2*) basic strip domains of the *A*-type, and (*3*) dynamic magnetization-reversal centers.

The second method of lowering P_t consists in producing a layered inhomogeneity of the tape bulk, e.g., in the formation of a two-layer grain structure in the tape cross section [8], or by uniform ion-beam treatment of the tape surface [9]. The latter method is more efficient. In anisotropic electrical-sheet steels with a two-layer grain structure (where there are two grains rather than one grain in the tape cross section), the decrease in P_t is only 15–16%. In the case of finemets, the decrease in P_t reaches 35–40% after the ion-beam surface treatment in the optimum regime (with an argon-ion exposure dose of 10^{16} ion/cm^2) and 20–25% after treatment by the first method, with the quantity B remaining essentially invariable.

The uniform ion-beam surface treatment of a ferromagnetic alloy leads to restructuring of the magnetic domains in it. A magnetic structure with very narrow transversely oriented regions is formed on the alloy surface (see Fig. 2a), and the width of the basic intrabulk domains decreases. The magnetic-flux closure scheme for a ferromagnet with such a finely divided layered magnetic structure is shown in Fig. 2b.

For a layered magnetic structure, a decrease in P_t is caused by the fragmentation of basic domains and the presence of wedge-shaped magnetic-reversal centers (as in the case of local magnetic-domain structure irregularities), as well as by a change in the dynamics of domain-boundary displacements in a variable magnetic field. The latter effect leads to an additional decrease in P_e.[1] In particular, the decrease in bending of the

[1] In an alternating magnetic field, domain boundaries experience vibrations, translations, and bending with the formation of amplitude and phase shifts in the tape cross section [10]. These dynamic features determine both the mean velocities of motion of domain boundaries and the quantity P_e [11].

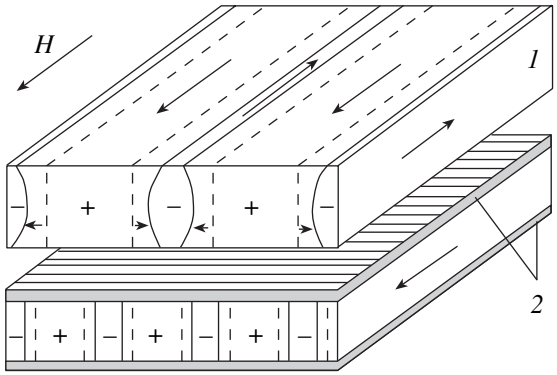


Fig. 3. Profiles of the planes of domain boundaries dislocated (under the action of a magnetic field) in the cross section of ferromagnetic tapes before (above) and after (below) the ion-beam treatment: (*1*) basic strip domains and (*2*) labyrinth domains. Dashed lines show initial positions of the domain boundaries.

domain-boundary planes, which is caused by their displacement, is retarded by the surface magnetic structure. Moreover, in this case, the effective thickness of the ferromagnetic sample decreases (see Fig. 3). In the interior of the material, where ion-beam implantation causes tensile stresses and leads to the elevation of the uniaxial magnetic anisotropy (in much the same way as the effect of magnetic coatings does [3]), the domain-boundary mobility increases under magnetization reversal. This flattening of moving domain boundary planes results in a decrease in the intensity of the most energy-consuming component of eddy currents, which is parallel to the tape plane [12].

Thus, by forming irregularities of a certain size and orientation in the crystalline and magnetic structures of a ferromagnet, we can control its magnetic properties and, in particular, efficiently reduce the energy loss for magnetization reversal in it.

ACKNOWLEDGMENTS

The authors are grateful to N.V. Gavrilov and B.Yu. Goloborodskiĭ for their help in performing the experiments.

This work was supported in part by the Russian Foundation for Basic Research, project nos. 02-02-16443 and 01-02-96407, and by the Government of Sverdlovsk oblast.

REFERENCES

1. S. V. Vonsovskiĭ, *Magnetism, Ser. Problems of Scientific and Technological Progress* (Nauka, Moscow, 1984).
2. R. I. Yanus and V. V. Druzhinin, Izv. Akad. Nauk SSSR, Ser. Fiz. **11**, 687 (1947).
3. Yu. N. Dragoshanskiĭ, V. V. Shulika, and A. P. Potapov, Dokl. Akad. Nauk **353**, 44 (1997) [Phys.–Dokl. **42**, 125 (1997)].
4. Yu. N. Dragoshanskii, E. V. Bratuseva, V. V. Gubernatorov, *et al.*, Phys. Met. Metallogr. **83**, 270 (1997).
5. R. H. Pry and C. P. Bean, J. Appl. Phys. **29**, 532 (1958).
6. Yu. N. Dragoshanskii, Candidate's Dissertation in Mathematical Physics (Sverdlovsk, 1968).
7. B. K. Sokolov, V. V. Gubernatorov, M. M. Noskov, *et al.*, Pat. SSSR No. 652230, Byull. Izobret., No. 10, 115 (1979).
8. S. V. Zhakov, Yu. N. Dragoshanskii, B. K. Sokolov, *et al.*, Fiz. Met. Metalloved. **63**, 502 (1987).
9. B. K. Sokolov, V. V. Gubernatorov, Yu. N. Dragoshanskii, *et al.*, Phys. Met. Metallogr. **89** (4), 32 (2000).
10. Yu. N. Dragoshanskii and A. Yu. Dragoshanskii, *Nonlinear Electromagnetic Systems* (Elsevier Press, Amsterdam, 1998).
11. Yu. N. Dragoshanskii, Phys. Met. Metallogr. **89**, Suppl. 1, S73 (2000).
12. M. Imamura, T. Sasaki, and T. Yamaguchi, IEEE Trans. Mag. **20**, 2120 (1984).

Translated by V. Chechin

Doklady Physics, Vol. 47, No. 4, 2002, pp. 305–307. Translated from Doklady Akademii Nauk, Vol. 383, No. 6, 2002, pp. 764–766.

TECHNICAL PHYSICS

Mechanical Activation of Combustion Processes in Certain Low-Energy Systems

A. Yu. Postnikov, P. I. Gavrilov, E. V. Levakov, A. Ya. Malyshev, A. V. Sten'gach, and I. K. Kremzukov

Presented by Academician V.V. Osiko November 27, 2001

Received November 8, 2001

It is well known that when the thermal effect of a combustion reaction is small, the combustion wave in the system sometimes does not propagate [1]. For example, the synthesis of certain high-melting compounds from simple species is weakly exothermic. Therefore, in order to produce these compounds in the self-propagating high-temperature synthesis (SHS) regime, various approaches are employed, namely, heating of the reacting batch [2], performing parallel SHS-processes [3], and introducing various additives [4]. The mechanical activation of solids is a well-known approach to enhancing their reactivity [5]. It is noted that the combustion rate, combustion temperature, and the reaction-product yield are elevated with increasing preliminary treatment time of an SHS batch, [6]. In a number of low-energy systems, combustion is possible only after preliminary mechanical activation of the system components. In this case, a change in the reactivity of solids can be a consequence of both the accumulation of microdistortions in crystallites and the variation of particle sizes.

As was shown in [7] using the Ti–C (carbon black) system as an example, the elevation of the combustion rate on increasing the duration of the mixture treatment is mainly associated with the mechanical activation of the titanium powder. To determine the relation between activation of the combustion process and structural changes in the metal lattice, which had been observed under mechanical treatment, measurements of exoelectron emission from titanium powders were carried out. The results obtained [7] show a direct dependence between the exoelectron-emission intensity and the metal chemical activity. However, it should be taken into account that the thickness of the solid layers from which the exoelectron emission occurred was about 100 Å [8]. Therefore, the information obtained by exoelectron-emission methods relates only to the state of the powder surface [7].

All-Russia Research Institute of Experimental Physics, Sarov, Nizhni Novgorod oblast, Russia

In this paper, we suggest using X-ray diffraction analysis to study and control the suitability of the SHS batch in the process of its mechanical activation. It is known [9] that plastic deformation of a crystal results in the grinding of mosaic blocks, which is followed by the appearance of microstresses. They, in turn, are a cause of the increase in the integral line width (B) in the X-ray diffraction pattern. The magnitude of B can be used as one of the objective criteria in estimating SHS-batch suitability.

The studies we performed also showed that in the case of long-term mechanical treatment of initial components, the SHS process is possible even in such low-energy systems as Si–C and W–B.

The initial batch was prepared by mixing the initial components in a ball mill in the nitrogen ambient. W(PWT), B(B99A), C(P234), and Si(KR00) powders were used to prepare the reaction mixture. The samples intended for combustion were pressed in cylindrical steel pots 20 mm in diameter. The combustion reaction was initiated by a combustible compound, which, in turn, had been ignited by a nichrome heating bridge. The average combustion rate was determined according to the readings of tungsten–rhenium thermocouples pressed into the sample at a fixed distance from each other. X-ray analysis was carried out with DRON-type equipment.

The variation of the combustion rate U, integral line width B, and line intensity I on the X-ray diffraction pattern, as well as the specific surface S of the mixture as a function of the mechanical-activation time τ, is presented in Figs. 1 and 2. At the initial stage, a progressive increase in the specific surface of the Si–C system was observed (Fig. 1). After 50 h of mixture treatment, the growth rate of the specific surface significantly decreased, which could be a result of the action of two opposite processes. These are a decrease in particle sizes during fracturing and the production of aggregated particles [10]. The integral line width monotonically increases with the mechanical-activation time of the preliminary mixture treatment during the entire time period under study (Figs. 1, 2). The elevation of the combustion rate of the Si–C mixture, which was

1028-3358/02/4704-0305$22.00

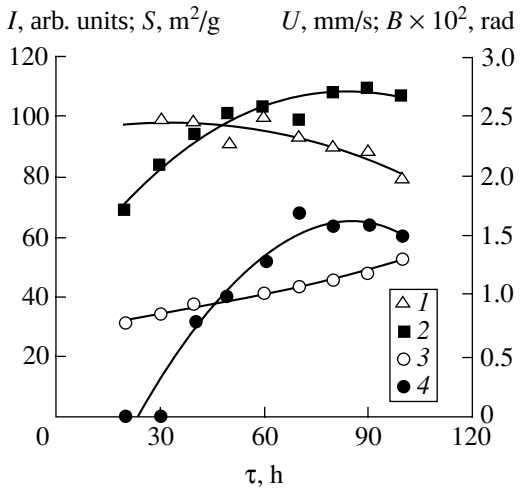


Fig. 1. (*1*) Intensity of lines for silicon, (*2*) specific surface of the mixture, (*3*) integral line width, (*4*) combustion rate for the Si–C composition as a function of the mixing time τ.

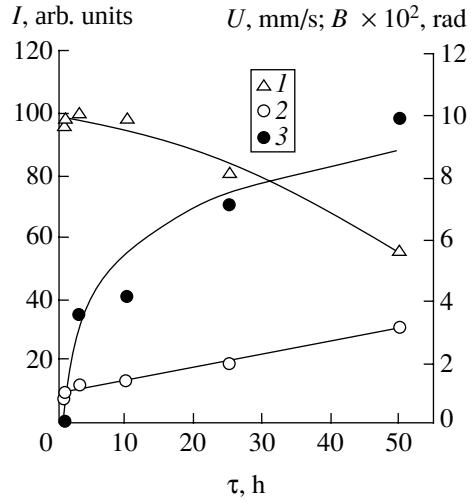


Fig. 2. (*1*) Intensity of lines for tungsten, (*2*) integral line width, (*3*) combustion rate for the W–B composition as a function of the mixing time τ.

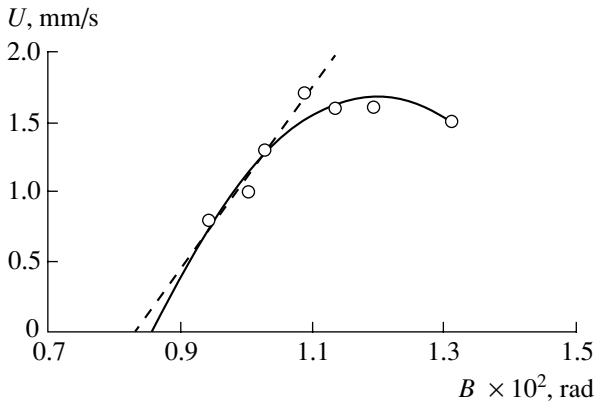


Fig. 3. Dependence of the combustion rate for the Si–C composition on the integral line width of silicon. Solid and dashed lines are plotted with and without allowance for mechanochemical activation, respectively.

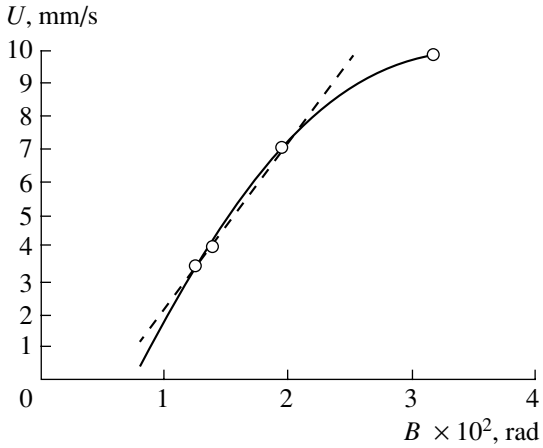


Fig. 4. Dependence of the combustion rate for the W–B composition on the integral line width of tungsten. Solid and dashed lines are plotted with and without allowance for mechanochemical activation, respectively.

observed while increasing the mechanical-activation time to 70 h, as well as the behavior of the curves for both the specific surface and the variation of the integral line width, show that, at the initial period of the mechanical-activation time, the change in the reactivity of the system is caused by an increase in the surface. Furthermore, the growth of microstresses in the crystal lattice can acquire a substantial significance. The data obtained agree well with the model for the state of plastic crystal flow. Plastic flow is the cause of particle aggregation and is responsible for the intensely proceeding process of defect formation [9].

With continuous elevation of the combustion rate, the intensity of lines on the X-ray diffraction patterns remains almost unchanged for both the Si–C and the W–B mixtures in the process of mechanical activation up to 70 h for the former (Fig. 1) and up to 10 h for the latter (Fig. 2). With a further increase in the activation time of the mixtures, the line intensity decreases, the decrease being sharper for the W–B system (Figs. 1, 2). This fact is evidence of both amorphization of the components and the formation of products as a result of mechanochemical reactions. In this case, both the initial components and the reaction products can be subjected to amorphization. With increasing mechanical-activation time, the combustion rate for the Si–C system grows linearly up to 70 h of mixture treatment, after which the combustion rate drops (Fig. 1). This fact, the X-ray phase analysis of slags, and additional calorimetric measurements testify to the formation of an amorphous product in the process of long-term mechanical activation of the Si–C mixture (after 70 h of treatment). For the W–B system, the combustion rate increases, as before, under the intensely proceeding

amorphization of the sample (20–50 h of mechanical activation, Fig. 2). This indicates that at this stage of mechanical activation, mechanochemical processes have a weaker effect than in the Si–C system within the range of 90–100 h of treatment of the mixture. However, in the W–B system, we also note a certain slowing down of the increase in the combustion rate at the end of the region of its dependence on the mechanical-activation time as a possible manifestation of the effect of development of mechanochemical processes.

Using the combustion rate as the measure of the reactivity of the SHS-batch components, we present it as a function of the integral line width for one of the components (Figs. 3, 4). It is worthwhile noting that the mechanical activation can be a cause of the change in the SHS-batch composition as a result of the progress of a mechanochemical reaction. In this case, dilution of the batch by the products of mechanochemical reactions promotes the reduction of the combustion rate.

Thus, if we exclude the regions corresponding to the dependence of the combustion rate on the integral line width (solid lines in Figs. 3 and 4) in the curves, which correspond to mechanochemical reactions proceeding in a condensed system as a result of mechanical activation, we arrive at a linear dependence (dashed line, Figs. 3, 4). The employment of this dependence allows us to predict the kinetics of solid-phase interactions of components in the combustion regime. This also makes it possible to determine the duration of mechanical activation necessary to stimulate the SHS process in a low-energy system. The importance of the correlation dependence obtained also consists in the possibility of quality control of SHS-batch suitability at its preparation stage. As was noted above, the integral line width is an objective criterion of SHS-batch suitability. For example, in contrast to mechanical activation, (whose procedure time exhibits a similar relation to the combustion rate), X-ray analysis eliminates mistakes and random errors at the mixture preparation stage.

We conclude that the magnitude of the integral line width can serve as a quantitative measure for characterizing the degree of mechanical activation and, correspondingly, the reactivity of solids as applied to SHS processes. However, we should remember that the linear dependences obtained relate to systems in which the basic reaction proceeds in the condensed phase. Therefore, the study of the dependences described in systems that differ in their mechanism of component interaction is very useful.

REFERENCES

1. A. G. Merzhanov, in *Combustion Processes and Synthesis of Materials* (Chernogolovka, 1998), pp. 214–216.
2. Yu. S. Naĭborodenko, V. I. Itin, A. G. Merzhanov, *et al.*, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 6, 145 (1973).
3. I. P. Borovinskaya, Pure Appl. Chem. **64**, 919 (1992).
4. I. P. Novikov, I. P. Borovinskaya, and V. V. Boldyrev, Fiz. Gorenia Vzryva, No. 3, 342 (1977).
5. V. V. Aleksandrov and M. A. Korchagin, Int. J. Self-Propag. High-Temp. Synth. **1**, 417 (1992).
6. A. P. Borisov, V. D. Makhaev, V. V. Aleshin, and L. A. Petrova, in *Transactions of the 21st International Pyrotechnical Seminar, Moscow, Russia, 1995* (Inst. Chem. Phys. Publ., Moscow, 1995), p. 70.
7. S. I. Voyuev, V. M. Maslov, S. V. Popetaev, *et al.*, Fiz. Gorenia Vzryva, No. 3, 18 (1983).
8. R. I. Mints, I. I. Mil'man, and V. I. Kryuk, Usp. Fiz. Nauk **119**, 4 (1976) [Sov. Phys. Usp. **19**, 697 (1976)].
9. N. N. Kachanov and L. I. Mirkin, *The X-ray Structural Analysis of Polycrystals. Practical Guide* (Mashgiz, Moscow, 1960).
10. E. G. Avvakumov, *Mechanical Methods of Activating Chemical Processes* (Nauka, Novosibirsk, 1986).

Translated by T. Galkina

Doklady Physics, Vol. 47, No. 4, 2002, pp. 308–311. Translated from Doklady Akademii Nauk, Vol. 383, No. 4, 2002, pp. 492–495.

MECHANICS

Singular Solutions for Plane Plastic Flow of Pressure-Dependent Materials

S. E. Alexandrov and E. A. Lyamina

Presented by Academician A.Yu. Ishlinskii, September 17, 2001

Received September 25, 2001

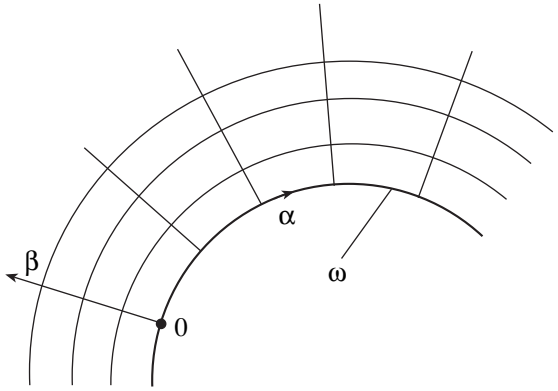
Models of plastically incompressible materials with a pressure-dependent yield criterion are used mainly in the mechanics of soil and granular materials [1, 2]. However, for a group of alloys, the pressure dependence of the yield criterion is also of certain importance [3, 4]. In contrast to soil, these alloys are usually subjected to plastic metal forming based on the frictional law at sliding. In the present paper, we show that in the case where model [2] and the maximum-friction law are applied, the velocity field becomes singular (the equivalent strain rate approaches infinity at the friction surface). Note that the maximum-friction law is often used in modeling plastic metal-forming processes [5]. In the framework of the classical plasticity model, a similar result was obtained in [6–8]. In [6] and [7], plane flow and three-dimensional deformation of a material obeying an arbitrary smooth-yield criterion independent of the mean stress, respectively, were considered. In [8], the axially symmetric deformation of a material obeying the Tresca plastic-flow condition was analyzed. In all of these cases, the equivalent strain rate was shown to be inversely proportional to the square root of the distance from the friction surface. Knowledge of the singular behavior of the velocity field made it possible to improve the solutions to problems of structural [9] and technological [10] plasticity that were based on the upper-bound estimates. In addition, the equivalent strain rate enters into the equations determining many physical fields and structural properties of materials. In particular, the effect of the maximum-friction surface on the temperature field was studied in [11]. It is also well known [12] that surface properties of a material change significantly in the deformation process. This is in qualitative agreement with the fact that the equivalent strain rate approaches infinity at the friction surface, although no quantitative estimates for the effect of such behavior of the velocity field on the structure of materials were obtained.

Two particular solutions based on model [2] show that the velocity fields for such a material can also be singular in the vicinity of the maximum-friction surfaces. These solutions are found in [13], where material flow through a planar converging channel was considered, and in [14], where plane strain compression between two parallel rough plates was studied. In the present paper, general conditions are found whose validity provides the singular velocity fields in the case of plane flow of a material [2].

Following [6], we introduce a curvilinear orthogonal coordinate system α, β such that the β-coordinate lines are straight and orthogonal to the friction surface ω (Fig. 1) for which $\beta = 0$. In this coordinate system, the equilibrium equations have the form [6]

$$\begin{aligned} \frac{\partial \sigma_{\alpha\alpha}}{\partial \alpha} + H\frac{\partial \tau_{\alpha\beta}}{\partial \beta} + \frac{2\tau_{\alpha\beta}}{R} &= 0, \\ \frac{\partial \tau_{\alpha\beta}}{\partial \alpha} + H\frac{\partial \sigma_{\beta\beta}}{\partial \beta} + \frac{\sigma_{\beta\beta} - \sigma_{\alpha\alpha}}{R} &= 0, \end{aligned} \tag{1}$$

where $\sigma_{\alpha\alpha}$, $\sigma_{\beta\beta}$, $\tau_{\alpha\beta}$ are the stress-tensor components, R is the radius of curvature for the α-coordinate line at $\beta = 0$, and $H = 1 + \dfrac{\beta}{R(\alpha)}$ is the Lamé coefficient for the α-coordinate lines.



Coordinate system in the vicinity of the friction surface.

Institute for Problems in Mechanics,
Russian Academy of Sciences,
pr. Vernadskogo 101, Moscow, 117526 Russia

1028-3358/02/4704-0308$22.00

Under the plane-strain conditions, the model of a material that was proposed in [2] is described by the equations

$$(\sigma_{\alpha\alpha} + \sigma_{\beta\beta})\sin\varphi + [(\sigma_{\alpha\alpha} - \sigma_{\beta\beta})^2 + 4\tau_{\alpha\beta}^2]^{1/2} = 2k\cos\varphi, \quad (2)$$

$$\xi_{\alpha\alpha} + \xi_{\beta\beta} = 0, \quad (3)$$

$$\sin 2\psi(\xi_{\alpha\alpha} - \xi_{\beta\beta}) - 2\cos 2\psi\xi_{\alpha\beta} + 2\sin\varphi\left(\omega_{\alpha\beta} - \frac{d\psi}{dt}\right) = 0 \quad (4)$$

for an arbitrary orthogonal coordinate system, including the α, β coordinate system introduced above. Here $\xi_{\alpha\alpha}$, $\xi_{\beta\beta}$, $\xi_{\alpha\beta}$ are the strain-rate tensor components; $\omega_{\alpha\beta}$ is the rotation-strain tensor component; ψ is the angle between the largest principal stress and the tangent to the α-coordinate line; φ is the angle of internal friction; and k is the cohesion coefficient. The yield criterion (2) is satisfied by the substitution

$$\sigma_{\alpha\alpha} = \sigma + (k\cos\varphi - \sigma\sin\varphi)\cos 2\psi,$$
$$\sigma_{\beta\beta} = \sigma - (k\cos\varphi - \sigma\sin\varphi)\cos 2\psi, \quad (5)$$
$$\tau_{\alpha\beta} = (k\cos\varphi - \sigma\sin\varphi)\sin 2\psi,$$

where $\sigma = \frac{1}{2}(\sigma_{\alpha\alpha} + \sigma_{\beta\beta})$ is the mean stress. Substituting (5) into the equilibrium equations (1), we arrive at

$$(1 - \sin\varphi\cos 2\psi)\frac{\partial\sigma}{\partial\alpha} - 2\sin 2\psi(k\cos\varphi - \sigma\sin\varphi)\frac{\partial\psi}{\partial\alpha} - H\sin\varphi\sin 2\psi\frac{\partial\sigma}{\partial\beta} + 2H\cos 2\psi(k\cos\varphi - \sigma\sin\varphi)\frac{\partial\psi}{\partial\beta} + \frac{2}{R}\sin 2\psi(k\cos\varphi - \sigma\sin\varphi) = 0,$$

$$-\sin\varphi\sin 2\psi\frac{\partial\sigma}{\partial\alpha} + 2\cos 2\psi(k\cos\varphi - \sigma\sin\varphi)\frac{\partial\psi}{\partial\alpha} + H(1 + \sin\varphi\cos 2\psi)\frac{\partial\sigma}{\partial\beta} + 2H\sin 2\psi(k\cos\varphi - \sigma\sin\varphi)\frac{\partial\psi}{\partial\beta} - \frac{2}{R}\cos 2\psi(k\cos\varphi - \sigma\sin\varphi) = 0.$$

After certain algebraic transformations, these equations can be reduced to the form

$$(\cos 2\psi - \sin\varphi)\frac{\partial\sigma}{\partial\alpha} + H\sin 2\psi\frac{\partial\sigma}{\partial\beta} + 2H(k\cos\varphi - \sigma\sin\varphi)\frac{\partial\psi}{\partial\beta} = 0, \quad (6)$$

$$\sin 2\psi\frac{\partial\sigma}{\partial\alpha} - H(\cos 2\psi + \sin\varphi)\frac{\partial\sigma}{\partial\beta} - 2(k\cos\varphi - \sigma\sin\varphi)\frac{\partial\psi}{\partial\alpha} + \frac{2}{R}(k\cos\varphi - \sigma\sin\varphi) = 0. \quad (7)$$

In the chosen α, β coordinate system, the strain-rate tensor components and the rotation-tensor components are given by the following expressions:

$$\xi_{\alpha\alpha} = \frac{1}{H}\left(\frac{\partial u}{\partial\alpha} + \frac{v}{R}\right), \quad \xi_{\beta\beta} = \frac{\partial v}{\partial\beta},$$
$$\xi_{\alpha\beta} = \frac{1}{2}\left[\frac{1}{H}\left(\frac{\partial v}{\partial\alpha} - \frac{u}{R}\right) + \frac{\partial u}{\partial\beta}\right], \quad (8)$$
$$\omega_{\alpha\beta} = \frac{1}{2}\left[\frac{1}{H}\left(\frac{\partial v}{\partial\alpha} - \frac{u}{R}\right) - \frac{\partial u}{\partial\beta}\right].$$

Substituting these expressions into (3) and (4), we obtain

$$\frac{1}{H}\left(\frac{\partial u}{\partial\alpha} + \frac{v}{R}\right) + \frac{\partial v}{\partial\beta} = 0, \quad (9)$$

$$\frac{\sin 2\psi}{H}\frac{\partial u}{\partial\alpha} - \frac{\cos 2\psi + \sin\varphi}{2}\frac{\partial u}{\partial\beta} + \frac{1}{2H}(\sin\varphi - \cos 2\psi)\frac{\partial v}{\partial\alpha} + \frac{\sin 2\psi}{H}\frac{v}{R} + \frac{1}{2H}(\cos 2\psi - \sin\varphi)\frac{u}{R} - \sin\varphi\frac{d\psi}{dt} = 0. \quad (10)$$

On the friction surface, the boundary conditions allow for the friction law and the impenetrability condition. Since the tool employed is assumed to be absolutely rigid, the friction surface can be regarded as immobile. Therefore, in the case of $\beta = 0$, the impenetrability condition yields

$$v = 0. \quad (11)$$

We also assume that the function v obeys the condition

$$\frac{\partial v}{\partial\alpha} = 0 \quad (12)$$

for $\beta = 0$. In classical plasticity theory, the maximum-friction law implies that friction stress with sliding is equal to the shear yield stress. For pressure-dependent materials, various modifications of this law are possible. Following [13, 14] and assuming, without loss of generality, that the direction of sliding is opposite to the positive direction of the α-coordinate line, we consider that

$$\psi = \psi_W = \frac{\pi}{4} + \frac{\varphi}{2} \quad (13)$$

for $\beta = 0$. This law corresponds to the maximum possible friction stress.

We now consider the behavior of the solution in the vicinity of the surface $\beta = 0$. In accordance with (11), an arbitrary material point located at an arbitrary instant on the friction surface can move only along this surface. Since condition (13) is satisfied at any point on the friction surface, we have

$$\frac{d\psi}{dt} = 0 \tag{14}$$

for any material point on this surface. Moreover, it follows from (13) that

$$(\cos 2\psi - \sin\varphi)\frac{u}{2HR} = -\frac{\sin\varphi}{R}u \tag{15}$$

on the friction surface. Substituting expressions (11), (12), (14), and (15) into Eq. (10), we arrive at

$$\cos\varphi\frac{\partial u}{\partial\alpha} - \frac{\sin\varphi}{R}u - \frac{(\cos 2\psi + \sin\varphi)}{2}\frac{\partial u}{\partial\beta} = 0 \tag{16}$$

for $\psi = \frac{\pi}{4} + \frac{\pi}{2}$ (or $\beta = 0$). If $\left|\frac{\partial u}{\partial\beta}\right| < \infty$, it follows from (16) that $\cos\varphi\frac{\partial u}{\partial\alpha} - u\frac{\sin\varphi}{R} = 0$ for $\psi = \frac{\pi}{4} + \frac{\varphi}{2}$. Thus, the distribution of the quantity u on the friction surface is determined by its shape and is independent of other reasons. It is clear that such a special distribution of u cannot occur in most cases as, e.g., for the solutions found in [13, 14]. Therefore, it is of interest to analyze the case $\frac{\partial u}{\partial\alpha} \neq 0$. Then, it follows from Eq. (16) that

$$\frac{\partial u}{\partial\beta} \longrightarrow \infty \tag{17}$$

for $\psi = \frac{\pi}{4} + \frac{\varphi}{2}$. Thus, the velocity field in such solutions must be singular.

We also consider the behavior of the solution in the vicinity of the friction surface. To this end, we represent u in the form

$$u = u_0 + O(\beta^\gamma), \tag{18}$$

where $\gamma \geq 0$ as $|u| < \infty$. It follows from (18) that

$$\frac{\partial u}{\partial\beta} = O(\beta^{\gamma-1}). \tag{19}$$

To satisfy (17), it is necessary to put $\gamma < 1$. Thus, we obtain

$$0 \leq \gamma < 1. \tag{20}$$

The expansion of the expression $(\cos 2\psi + \sin\varphi)$ in the Taylor series in the vicinity of the point $\psi = \frac{\pi}{4} + \frac{\varphi}{2}$ has the form

$$\cos 2\psi + \sin\varphi = 2\cos\varphi\left(\frac{\pi}{4} + \frac{\varphi}{2} - \psi\right) + O\left[\left(\frac{\pi}{4} + \frac{\varphi}{2} - \psi\right)^2\right]. \tag{21}$$

Since the first and second terms in Eq. (16) are finite, substituting (19) and (20) into the third term of Eq. (16) yields

$$2\cos\varphi\left(\frac{\pi}{4} + \frac{\varphi}{2} - \psi\right)O(\beta^{\gamma-1}) = O(1).$$

Thus, in the vicinity of the friction surface,

$$\frac{\pi}{4} + \frac{\varphi}{2} - \psi = O(\beta^{1-\gamma}). \tag{22}$$

It follows from expressions (20) and (22) that the derivative $\frac{\partial\psi}{\partial\beta}$ approaches infinity as $\beta \longrightarrow 0$. In Eq. (6), apart from this derivative, only the derivative $\frac{\partial\sigma}{\partial\beta}$ can approach infinity. It is clear that

$$O\left(\frac{\partial\sigma}{\partial\beta}\right) = O\left(\frac{\partial\psi}{\partial\beta}\right) = O(\beta^{-\gamma}), \tag{23}$$

where Eq. (22) is taken into account. It follows from (23) that

$$\sigma = O(\beta^{1-\gamma}). \tag{24}$$

Using Eqs. (21)–(24), it is possible to evaluate the order of each term in Eq. (7). As a result, we obtain

$$O(\beta^{1-\gamma}) + O(\beta^{1-2\gamma}) + O(1) + O(\beta^{1-\gamma}) + O[\beta^{2(1-\gamma)}] = 0.$$

By virtue of condition (20), the only possibility to compensate the term $O(1)$ in this equation is to assume that

$$\gamma = \frac{1}{2}. \tag{25}$$

It follows from conditions (19) and (25) that, in the vicinity of the maximum-friction surface, the shear strain rate and therefore the equivalent strain rate ξ_{eq} are given by the expressions

$$\xi_{\alpha\beta} = O(\beta^{-1/2}), \quad \xi_{eq} = O(\beta^{-1/2}).$$

This result coincides with the solutions to the particular problems obtained in [13, 14] and with the behavior of ξ_{eq} in classical plasticity [7]. In addition,

expressions (22), (24) and (25) make it possible to find that

$$\frac{\pi}{4} + \frac{\varphi}{2} - \psi = O(\beta^{1/2}), \quad \sigma = O(\beta^{1/2}).$$

It is worth noting that the solution for stresses in the vicinity of the coordinate line $\beta = 0$ was studied in [1].

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 01-01-06118.

REFERENCES

1. V. V. Sokolovskiĭ, *Statics of a Granular Medium* (Fizmatgiz, Leningrad, 1960).
2. A. J. M. Spencer, in *Mechanics of Solids. The Rodney Hill 69th Anniversary Volume* (Pergamon, Oxford, 1982), pp. 607–652.
3. W. A. Spitzig, R. J. Sober, and O. Richmond, Metallurg. Trans. **7A**, 1703 (1976).
4. A. S. Kao, H. A. Kuhn, W. A. Spitzig, and O. Richmond, Trans. ASME J. Eng. Mater. Technol. **112**, 26 (1990).
5. S. Kobayashi, S.-I. Oh, and T. Altan, *Metal Forming and the Finite-Element Method* (Oxford Univ. Press, New York, 1989).
6. V. V. Sokolovskiĭ, Prikl. Mat. Mekh. **20**, 328 (1956).
7. S. Alexandrov and O. Richmond, Int. J. Nonlinear Mech. **36**, 1 (2001).
8. S. E. Aleksandrov and O. Richmond, Dokl. Akad. Nauk **360**, 480 (1998) [Phys.–Dokl. **43**, 105 (1998)].
9. S. Alexandrov and O. Richmond, Int. J. Solids Struct. **37**, 669 (2000).
10. S. Alexandrov, G. Mishuris, W. Mishuris, and R. E. Sliwa, Int. J. Mech. Sci. **43**, 367 (2001).
11. S. Alexandrov and O. Richmond, in *Proceedings 3rd Int. Congress on Thermal Stresses, Cracow, Poland, 1999* (Cracow Univ. of Technology, Cracow, 1999), p. 153.
12. T. Aukrust and S. LaZghab, Int. J. Plast. **16**, 59 (2000).
13. C. S. Pemberton, J. Mech. Phys. Solids **13**, 351 (1965).
14. E. A. Marshall, Acta Mech. **3**, 82 (1967).

Translated by S. Alexandrov

Doklady Physics, Vol. 47, No. 4, 2002, pp. 312–315. Translated from Doklady Akademii Nauk, Vol. 383, No. 4, 2002, pp. 496–499.

MECHANICS

On the Formation of Loops in Ropes or Filaments

M. V. Los'* and A. E. Ordanovich**

Presented by Academician A.Yu. Ishlinskiĭ December 4, 2001

Received December 7, 2001

Investigation of the formation of loops (usually undesirable) in ropes and filaments is of considerable practical and scientific interest. A suitable model for theoretically studying this phenomenon is a thin elastic rod that has ends constrained in the undeformed rod axis and is subjected to a compressing force and a torsion torque.

Previously, a simpler methodological problem of forming a loop in a thin elastic rod with simply supported ends was investigated. The result of these studies is briefly as follows. When such a rod is subjected to an external force Q^0 directed along the undeformed axis, the straight rod loses its stability under certain conditions and its ends approach each other. In this case, a rod subjected to sufficiently strong forces takes a loop-like shape (Euler elastics) and is stable against small disturbances. However, in order for the rod to take this configuration, the displacement of the rod ends must be commensurable with the undeformed rod length. This situation is rarely realized in practice. As a rule, the formation of loops is attributed to twisting of the rod. Indeed, on being additionally exposed to an external moment M^0 directed along the original axis, the rod takes a spiral shape [1, 2]. However, as was ascertained, the new spatial configuration is also stable against small disturbances and loops are not formed. Previously, we showed that, in order to form a loop, the rod ends must be forced to approach each other at a given external moment M^0 [3–6]. In this case, the spiral configuration loses stability and can form a loop. Thus, to find the conditions of loop formation, it is necessary to consider a new mathematical problem, where the determining parameters are an external moment M^0 and a distance z^0 between the rod ends.

In this study, qualitative results obtained previously are generalized for a rod with ends constrained in the undeformed rod axis. This problem is closer to actual conditions.

Mathematically, we formulate the problem as follows. We analyze the static configuration and stability of a thin, originally rectilinear rod that has a length L and ends constrained in the undeformed rod axis. The rod is subjected to a force Q^0 and torque M^0 acting along the undeformed rod axis. Under certain conditions, the rod takes a three-dimensional shape and the rod ends, which remain in the undeformed rod axis and conserve their orientation, can approach each other. We consider the Cartesian coordinate system $OXYZ$ whose OZ axis is directed along the undeformed rod axis and whose origin O coincides with the left-hand rod end. The arc length s of the axis line is measured from the origin. It is assumed that the rod is inextensible and axisymmetric and has stiffness A and density ρS per unit length. The rotary inertia is ignored. The equations of motion of the thin-rod axis have the form [3–6]

$$\frac{\partial \mathbf{Q}}{\partial s} = \frac{\partial^2 \mathbf{r}}{\partial t^2}, \quad \frac{\partial \mathbf{M}}{\partial s} = [\mathbf{Q} \times \boldsymbol{\tau}], \quad \frac{\partial \boldsymbol{\tau}}{\partial s} = [\mathbf{M} \times \boldsymbol{\tau}], \quad \boldsymbol{\tau} = \frac{\partial \mathbf{r}}{\partial s}, \tag{1}$$

where t is the time, $\mathbf{Q} = (Q_1, Q_2, Q_3)$ is the force vector, $\mathbf{M} = (M_1, M_2, M_3)$ is the internal moment vector, $\mathbf{r} = (x, y, z)$ is the radius vector of the point s, and $\boldsymbol{\tau} = (l_{11}, l_{12}, l_{13})$ is the unit tangent vector to the axis line of the rod. Equations (1) are written in dimensionless form, where distances, forces, moments, and times are measured in the rod length L, $Q_m = \frac{A}{L^2}$, $M_m = \frac{A}{L}$, and $T_m = \left(\rho \frac{SL^4}{A}\right)^{1/2}$, respectively.

Two types of boundary conditions are considered.

1. The classical formulation of the problem implies that $\mathbf{r}(0, t) = 0$ and $\boldsymbol{\tau}(0, t) = \mathbf{k}$, where $\mathbf{k}$ is the unit vector of the OZ axis, at the left-hand end of the rod and $\mathbf{r}(1, t) \times \mathbf{k} = 0$ and $\boldsymbol{\tau}(1, t) = \mathbf{k}$ at the right-hand end ($s = 1$) of the rod. The force $Q_3(1, t) = Q^0$ and moment $M_3(1, t) = M^0$ acting along the undeformed rod (OZ axis) are given external parameters.

* *Moscow State Aviation Institute (Technical University), Volokolamskoe sh. 4, Moscow, 125080 Russia*

** *Moscow State University, Vorob'evy gory, Moscow, 119899 Russia*

1028-3358/02/4704-0312$22.00

A solution to the problem is the static configuration of the rod $\mathbf{r}(s)$, including the distance z^0 between the rod ends and the tie responses $Q_1(0)$, $Q_2(0)$, $M_1(0)$, and $M_2(0)$ at the $s = 0$ end and $-Q_1(1)$, $-Q_2(1)$, $-M_1(1)$, and $-M_2(1)$ at the $s = 1$ end, as well as the stability of this configuration. We note that $\mathbf{Q} = (Q_1, Q_2, Q_3) = \text{const}$ in any section s in static configurations.

2. The second formulation of the problem implies that the $s = 1$ end of the rod is fixed and $\mathbf{r}(1, t) = z^0\mathbf{k}$ is taken instead of $\mathbf{r}(1, t) \times \mathbf{k} = 0$. In this case, the distance z^0 between the rod ends and the moment value $M_3(1, t) = M^0$ (acting along the undeformed rod axis) are given external parameters. In this case, the force value $Q_3(1) = Q^0$ is determined as the response of the ties when numerically solving the problem by minimizing the residual function. The other boundary conditions remain unchanged.

First, we consider the case of the absence of the torque $M_3(1) = M^0 = 0$. In this case, the static configuration is plane. Let a solution be in the OXZ plane. Then, we have $l_{12} \equiv 0$, $M_1 \equiv 0$, and $Q_2 = 0$.

In this case, the set of equations (1) has the form

$$\frac{d^2x}{ds^2} = (Q^0x - Q_1z + M_2^0)l_{13},$$
$$\frac{dz}{ds} = l_{13}, \quad l_{13} = \sqrt{1 - \left(\frac{dx}{ds}\right)^2}, \tag{2}$$

where the force Q^0 acting on the rod end is a known (given) quantity, whereas the responses Q_1 and $M_2^0 = M_2(0)$ of the ties are unknown constants to be determined in the numerical solution.

The boundary conditions have the form

$$x(0) = 0, \quad z(0) = 0, \quad x'(0) = 0,$$
$$x(1) = 0, \quad x'(1) = 0. \tag{3}$$

The boundary value problem was solved by the shooting method. The conditions at $s = 0$ are specified, whereas the conditions at $s = 1$ are satisfied by choosing the constants Q_1 and M_2^0 that minimize the residual function

$$N_1(Q_1, M_2^0) = x(1)^2 + x'(1)^2$$

with respect to two variables.

When integrating the equation of moments, we have

$$M_2(s) = Q^0x - Q_1z + \text{const},$$

where $\text{const} = M_2(0)$, because $x(0) = 0$ and $z(0) = 0$. For symmetric forms, the equality $M_2(0) = M_2(1)$ should be satisfied and therefore $Q_1 = 0$. Figure 1 shows the numerical solution of the problem, i.e., Eqs. (2) and (3),

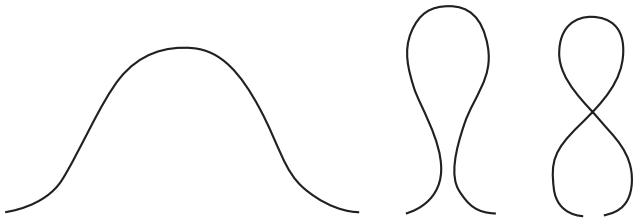


Fig. 1. Elastics $x(z)$.

for increasing the force ($Q^0 < 0$) acting on the end. As in the classical Euler problem of the stability of a rod with simply supported ends, nontrivial solutions are possible for $Q^0 < -4\pi^2$ (analogs of the Euler elastics) [7]. The plotted lines correspond to the parameters $Q^0 = -47$, -70, and -90 and $M_2^0 = 7.5105$, 14.1079, and 17.43374, respectively. In this case, the rod ends approach each other, the dependence of $z(1)$ on Q^0 is monotonic, and the transfer from the first formulation to the second formulation of the problem is elementary.

Now, we consider the three-dimensional case where $M_3(1) = M^0 \neq 0$. Let the rod take an arclike shape (Fig. 1) under the action of the force $Q^0 = -47$ and be in a stressed–strained state. We fix the distance $z(1) = z^0$ ($z^0 = 0.6866$) between the supports and increase the torque M^0. The case of simply supported ends was similarly studied in [3–6].

The set of equations for determining the static configurations of the rod axis has the form

$$\frac{d^2x}{ds^2} = (Q^0x - Q_1z + M_2^0)l_{13} - (Q_1y - Q_2x + M^0)\frac{dy}{ds},$$
$$\frac{d^2y}{ds^2} = (Q_1y - Q_2x + M^0)\frac{dx}{ds} - (Q_2z - Q^0y + M_1^0)l_{13}, \tag{4}$$
$$\frac{dz}{dz} = l_{13},$$
$$l_{13} = \sqrt{1 - \left(\frac{dx}{ds}\right)^2 - \left(\frac{dy}{ds}\right)^2}.$$

For a configuration in which $l_{13}(s)$ changes sign, the differential equation is more convenient.

The boundary conditions have the form

$$x(0) = y(0) = z(0) = 0,$$
$$x'(0) = y'(0) = 0,$$
$$x(1) = y(1) = 0, \quad x'(1) = y'(1) = 0, \tag{5}$$
$$z(1) = z^0.$$

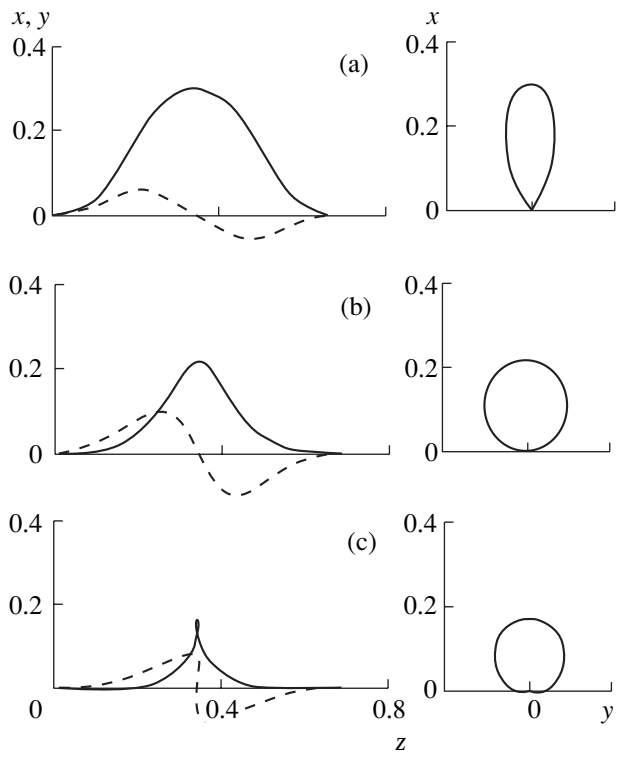


Fig. 2. Left-hand panels: functions (solid lines) $x(z)$ and (dashed lines) $y(z)$; right-hand panels: the function $x(y)$. Corresponding results of calculations are presented in the table.

The equations are independent of rotation of the coordinate system about the OZ axis, and therefore solutions of these equations are determined up to an arbitrary angle of rotation about this axis. For definiteness, we suppose that the form $x(z)$ is symmetric about the point $\frac{z^0}{2}$ and therefore $Q_2 = 0$. In this case, only three parameters, Q_1, M_1^0, and M_2^0, should be calculated.

In the classical formulation, the external conditions are the force Q^0 and the torque M^0. In this case, the right-end conditions $x(1) = y(1) = x'(1) = y'(1) = 0$ are satisfied by choosing the constants Q_1, M_1^0, and M_2^0 that minimize the residual function

$$N_2(Q_1, M_1^0, M_2^0) = x(1)^2 + y(1)^2 + x'(1)^2 + y'(1)^2.$$

It is easy to prove that the residual function is also invariant under rotation of the coordinate system about the OZ axis. Therefore, three constants are enough to find the minimum of this function.

In the second formulation, the external conditions are the torque M^0 and the distance z^0 between the supports and we have the problem specified by Eqs. (4) and (5). In this case, the force Q^0 is the response of the ties and is found by minimizing the complicated residual function

$$N_3(Q^0, Q_1, M_1^0, M_2^0)$$
$$= x(1)^2 + y(1)^2 + x'(1)^2 + y'(1)^2 + (z(1) - z^0)^2.$$

Because of symmetry, we have

$$M_1(0)^2 + M_2(0)^2 = M_1(1)^2 + M_2(1)^2,$$

and therefore

$$M_2(0) = \frac{Q_1 z^0}{2} \text{ and } M_2(1) = -\frac{Q_1 z^0}{2}.$$

A solution of Eqs. (4) and (5) is represented by the functions $x(s)$, $y(s)$, and $z(s)$. Figure 2 shows the functions $x(z)$ and $y(z)$ varying with increasing torque $M_3(1) = M^0$. The solid and dashed lines in the left-hand panels are the projections $x(z)$ and $y(z)$ of the rod shape onto the OXZ and OYZ planes, respectively; the right-hand panels show the projection $y(x)$ of the rod shape onto the OXY plane. Note that the change in the rod shape with increasing torque M^0 for a given distance between the ends is in general similar to the change found in [3–6]. As is seen, the rod shape varies from a smooth spiral (Fig. 2a) to the characteristic preloop shape at $M^0 = M_{cr}^0$ (Fig. 2c). The equations of small oscillations about the resulting equilibrium states are analyzed by the method proposed in [8].

We also calculated the dependence $Q^0(M^0)$. This function is many-valued, and the function $M^0(Q^0)$ has an extremum M_{cr}^0 at a certain value $Q^0 = Q_{cr}^0$. Thus, there are no static solutions for $M^0 > M_{cr}^0$. This fact indicates that stability is lost in this case and the formation of a loop becomes possible. A similar pattern of loop formation was observed for a rod with simply supported ends [3–6].

Calculations were performed for various z^0 values in the interval $0.6 < z^0 < 1$. An analysis of the data showed that the character of variation in rod shape in the transition to the preloop state is the same for different z^0 values.

In summary, we formulated a procedure for determining the region of external parameters for which the

Table

Fig. 2	M^0	Q^0	Q_1	M_1^0	M_2^0
a	1	−39.29	12.272969	−5.8864194	4.2133142
b	8	19.38	14.072554	0.5219528	4.8311114
c	12.35	86.87	3.3498208	1.1300019	1.1499947

formation of a loop in a rod is possible. An important feature of this procedure is the necessity of solving a new boundary value problem, where the distance between the ends of the deformed rod is *a priori* specified. It was shown that the formation of the loop in the rod is determined by two factors: a variation in the distance between the rod ends and the presence of a torque.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 99-01-01041.

REFERENCES

1. E. L. Nikolai, *Transactions on Mechanics* (Gostekhizdat, Moscow, 1955).
2. A. S. Vol'mir, *The Stability of Deformable Systems* (Nauka, Moscow, 1967).
3. M. V. Los' and A. E. Ordanovich, Vestn. Mosk. Univ., Ser. 1: Mat., Mekh., No. 5, 34 (1994).
4. M. V. Los' and A. E. Ordanovich, Vestn. Mosk. Univ., Ser. 1: Mat., Mekh., No. 3, 62 (1998).
5. M. V. Los' and A. E. Ordanovich, Vestn. Mosk. Univ., Ser. 1: Mat., Mekh., No. 6, 33 (2000).
6. M. V. Los', Candidate's Dissertation in Mathematical Physics (Moscow, 2000).
7. L. Collatz, *Eigenwertaufgaben mit technischen Anwentungen* (Geest und Portig, Leipzig, 1949; Nauka, Moscow, 1968).
8. V. A. Svetlitskiĭ, *The Mechanics of Bars* (Vysshaya Shkola, Moscow, 1987).

Translated by R. Tyapaev

Doklady Physics, Vol. 47, No. 4, 2002, pp. 316–318. Translated from Doklady Akademii Nauk, Vol. 383, No. 5, 2002, pp. 635–637.

MECHANICS

Generalized Euler's Equations Describing the Motion of a Rigid Body with a Fixed Point in $\mathbb{R}^n$

D. V. Georgievskiĭ and M. V. Shamolin

Presented by Academician V.V. Kozlov December 17, 2001

Received December 19, 2001

1. Let two Cartesian coordinate systems with the common origin O be set in $\mathbb{R}^n$. One of them (Ox_i) with the unit vectors $\mathbf{e}_i$, $i = 1, 2, \ldots, n$, is fixed, while the other (Ox_i') with the unit vectors $\mathbf{e}_i'$ is referred to an n-dimensional, perfectly rigid body. The point O is fixed when the rigid body moves in $\mathbb{R}^n$.

The components of any tensor $\mathbf{D}^{\{m\}}$ of rank m, which is representable in the above bases in the form

$$\begin{aligned}\mathbf{D}^{\{m\}} &= D_{i_1\ldots i_m}\mathbf{e}_{i_1}\otimes\ldots\otimes\mathbf{e}_{i_m}\\ &= D'_{i_1\ldots i_m}\mathbf{e}'_{i_1}\otimes\ldots\otimes\mathbf{e}'_{i_m},\end{aligned} \tag{1}$$

are transformed as[1]

$$D_{i_1\ldots i_m} = \Gamma_{i_1j_1}\ldots\Gamma_{i_mj_m}D'_{j_1\ldots j_m}, \tag{2}$$

where $\mathbf{\Gamma}^{\{2\}}(t)$ is the orthogonal transition matrix ($\mathbf{e}_i = \Gamma_{ij}\mathbf{e}_j'$). In view of the orthogonality of this matrix, the tensor

$$\mathbf{\Xi}^{\{2\}} = \dot{\mathbf{\Gamma}}^{\{2\}}\cdot(\mathbf{\Gamma}^{\{2\}})^{-1} = \Xi_{ij}\mathbf{e}_i\otimes\mathbf{e}_j = \Xi'_{ij}\mathbf{e}'_i\otimes\mathbf{e}'_j \tag{3}$$

is antisymmetric (see, e.g., [1]). Therefore, by differentiating Eq. (2) with respect to t, we can easily obtain the relation between the absolute

$$\frac{d\mathbf{D}^{\{m\}}}{dt} = \dot{D}_{i_1\ldots i_m}\mathbf{e}_{i_1}\otimes\ldots\otimes\mathbf{e}_{i_m} \tag{4}$$

and relative (local)

$$\frac{\delta\mathbf{D}^{\{m\}}}{\delta t} = \dot{D}'_{i_1\ldots i_m}\mathbf{e}'_{i_1}\otimes\ldots\otimes\mathbf{e}'_{i_m} \tag{5}$$

derivatives of the tensor $\mathbf{D}^{\{m\}}$ with respect to time:

$$\begin{aligned}\dot{D}_{i_1\ldots i_m} &= \Gamma_{i_1j_1}\ldots\Gamma_{i_mj_m}\dot{D}'_{j_1\ldots j_m}\\ &+ \Xi_{i_1j_1}D_{j_1i_2\ldots i_m} + \Xi_{i_mj_m}D_{i_1\ldots i_{m-1}j_m},\end{aligned} \tag{6}$$

or, by contracting both sides with the polyade $\mathbf{e}_{i_1}\otimes\ldots\otimes\mathbf{e}_{i_m}$, we obtain

$$\begin{aligned}\frac{d\mathbf{D}^{\{m\}}}{dt} &= \frac{\delta\mathbf{D}^{(m)}}{\delta t}\\ &+ (\Xi_{i_1j_1}D_{j_1i_2\ldots i_m} + \Xi_{i_mj_m}D_{i_1\ldots i_{m-1}j_m})\mathbf{e}_{i_1}\otimes\ldots\otimes\mathbf{e}_{i_m}.\end{aligned} \tag{7}$$

Expression (7) is known as the rule of differentiating a tensor of the mth rank with respect to time.

When $\mathbf{D}^{\{m\}}$ is the radius vector $\mathbf{r} = x_i\mathbf{e}_i = x_i'\mathbf{e}_i'$, i.e., $m = 1$ and $\dot{D}_i' = \dot{x}_i' \equiv 0$, Eq. (6) yields

$$\dot{x}_i = \Xi_{ij}x_j. \tag{8}$$

On the other hand, we have the generalized Euler's kinematic formula

$$\dot{x}_i = \epsilon_{j_1\ldots j_{n-2}ki}\Omega_{j_1\ldots j_{n-2}}x_k = \omega_{ki}x_k, \tag{9}$$

where $\Omega_{j_1\ldots j_{n-2}}$ and ω_{ki} are the fixed-basis components of the angular-velocity tensor $\mathbf{\Omega}^{\{n-2\}}$ and dual angular-velocity tensor $\boldsymbol{\omega}^{\{2\}}$, respectively [2]. These tensors are antisymmetric in any pair of their indices and have $\frac{1}{2}n(n-1) = N$ independent components. A comparison of Eqs. (8) and (9) provides

$$\mathbf{\Xi}^{\{2\}} = -\boldsymbol{\omega}^{\{2\}}. \tag{10}$$

2. Previously [2], we introduced the angular-momentum tensor $\mathbf{K}^{\{n-2\}} = K'_{i_1\ldots i_{n-2}}\mathbf{e}'_{i_1}\otimes\mathbf{e}'_{i_{n-2}}$ and its dual tensor $\mathbf{k}^{\{2\}} = k'_{ij}\mathbf{e}'_i\otimes\mathbf{e}'_j$ in $\mathbb{R}^n$:

$$\begin{aligned}K'_{i_1\ldots i_{n-2}} &= \delta^{j_1\ldots j_{n-2}ml}_{i_1\ldots i_{n-2}kl}\Omega'_{j_1\ldots j_{n-2}}b'_{km}\\ &= \frac{1}{n-2}\sum_{k=1}^{n-2}J'_{i_kl}\Omega'_{i_1\ldots i_{k-1}li_{k+1}\ldots i_{n-2}},\end{aligned} \tag{11}$$

[1] Below, the summation with respect to the known indices repeated in a monomial is from 1 to n. Free indices also range from 1 to n. There is no summation with respect to the repeated Greek indices.

Moscow State University,
Vorob'evy gory 1, Moscow, 119899 Russia

1028-3358/02/4704-0316$22.00

$$k'_{ij} = (n-2)!(b'_{im}\omega'_{mj} - b'_{jm}\omega'_{mi}). \quad (12)$$

Here, $\delta^{j_1\ldots j_n}_{i_1\ldots i_n}$ is the generalized $2n$-index Kronecker delta,

$$b'_{km} = \int_V \rho x'_k x'_m dV,$$
$$J'_{ij} = (n-2)![b'_{kk}\delta_{ij} - (n-2)b'_{ij}], \quad (13)$$

where $\rho(\mathbf{r})$ is the density of an n-dimensional body occupying a volume V and the components J'_{ij} of the symmetric inertia tensor $\mathbf{J}^{\{2\}}$ characterizing the mass geometry are constant in the movable basis.

Let an $(n-2)$-rank tensor $\mathbf{M}^{\{n-2\}}$ be the moment of external forces with respect to the point O. Let us write the relationship

$$\frac{d\mathbf{K}^{\{n-2\}}}{dt} = \mathbf{M}^{\{n-2\}} \quad (14)$$

generalizing the classical law of angular-momentum variation. In order to deal with second-rank objects for any n value, it is convenient to represent Eq. (14) in terms of dual tensors $\mathbf{k}^{\{2\}}$ and $\mathbf{m}^{\{2\}}$:

$$\frac{d\mathbf{k}^{\{2\}}}{dt} = \mathbf{m}^{\{2\}}. \quad (15)$$

It is generally difficult to write this expression in the operator form. For example, when a four-dimensional rigid body subjected to the moment of external forces moves around a fixed point, Eq. (15) [or Eq. (14)] takes the form (cf. [3–5])

$$\dot{\Omega}\Lambda + \Lambda\dot{\Omega} + [\Omega, \Omega\Lambda + \Lambda\Omega] = M, \quad (16)$$

where $\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$; $\lambda_1 = \frac{1}{2}(-I_1 + I_2 + I_3 + I_4)$, …, $\lambda_4 = \frac{1}{2}(I_1 + I_2 + I_3 - I_4)$; the moment M of the external forces acting on the body in $\mathbb{R}^4$ is projected on the "natural axes" in SO(4); $[\cdot, \cdot]$ is the commutator in SO(4); I_1, …, I_4 are the principal moments of inertia; and $\Omega \in$ SO(4) is the "angular velocity" of the body. Equation (16) on the SO(3) group takes a similar form, but it is usually written in $\mathbb{R}^3$, since $\mathbb{R}^3 \simeq$ SO(3).

Transforming the left-hand side of Eq. (15) according to Eq. (7) and taking Eq. (10) into account, we write this law for each component in the movable basis:

$$\dot{k}'_{ij} - \omega'_{ik}k'_{kj} + \omega'_{jk}k'_{ki} = m'_{ij}. \quad (17)$$

Substituting Eq. (12), which specifies the relation between $\mathbf{k}^{\{2\}}$ and $\boldsymbol{\omega}^{\{2\}}$, into Eq. (17), we arrive at the equations of motion of a rigid body with a fixed point in $\mathbb{R}^n$:

$$b'_{im}\dot{\omega}'_{mj} - b'_{jm}\dot{\omega}'_{mi} + b'_{km}(\omega'_{mi}\omega'_{jk} - \omega'_{mj}\omega'_{ik}) + (b'_{mj}\omega'_{ik} - b'_{mi}\omega'_{jk})\omega'_{mk} = m'_{ij}. \quad (18)$$

If (Ox'_i) are the principal inertia axes, we have $b'_{\alpha\beta} = J'_{\alpha\beta} = 0$, $b'_{\alpha\alpha} = B_\alpha > 0$, and

$$J'_{\alpha\alpha} = (n-2)!\left(\sum_{\beta=1}^{n} B_\beta - (n-2)B_\alpha\right). \quad (19)$$

As is seen from Eq. (19), the tensor $\mathbf{J}^{\{2\}}$ for $n \geq 4$ is not necessarily positive-definite. However, there is no contradiction here, since we showed [2] that the kinetic energy T of a body is evaluated by the formulas

$$T = \frac{1}{2}J_{kl}\Omega_{i_1\ldots i_{n-3}k}\Omega_{i_1\ldots i_{n-3}l}, \quad n \geq 3 \quad (20)$$

or

$$T = \frac{1}{2}b'_{kl}\omega'_{ki}\omega'_{li} \quad (21)$$

and is always nonnegative. Substituting the expressions

$$b'_{im} = \sum_{\alpha=1}^{n} B_\alpha \delta_{\alpha i}\delta_{\alpha m} \quad (22)$$

into Eq. (18), we obtain the generalized Euler's equations:

$$\sum_{\alpha=1}^{n} B_\alpha[(\dot{\omega}'_{\alpha j}\delta_{\alpha i} - \dot{\omega}'_{\alpha i}\delta_{\alpha j}) + (\omega'_{ik}\delta_{\alpha j} - \omega'_{jk}\delta_{\alpha i})\omega'_{\alpha k}] = m'_{ij}. \quad (23)$$

It is easy to show that these equations can be rewritten in one of the two forms

$$(B_\alpha + B_\beta)\dot{\omega}'_{\alpha\beta} - (B_\alpha - B_\beta)\omega'_{\alpha k}\omega'_{\beta k} = m'_{\alpha\beta}, \quad \alpha, \beta = 1, 2, \ldots, n, \quad (24)$$

$$(B_\alpha + B_\beta)\dot{\omega}'_{\alpha\beta} + (B_\alpha - B_\beta)(\omega^2)'_{\alpha\beta} = m'_{\alpha\beta}, \quad \alpha, \beta = 1, 2, \ldots, n. \quad (25)$$

For example, for a four-dimensional rigid body ($N = 6$) moving in a resisting medium, i.e., in a certain particular field of forces [5], Eqs. (24) and (25) take the form (ef., [6, 7])

$$\begin{aligned}
(\lambda_4 + \lambda_3)\dot{\omega}_1 + (\lambda_3 - \lambda_4)(\omega_3\omega_5 + \omega_2\omega_4) &= 0,\\
(\lambda_2 + \lambda_4)\dot{\omega}_2 + (\lambda_2 - \lambda_4)(\omega_3\omega_6 - \omega_1\omega_4) &= 0,\\
(\lambda_4 + \lambda_1)\dot{\omega}_3 + (\lambda_4 - \lambda_1)(\omega_2\omega_6 + \omega_1\omega_5) &= X_{4L}S,\\
(\lambda_3 + \lambda_2)\dot{\omega}_4 + (\lambda_2 - \lambda_3)(\omega_5\omega_6 + \omega_1\omega_2) &= 0,\\
(\lambda_1 + \lambda_3)\dot{\omega}_5 + (\lambda_3 - \lambda_1)(\omega_4\omega_6 + \omega_1\omega_3) &= -X_{3L}S,\\
(\lambda_1 + \lambda_2)\dot{\omega}_6 + (\lambda_1 - \lambda_2)(\omega_4\omega_5 + \omega_2\omega_3) &= X_{2L}S,
\end{aligned} \quad (26)$$

where, according to the traditional notation [5],

$$\mathrm{SO}(3) \ni \Omega = \begin{pmatrix} 0 & -\omega_6 & \omega_5 & -\omega_4 \\ \omega_6 & 0 & -\omega_4 & \omega_2 \\ -\omega_5 & \omega_4 & 0 & -\omega_1 \\ \omega_3 & -\omega_2 & \omega_1 & 0 \end{pmatrix}, \qquad (27)$$

Ω is the angular-velocity matrix, S is the resistance force magnitude, and $L = (0, X_{2L}, X_{3L}, X_{4L})$ is its point of application.

System (23) [or (24) and (25)] contains N independent equations with respect to N components ω'_{ij} of the antisymmetric tensor $\boldsymbol{\omega}^{\{2\}}$. If the moments m'_{ij} of external forces depend only on t and ω'_{kl} and are independent of the generalized Euler's angles $\varphi_1, \ldots, \varphi_N$ [5, 7], this system is closed and, to be integrated, must be complemented by the initial conditions $\omega'_{ij}(0) = \omega^0_{ij}$. In the three-dimensional case, in terms of the traditional notation $\omega'_{23} = \Omega'_1 = p$, $\omega'_{31} = \Omega'_2 = q$, $\omega'_{12} = \Omega'_3 = r$, $B_2 + B_3 = A$, $B_3 + B_1 = B$, and $B_1 + B_2 = C$, Eqs. (24) reduce to the three ($N = 3$) classical Euler's equations.

3. Now we consider in more detail the generalized Euler's case where the moments of external forces are absent (see, [8–10]). Taking $m'_{\alpha\beta} = 0$ in Eq. (24), we have

$$(B_\alpha + B_\beta)\dot{\omega}'_{\alpha\beta} - (B_\alpha - B_\beta)\omega'_{\alpha k}\omega'_{\beta k} = 0. \qquad (28)$$

System (28) has a number of first integrals. To obtain them, each of N equations (28) must be multiplied by $F(\alpha; \beta; M)\omega'_{\alpha\beta}$, where

$$2F(\alpha; \beta; M) = B_\alpha^M + B_\alpha^{M-1}B_\beta + \ldots + B_\alpha B_\beta^{M-1} + B_\beta^M, \qquad (29)$$

and M is an arbitrary nonnegative integer, and summed up. Since $2(B_\alpha - B_\beta)F(\alpha; \beta; M) = B_\alpha^{M+1} - B_\beta^{M+1}$ and

$$\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}(B_\alpha^{M+1} - B_\beta^{M+1})\omega'_{\alpha k}\omega'_{\beta k}\omega'_{\alpha\beta} = 0, \qquad (30)$$

we obtain the denumerable set of the first integrals for Eq. (28):

$$\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}(B_\alpha + B_\beta)F(\alpha; \beta; M)\omega'^2_{\alpha\beta} = C_M^2, \qquad C_M = \text{const}. \qquad (31)$$

(i) $M = 0$. According to Eq. (29), $F(\alpha; \beta; 0) = 1$ and Eq. (31) is the energy integral:

$$\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}(B_\alpha + \beta_\beta)\omega'^2_{\alpha\beta} = C_0^2. \qquad (32)$$

Indeed, it follows from Eq. (21) that the left-hand side of Eq. (32) is the quadruplicate kinetic energy of a body. This result can be obtained by multiplying homogeneous equations (23) by ω'_{ij} and summing with respect to i and j.

(ii) $M = 1$ and $F(\alpha; \beta; 1) = B_\alpha + B_\beta$. Using Eq. (31), we obtain a first integral known as the integral of moments in the three-dimensional case:

$$\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}(B_\alpha + B_\beta)^2\omega'^2_{\alpha\beta} = C_1^2. \qquad (33)$$

In the n-dimensional case, the left-hand side of Eq. (33) up to a factor is also the sum of the squares of all components $K'_{i_1 \ldots i_{n-2}}$ [see Eqs. (11), (12)]; i.e., Eq. (33) can be written in the form $\|\mathbf{K}^{\{n-2\}}\|^2 = \tilde{C}_1^2$ as well.

(iii) The first $n - 1$ first integrals of Eq. (31) (i.e., for $M = 0, 1, \ldots, n - 2$) are linearly independent in $\mathbb{R}^n$. Thus, the number of independent first integrals in the Euler's case is less than the number of the angular-velocity components by $\frac{1}{2}(n-2)(n-1)$.

REFERENCES

1. A. P. Markeev, *Theoretical Mechanics* (CheRo, Moscow, 1999).
2. D. V. Georgievskiĭ and M. V. Shamolin, Dokl. Akad. Nauk **380**, 47 (2001) [Dokl. Phys. **46**, 663 (2001)].
3. O. I. Bogoyavlenskiĭ, Dokl. Akad. Nauk SSSR **287**, 1105 (1986) [Sov. Phys. Dokl. **31**, 309 (1986)].
4. O. I. Bogoyavlenskiĭ, Dokl. Akad. Nauk SSSR **272**, 1364 (1983) [Sov. Phys. Dokl. **28**, 858 (1983)].
5. M. V. Shamolin, Dokl. Akad. Nauk **375**, 343 (2000) [Dokl. Phys. **45**, 632 (2000)].
6. V. V. Trofimov and A. T. Fomenko, Itogi Nauki Tekh., Ser.: Sovr. Probl. Mat., Noveish. Dostizh. **29**, 3 (1986).
7. B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, *Modern Geometry—Methods and Applications* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984, 1985, 1990), Parts 1–3.
8. A. P. Veselov, Dokl. Akad. Nauk SSSR **270**, 1298 (1983).
9. S. V. Manakov, Funkts. Analiz. Ego Prilozh. **10**, 93 (1976).
10. M. A. Ol'shanetskiĭ and A. M. Perelomov, Funkts. Analiz. Ego Prilozh. **11**, 75 (1977).

Translated by Yu. Vishnyakov

Doklady Physics, Vol. 47, No. 4, 2002, pp. 319–323. Translated from Doklady Akademii Nauk, Vol. 383, No. 5, 2002, pp. 638–642.

MECHANICS

Stress Characteristic Relations for the Three-Dimensional Problem of the Full Limit Equilibrium of a Soil Continuum

D. D. Ivlev*, Academician A. Yu. Ishlinskiĭ, and R. I. Nepershin*****

Received December 18, 2001

We present the characteristic relationships for stresses in the three-dimensional problem for a soil continuum under the full limit equilibrium according to the Coulomb criterion. These relationships involve the limiting cases of the plane and axisymmetric problems of the limit equilibrium. We considered the problem of the limit equilibrium of a half-space subjected to the pressure of a smooth flat die with an elliptic contour in plan.

Let us consider the three-dimensional limit equilibrium of a soil continuum satisfying the Coulomb criterion

$$\max|\tau_n| = k + \sigma_n \tan\rho, \tag{1}$$

where k is the cohesion, ρ is the angle of internal friction, and τ_n and σ_n are the tangential and normal stresses in areas along which a material slips in the limit-equilibrium case. The positive sign is assigned to compressive normal stresses.

Criterion (1) can be represented by a hexahedral pyramid in the principal-stress space. For $\rho = 0$, this pyramid degenerates into the Tresca prism in the case of a perfectly plastic body. The three-dimensional slip under criterion (1) is possible only at the Coulomb-pyramid edges [1, 2]. The stressed state corresponding to the Coulomb-pyramid edges is referred to as the state of the full limit equilibrium for which the principal stresses satisfy the relationships

$$\sigma_1 = \sigma_2, \quad \sigma_3 > \sigma_1, \text{ or } \sigma_3 < \sigma_1. \tag{2}$$

For the full limit equilibrium, the three-dimensional problem of a soil continuum is statically definable and hyperbolic [3, 4]. The axis of the characteristic cone coincides with the direction of the stress σ_3. The slip surfaces are tangent to the characteristic cone and inclined to the σ_3 direction at the angle

$$\mu = \frac{\pi}{4} - \frac{\rho}{2}. \tag{3}$$

We specify the direction of the stress σ_3 in the Cartesian coordinates $\{x, y, z\}$ by the unit vector $\mathbf{n}$:

$$n_1 = \cos\varphi, \quad n_2 = \sin\varphi\sin\theta, \quad n_3 = \sin\varphi\cos\theta, \tag{4}$$

where φ is the angle between the x-axis and the vector $\mathbf{n}$, whereas θ is the angle between the z-axis and the component of the vector $\mathbf{n}$ in the $\{y, z\}$ plane. The unit vectors of the directions of the stresses σ_1 and σ_2 are denoted as $\mathbf{l}$ and $\mathbf{m}$, and the surface orthogonal to the vector $\mathbf{m}$ is considered. The characteristic cone intersects this surface along the directions of slip lines α and β forming the angles $\pm\mu$ with the vector $\mathbf{n}$. The unit vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of the slip lines are defined from the scalar products

$$\boldsymbol{\alpha}\cdot\mathbf{n} = \cos\mu, \quad \boldsymbol{\alpha}\cdot\mathbf{l} = \sin\mu, \quad \boldsymbol{\alpha}\cdot\mathbf{m} = 0; \tag{5}$$

$$\boldsymbol{\beta}\cdot\mathbf{n} = \cos\mu, \quad \boldsymbol{\beta}\cdot\mathbf{l} = -\sin\mu, \quad \boldsymbol{\beta}\cdot\mathbf{m} = 0, \tag{6}$$

which specify the slip surfaces for known directions of principal stresses.

Following [5], we introduce the reduced average pressure σ equal to the distance from the center of the Mohr's circle to the Coulomb-pyramid vertex in the stress plane $\{\sigma_n, \tau_n\}$:

$$\sigma = \frac{1}{2}(\sigma_1 + \sigma_3) + h, \quad h = k\cot\rho. \tag{7}$$

The center and radius of the Mohr's circle are defined by the relationships

$$\frac{1}{2}(\sigma_1 + \sigma_3) = \sigma - h, \quad \frac{1}{2}(\sigma_3 - \sigma_1) = \lambda\sigma\sin\rho, \tag{8}$$

where $\lambda = +1$ if $\sigma_3 > \sigma_1$ and $\lambda = -1$ if $\sigma_3 < \sigma_1$, from

* *Yakovlev Chuvash State Pedagogical University, pr. K. Marksa 38, Cheboksary, 428000 Russia*

** *Institute for Problems of Mechanics, Russian Academy of Sciences, pr. Vernadskogo 101, Moscow, 117526 Russia*

*** *Moscow State Academy of Instrumentation Engineering and Computer Sciences, ul. Stromynka 20, Moscow, 107846 Russia*

1028-3358/02/4704-0319$22.00

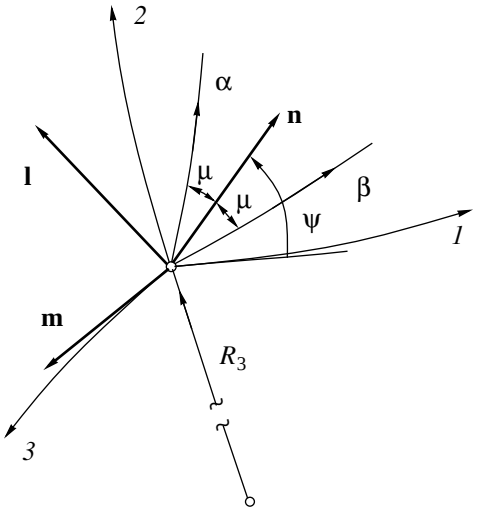


Fig. 1. Curvilinear coordinates and slip lines at the surface orthogonal to the stress σ_2.

which we find the principal stresses

$$\sigma_1 = \sigma_2 = \sigma(1-\lambda\sin\rho) - h,$$
$$\sigma_3 = \sigma(1+\lambda\sin\rho) - h. \quad (9)$$

We introduce the local orthogonal system of coordinates {*1*, *2*, *3*} so that line *3* is directed along the unit vector **m** of the stress σ_2 and the direction of line *2* is opposite to the direction of the principal normal to line *3* (Fig. 1). The principal normals to coordinate lines *1* and *2* are directed along the vector –**m**. The direction cosines of the principal stresses in the coordinates {*1*, *2*, *3*} have the form

$$n_1 = \cos\psi, \quad n_2 = \sin\psi, \quad n_3 = 0,$$
$$l_1 = -\sin\psi, \quad l_2 = \cos\psi, \quad l_3 = 0, \quad (10)$$
$$m_1 = 0, \quad m_2 = 0, \quad m_3 = 1,$$

where ψ is the angle between the vector **n** and line *1*.

From Eqs. (9) and (10), we find the components of the stress tensor in the coordinates {*1*, *2*, *3*}:

$$\sigma_{11} = \sigma(1+\lambda\sin\rho\cos2\psi) - h,$$
$$\sigma_{22} = \sigma(1-\lambda\sin\rho\cos2\psi) - h, \quad (11)$$
$$\tau_{12} = \lambda\sigma\sin\rho\sin2\psi, \quad \sigma_{33} = \sigma(1-\lambda\sin\rho) - h.$$

We consider the equilibrium of an element of the continuum in the coordinates {*1*, *2*, *3*} with allowance for the body force **F**. We take the angles ξ_i between tangents and lines *1*, *2*, and *3* as curvilinear coordinates. The differentials of the arcs of the coordinate lines are related to the radii of curvature R_i as

$$dS_1 = R_1 d\xi_1, \quad dS_2 = R_2 d\xi_2, \quad dS_3 = R_3 d\xi_3, \quad (12)$$

which shows that the radii of curvature R_i are the Lamé parameters. The equilibrium equations in the coordinates ξ_i have the form

$$\frac{\partial\sigma_{11}}{\partial S_1} + \frac{\partial\tau_{12}}{\partial S_2} + (\sigma_{11}-\sigma_{22})\frac{\partial\ln R_2}{\partial S_1}$$
$$+ (\sigma_{11}-\sigma_{33})\frac{\partial\ln R_3}{\partial S_1} + \tau_{12}\left(2\frac{\partial\ln R_1}{\partial S_2} + \frac{\partial\ln R_3}{\partial S_2}\right) = F_1, \quad (13)$$

$$\frac{\partial\tau_{12}}{\partial S_1} + \frac{\partial\sigma_{22}}{\partial S_2} + (\sigma_{22}-\sigma_{11})\frac{\partial\ln R_1}{\partial S_2}$$
$$+ (\sigma_{22}-\sigma_{33})\frac{\partial\ln R_3}{\partial S_2} + \tau_{12}\left(2\frac{\partial\ln R_2}{\partial S_1} + \frac{\partial\ln R_3}{\partial S_1}\right) = F_2, \quad (14)$$

$$\frac{\partial\sigma_{33}}{\partial S_3} + (\sigma_{33}-\sigma_{11})\frac{\partial\ln R_1}{\partial S_3} + (\sigma_{33}-\sigma_{22})\frac{\partial\ln R_2}{\partial S_3} = F_3, \quad (15)$$

where F_1, F_2, and F_3 are the projections of the vector **F** onto directions *1*, *2*, and *3*.

The radii of curvature of coordinate lines *1*, *2*, *3* satisfy the relationships

$$\frac{\partial R_1}{\partial S_2} = \frac{\partial R_2}{\partial S_1} = \frac{\partial R_3}{\partial S_1} = 0, \quad \frac{\partial R_1}{\partial S_3} = \frac{\partial R_2}{\partial S_3} = \frac{\partial R_3}{\partial S_2} = 1, \quad (16)$$

which make it possible to represent equilibrium equations (13)–(15) in the form

$$\frac{\partial\sigma_{11}}{\partial S_1} + \frac{\partial\tau_{12}}{\partial S_2} + \frac{\tau_{12}}{R_3} = F_1, \quad (17)$$

$$\frac{\partial\tau_{12}}{\partial S_1} + \frac{\partial\sigma_{22}}{\partial S_2} + \frac{\sigma_{22}-\sigma_{33}}{R_3} = F_2, \quad (18)$$

$$\frac{\partial\sigma_{33}}{\partial S_3} = F_3 + (\sigma_{11}-\sigma_{33})\frac{1}{R_1} + (\sigma_{22}-\sigma_{33})\frac{1}{R_2}. \quad (19)$$

If coordinate lines *1*, *2*, and *3* are straight and the body-force vector lies in the plane {*1*, *2*}, Eqs. (17) and (18) go over into the limit-equilibrium equations for a soil continuum under plane strain with differential relationships for σ and ψ along the slip lines [5]. If only lines *1* and *2* are straight, all the planes {*1*, *2*} pass through axis *1*, and $F_3 = 0$, we obtain the equations of the full limit equilibrium for axisymmetric strain with the differential relationships along the slip lines, which were presented in [6]. In these cases, it follows from Eq. (19) that $\sigma_{33} = \sigma_2 = \text{const}$ along the normal to the plane {*1*, *2*}.

We write the characteristic relationships [5, 6] for σ and ψ along the slip lines α and β in the following form convenient for calculations in the planes {*1*, *2*} of the three-dimensional problem:

$$\cos\rho\, d\sigma + 2\sigma\sin\rho\, d\psi = (A_F + A_3)dS_\alpha$$
$$\text{along the } \alpha \text{ line}, \quad (20)$$

$$\cos\rho\, d\sigma - 2\sigma\sin\rho\, d\psi = (B_F + B_3)dS_\beta \quad \text{along the } \beta \text{ line}, \tag{21}$$

where dS_α and dS_β are the differentials of the arcs of the slip line, and the coefficients A and B are determined in terms of the body force and the curvature of coordinate line *3* as

$$\begin{aligned} A_F &= F_2\cos(\psi-\mu) - F_1\sin(\psi-\mu), \\ B_F &= F_1\sin(\psi+\mu) - F_2\cos(\psi+\mu), \\ A_3 &= \frac{\sigma\sin\rho}{2R_3}[\cos(\psi+\mu) - \lambda\cos(\psi-\mu)], \\ B_3 &= -\frac{\sigma\sin\rho}{2R_3}[\cos(\psi-\mu) - \lambda\cos(\psi+\mu)]. \end{aligned} \tag{22}$$

Equations (11) and (19) provide the following relationship along the normal to the surface $\{1, 2\}$:

$$\begin{aligned} &\frac{\partial\sigma}{\partial S_3}(1-\lambda\sin\rho) \\ &= F_3 + \sin\rho\left[\frac{1+\cos 2\psi}{R_1} + \frac{1-\cos 2\psi}{R_2}\right]. \end{aligned} \tag{23}$$

If the surface $\{1, 2\}$ has small curvature $\left(\frac{1}{R_1} \approx 0\right.$ and $\left.\frac{1}{R_2} \approx 0\right)$ and $F_3 = 0$, it follows from Eq. (23) that $\sigma \approx$ const and Eqs. (20) and (21), together with the differential equations of the slip lines, form a closed set with respect to σ and ψ and to the coordinates of the slip-line points. In a number of important applied problems, the surface $\{1, 2\}$ can be plane or nearly plane. In this case, the three-dimensional problem is solved as a sequence of two-dimensional problems at known surfaces $\{1, 2\}$ by using Eq. (23) for checking the calculation errors.

The general case of the three-dimensional problem with determination of the unknown surfaces $\{1, 2\}$ requires further investigation, because Eqs. (20)–(23) involve unknown curvatures $\frac{1}{R_1}$, $\frac{1}{R_2}$, and $\frac{1}{R_3}$. A similar situation arises in the three-dimensional problem for a perfectly plastic body under conditions of perfect plasticity [7].

We consider the problem of the full limit equilibrium for the $x \geq 0$ half-space subjected to the pressure of a flat smooth die with an elliptic contour in plan with allowance for gravity along the x-axis.

We take the length of the small ellipse semiaxis directed along the z-axis as the characteristic size $H = 1$. The length of the large semiaxis directed along the y-axis is designated as b. We consider the half-space in the region $y \geq 0$ and $z \geq 0$. At the boundary $x = 0$ ahead of the die contour, we specify the normal pressure p. The quantity $\gamma H = 1$, where γ is the volume weight of the soil continuum, is taken as the characteristic stress.

When the die is indented along the x-axis, the highest compressive stress σ_3 in the surface layer in front of the die is directed along the normal to the die boundary. Under the condition $\sigma_2 = \sigma_1 = p$ of full limit equilibrium in this layer, σ_1 is directed along the normal to the half-space boundary, $\sigma_3 > \sigma_1$ and $\lambda = +1$. Taking these conditions into account, we suppose that the surfaces $\{1, 2\}$ are planes that are orthogonal to the die contour and parallel to the x-axis. Due to the symmetry of the problem, this assumption is valid in the planes $y = 0$ and $z = 0$ and in the region in front of the die.

We write the parametric equation of the die boundary in the plane $x = 0$:

$$z_0 = \cos\xi, \quad y_0 = b\sin\xi, \quad 0 \leq \xi \leq \frac{\pi}{2}. \tag{24}$$

The radius of curvature R of the die boundary and the angle θ between the z-axis and the plane $\{1, 2\}$ passing through the point (z_0, y_0) are determined as

$$R = \frac{1}{b}(\sin^2\xi + b^2\cos^2\xi)^{3/2}, \quad \tan\theta = \frac{\tan\xi}{b}. \tag{25}$$

The plane $\{1, 2\}$ intersects the y-axis at a distance d from the die boundary

$$d = \frac{1}{b}\sqrt{\sin^2\xi + b^2\cos^2\xi}. \tag{26}$$

In the plane $\{1, 2\}$, we choose the Cartesian coordinates $\{x_1, x_2\}$ for which the origin is at the point of intersection of this plane with the y-axis and the x_1-axis is parallel to the x-axis of the basic system of coordinates $\{x, y, z\}$. The radius of curvature of coordinate line *3* normal to the plane $\{1, 2\}$ is defined as

$$R_3 = R - d + x_2. \tag{27}$$

The body-force vector has the components $F_1 = 1$, $F_2 = F_3 = 0$. The angle ψ coincides with the angle φ of the basic system of coordinates, and the differential equations of the slip lines in the plane $\{x_1, x_2\}$ have the form

$$\frac{dx_2}{dx_1} = \tan(\psi+\mu) \quad \text{for the } \alpha \text{ line}, \tag{28}$$

$$\frac{dx_2}{dx_1} = \tan(\psi-\mu) \quad \text{for the } \beta \text{ line}. \tag{29}$$

Figure 2 shows the slip-line field in the plane $\theta =$ const under the assumption that the material can slip along the smooth die boundary OA when the limit equilibrium is reached. The boundary AB is free of tangential stresses and is loaded by the pressure $p = \sigma_{11}$.

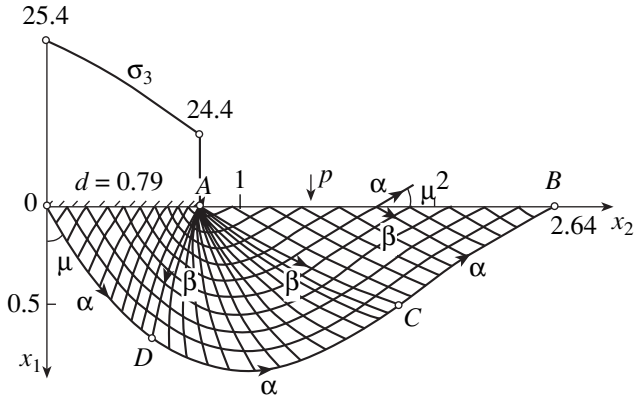


Fig. 2. Slip lines and the distribution of pressure over the boundary of the contact between the elliptic die and the half-space in the plane $\theta = 0.245$ for $b = 2$ and $p = 0.1$.

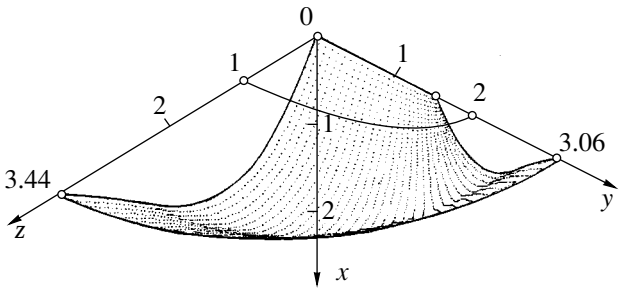


Fig. 3. Limit slip surface in the region $z \geq 0$ and $y \geq 0$ for the elliptic die pressure upon the half-space.

At this boundary, the direction σ_3 defining the angle $\psi = \frac{\pi}{2}$ is known. From the first of Eqs. (11), we find

$$\sigma = \frac{p+h}{1-\sin\rho}, \quad \psi = \frac{\pi}{2}, \qquad x_1 = 0, \quad x_2 \geq d \text{ at } AB. \tag{30}$$

At the boundary OA, the stress σ_3 is directed along the x_1-axis and $\tau_{12} = 0$. From the third of Eqs. (11), we find

$$\psi = 0, \quad x_1 = 0, \quad x_2 \leq d \text{ at } OA. \tag{31}$$

At point A, which is a degenerate slip α line, the angle ψ changes from 0 to $\frac{\pi}{2}$ according to conditions (30) and (31). Integrating Eq. (20) for $dS_\alpha = 0$ with allowance for condition (30), we find

$$\sigma = \frac{p+h}{1-\sin\rho}\exp[\tan\rho(\pi - 2\psi)], \qquad 0 \leq \psi \leq \frac{\pi}{2}, \quad x_1 = 0, \quad x_2 = d \tag{32}$$

at point A.

Boundary conditions (30)–(32), equations (28), (29), and relationships (20)–(22) determine the slip-line field in the region ABC from the solution to the Cauchy problem with conditions (30) at AB. In the region ACD, this field is determined by solving the Goursat problem with σ, ψ, x_1, and x_2 values specified at AC and at point A. In the region AOD, this field is found by solving the mixed problem for σ, ψ, x_1, and x_2 values specified at AD and conditions (31) at OA. The unknown length L of the boundary AB must satisfy the functional equation

$$x_2(L) = 0 \quad \text{at point } O, \tag{33}$$

because the boundary of the limit-equilibrium region $ODCB$ is unambiguously determined when the sequence of the above-indicated boundary value problems is solved.

Equations (28), (29) with allowance for relationships (20)–(22) are numerically solved by replacing differentials and unknown functions by finite differences and by average values of the functions, respectively. The length L of the boundary AB is found via solving Eq. (33) by Newton's method.

Executing similar calculations for the sequence of the planes $\{x_1, x_2\}$ defined by Eqs. (24), (25), we find slip surfaces and the pressure distribution over the die–half-space contact. The average pressure on the die can be found by integrating the contact-pressure distribution:

$$q = \frac{4}{\pi b}\int_0^{\pi/2}\int_0^d \sigma_3 R_3 dx_2 d\theta. \tag{34}$$

The principal stresses in the plane $\{x_1, x_2\}$ are found from Eqs. (9). We find the vector $\mathbf{n}$ in the basic system of coordinates $\{x, y, z\}$ from Eqs. (4) for $\varphi = \psi$. The components of the vector $\mathbf{m}$ are $m_1 = 0$, $m_2 = -\cos\theta$, and $m_3 = \sin\theta$. The vector $\mathbf{l}$ is found from the vector product $\mathbf{l} = \mathbf{m} \times \mathbf{n}$: $l_1 = -\sin\varphi$, $l_2 = -\cos\varphi\sin\theta$, and $l_3 = \sin\theta$. From the known principal stresses, we find the components of the stress tensor in the coordinates $\{x, y, z\}$, which are related to the coordinates $\{x_1, x_2\}$ as

$$x = x_1, \quad y = y_0 + (x_2 - d)\sin\theta, \quad z = x_2\cos\theta. \tag{35}$$

To solve the problem of the limit pressure of the elliptic die, we write a computer program. Figure 2 shows the slip-line field and the contact-pressure distribution over the cross section $\theta = 0.245$ for the elliptic die at $b = 2$ and $p = 0.1$ and for the soil-continuum parameters $\rho = 30°$, $k = 0.75$, and $\gamma H = 1$.

For the above parameters of the problem, Fig. 3 shows the isometric projection of the α-family limit slip surface, which passes through the plane of symmetry $z = 0$. The corresponding distribution of the limit pressure on the die is shown in Fig. 4 in the form of the isometric projection of the surface $\sigma_3(x, y)$ over the plane of the die–half-space contact. In this case, the

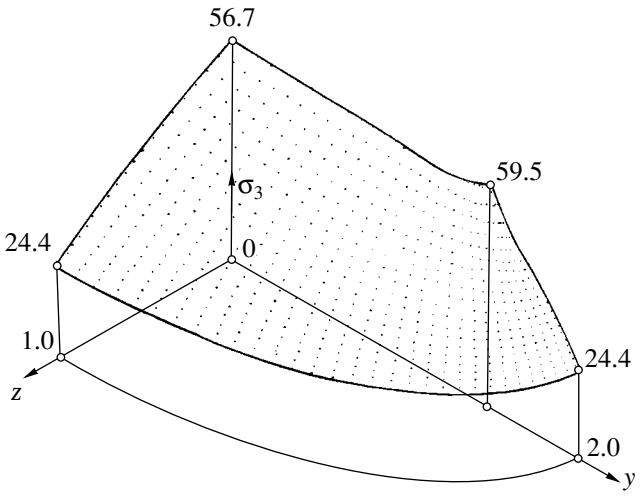


Fig. 4. Distribution of the normal pressure for a quarter of the plane of the contact between the elliptic die and half-space.

average pressure q on the die is equal to 39.87. The limit-equilibrium region decreases with increasing ellipse curvature in the cross sections θ = const and with increasing pressure on the die. The curvature $\frac{1}{R}$ is almost constant near the ellipse focus, and the problem is close to the axially symmetric problem where the condition σ = const is almost exactly satisfied along the normal to the planes θ = const. In other cross sections, the lines of the level σ = const deviate negligibly from the normals to the planes θ = const.

REFERENCES

1. A. Yu. Ishlinskiĭ, Izv. Akad. Nauk SSSR, Otd. Tekh. Nauk, No. 3, 250 (1945).
2. A. Yu. Ishlinskiĭ, *Applied Problems in Mechanics* (Nauka, Moscow, 1986), Vol. 1, pp. 62–83.
3. D. D. Ivlev, Prikl. Mat. Mekh. **22**, 90 (1958).
4. D. D. Ivlev, *Mechanics of Yielding Media* (Fizmatlit, Moscow, 2001), Vol. 1, pp. 5–14.
5. V. V. Sokolovskiĭ, *Statics of the Soil Continuum* (Gos. Izd. Tekh. Teor. Lit., Moscow, 1954).
6. V. G. Berezantsev, *Axisymmetric Problem of the Limit Equilibrium of the Soil Continuum* (Gos. Izd. Tekh. Teor. Lit., Moscow, 1952).
7. D. D. Ivlev, A. Yu. Ishlinskiĭ, and R. I. Nepershin, Dokl. Akad. Nauk **381**, 616 (2001) [Dokl. Phys. **46**, 890 (2001)].

Translated by V. Bukhanov

Doklady Physics, Vol. 47, No. 4, 2002, pp. 324–327. Translated from Doklady Akademii Nauk, Vol. 383, No. 6, 2002, pp. 767–770.

MECHANICS

Determination of Frequencies of Bending Magnetoelastic Vibrations in Plates in the Three-Dimensional and Averaged Formulations

Yu. S. Safaryan

Presented by Academician S.S. Grigoryan January 10, 2002

Received November 28, 2001

The linear problems of bending magnetoelastic vibrations in a plate were considered in [1–3]. Nonlinear modulation waves in a longitudinal field were studied in [4, 5]. These investigations were based on the classical averaged theory of thin plates. In this study, the three-dimensional problem of vibrations in a plate in transverse and longitudinal magnetic fields is analyzed first for infinitely large electric conductivity and second for finite electric conductivity. The results are compared with those obtained in the averaged theory. It is shown that both theories provide identical results for large electric conductivity in a longitudinal field, whereas these theories give quantitatively and qualitatively different results for a transverse field. In the three-dimensional formulation (as well as in the averaged one), the longitudinal field increases frequencies, whereas the transverse field decreases frequencies in the three-dimensional formulation and increases them in the averaged formulation. In the case of finite electric conductivity, these formulations provide different results for both transverse and longitudinal fields.

Experiments performed in weak magnetic fields from 0.05 to 0.5 T corroborate the results of the three-dimensional theory.

The problems investigated in this study may be of interest in designing setups for controlled thermonuclear reactions.

Since the wave problem for a plate is isotropic, it is sufficient to consider the plane problem. Let the x-axis be directed along the mean line of the plate along which the wave propagates; the z-axis be directed normally to this line; the unperturbed magnetic field H_0 be directed along the z-axis; u_x and u_z be the displacement components along the x- and z-axes, respectively; and $\mathbf{H} = \mathbf{H}_0 + \mathbf{h}$ be the magnetic-field vector. The induced field is represented as

$$h_x = H_0 H_x', \quad h_z = H_0 H_z'.$$

For quasi-monochromatic waves, we assume that

$$\begin{gathered} u_x = \frac{1}{2}U_x(z)\exp(i\tau) + \text{c.c.}, \\ u_z = \frac{1}{2}U_z(z)\exp(i\tau) + \text{c.c.}, \\ H_z' = \frac{1}{2}H_z(z)\exp(i\tau) + \text{c.c.}, \quad \tau = kx - \omega t, \\ H_x' = \frac{1}{2}H_x(z)\exp(i\tau) + \text{c.c.} \end{gathered} \tag{1}$$

In a transverse magnetic field, the basic equations in variables (1) have the form

$$\begin{gathered} \frac{b^2}{a^2}\frac{d^2U_x}{dz^2} - k^2U_x + \frac{\omega^2}{a^2}U_x + \xi ik\frac{dU_z}{dz} \\ = -\frac{a_1^2}{a^2}\left(\frac{dH_x}{dz} - ikH_z\right); \end{gathered} \tag{2}$$

$$\begin{gathered} \xi = 1 - \frac{b^2}{a^2}, \\ \xi ik\frac{dU_x}{dz} + \frac{d^2U_z}{dz^2} - \frac{b^2}{a^2}k^2U_z + \frac{\omega^2}{a^2}U_z = 0; \end{gathered} \tag{3}$$

$$-i\omega H_x + \nu_m k^2 H_x - \nu_m \frac{d^2H_x}{dz^2} = -i\omega\frac{dU_x}{dz}; \tag{4}$$

$$-i\omega H_z + \nu_m k^2 H_z - \nu_m \frac{d^2H_z}{dz^2} = -\omega k U_x. \tag{5}$$

Armenian State Engineering University, Goris Branch, ul. Avangard 4, Goris, 377830 Armenia

1028-3358/02/4704-0324$22.00 © 2002 MAIK “Nauka/Interperiodica”

Here, $a_1^2 = \frac{H_0^2}{4\pi\rho}$; a and b are the velocities of the longitudinal and transverse elastic waves, respectively; σ is the electric conductivity; c is the speed of light; and $\nu_m = \frac{c^2}{4\pi\sigma}$ is the magnetic viscosity.

The solution of the three-dimensional problem can be sought in the form [6]

$$\begin{aligned} U_z &= A_j \cosh \nu_j z, \quad U_x = B_j \sinh \nu_j z, \\ H_x &= C_j \cosh \nu_j z, \quad H_z = D_j \sinh \nu_j z, \end{aligned} \tag{6}$$

where summation over the repeated indices from 1 to 3 is implied. Equations (2)–(5) provide the relations between all constants entering into Eq. (6) via $B_{1.2.3}$ by the following formulas:

$$\begin{aligned} C_{1.2.3} &= -\frac{i\omega\nu_{1.2.3}B_{1.2.3}}{X_{1.2.3}}, \quad D_{1.2.3} = -\frac{\omega k B_{1.2.3}}{X_{1.2.3}}, \\ X_{1.2.3} &= -i\omega + \nu_m k^2 - \nu_m \nu_{1.2.3}^2, \\ \xi i k B_{1.2.3}\nu_{1.2.3} &+ \left(\nu_{1.2.3}^2 - \frac{b^2}{a^2}k^2 + \frac{\omega^2}{a^2}\right)A_{1.2.3} = 0; \\ \left(\frac{b^2}{a^2}\nu_{1.2.3}^2 - k^2 + \frac{\omega^2}{a^2}\right)&B_{1.2.3} + \xi i k \nu_{1.2.3} A_{1.2.3} \\ &= -\frac{a_1^2}{a^2}(C_{1.2.3}\nu_{1.2.3} - ikD_{1.2.3}). \end{aligned} \tag{7, 8}$$

Equations (7) and (8) provide the following representation of $\bar{\nu} = \nu_{1.2.3}$ in terms of ω:

$$\begin{aligned} \frac{b^2}{a^2}\bar{\nu}^2 - k^2 + \frac{\omega^2}{a^2} + \xi^2 \frac{k^2\bar{\nu}^2}{\bar{\nu}^2 - \frac{b^2}{a^2}k^2 + \frac{\omega^2}{a^2}} \\ = -\frac{a_1^2}{a^2}\frac{\bar{\nu}^2 - k^2}{1 + i\frac{k^2 - \bar{\nu}^2}{\omega}\nu_m}. \end{aligned} \tag{9}$$

For finite values of σ, all values of $\nu_{1.2.3}$ are also finite. For $\sigma = \infty$, it is possible to obtain two finite values of $\bar{\nu}^2 = \nu_{1.2}^2$. However, for large finite values of σ, there is a third value $\bar{\nu}^2 = \nu_3^2$ that is defined as

$$i\frac{\nu_3^2}{\omega}\nu_m = \frac{a_1^2}{b^2} + 1 \tag{10}$$

when $\frac{\omega}{\nu_m k^2} \gg 1$. In order to obtain a dispersion relation for $\omega(k)$ in the three-dimensional formulation, Eqs. (2)–(5) should be complemented by the boundary conditions $\sigma_z = \sigma_{xz} = 0$ at the plate–dielectric interfaces $z = \pm\frac{h}{2}$ and continuity conditions of $\bar{h}$. The induced magnetic field in the dielectric outside the plate is written in the form

$$\begin{aligned} \tilde{h}_x &= \frac{1}{2}(C_1' \exp(\theta) + \text{c.c.}), \\ \tilde{h}_z &= \frac{1}{2}(C_2' \exp(\theta) + \text{c.c.}), \quad \theta = i\tau \mp kz. \end{aligned} \tag{11}$$

Using the equation $\frac{\partial\tilde{h}_x}{\partial x} + \frac{\partial\tilde{h}_z}{\partial z} = 0$, we represent the above boundary conditions in the form

$$\begin{aligned} C_j \cosh \nu_j \frac{h}{2} &= -kC_j\nu_j^{-1}\sinh\nu_j\frac{h}{2}, \\ B_j\nu_j\cosh\nu_j\frac{h}{2} + ikA_j\cosh\nu_j\frac{h}{2} &= 0, \\ A_j\nu_j\sinh\nu_j\frac{h}{2} + \frac{a^2 - 2b^2}{a^2}ikB_j\sinh\nu_j\frac{h}{2} &= 0, \end{aligned} \tag{12}$$

where Eqs. (7)–(10) should be substituted. The determinant equation for Eqs. (12) has the form

$$\begin{vmatrix} \frac{1 + \frac{k}{\nu_1}\tanh\nu_1\frac{h}{2}}{\chi_1} & \frac{1 + \frac{k}{\nu_2}\tanh\nu_2\frac{h}{2}}{\chi_2} & \frac{1 + \frac{k}{\nu_3}\tanh\nu_3\frac{h}{2}}{\chi_3} \\ 1 + \frac{\xi}{\Delta_1}k^2 & 1 + \frac{\xi}{\Delta_2}k^2 & 1 + \frac{\xi}{\Delta_3}k^2 \\ \frac{\tanh\nu_1\frac{h}{2}}{\nu_1}\Gamma_1 & \frac{\tanh\nu_2\frac{h}{2}}{\nu_2}\Gamma_2 & \frac{\tanh\nu_3\frac{h}{2}}{\nu_3}\Gamma_3 \end{vmatrix} = 0. \tag{13}$$

Here,

$$\begin{aligned} \Delta_j &= \nu_j^2 - \frac{b^2}{a^2}k^2 + \frac{\omega^2}{a^2}, \quad \chi_j = 1 + i\frac{k^2 - \nu_j^2}{\omega}\nu_m, \\ \Gamma_j &= \frac{a^2 - 2b^2}{a^2} - \frac{\xi\nu_j^2}{\Delta_j}. \end{aligned} \tag{14}$$

For finite $\theta = \frac{i\omega}{\nu_m}$ values and small $\frac{a_1}{\omega}$, Eq. (9) has the solutions

$$\begin{aligned} \nu_1^2 &= k^2 - \frac{\omega^2}{a^2} + \frac{a_1^2k^2}{a^2}\left(1 + \frac{\omega^2}{a^2\theta}\right) - \frac{a_1^4k^4}{a^2\omega^2\xi}, \\ \nu_2^2 &= k^2 - \frac{\omega^2}{b^2} + \frac{a_1^2k^2}{b^2}\left(1 + \frac{\omega^2}{b^2\theta}\right) + \frac{a_1^2\omega^2}{b^4} - \frac{a_1^4k^4}{a^2\omega^2\xi}. \end{aligned} \tag{15}$$

It can be shown that the terms with $\frac{\omega^2}{\theta}$ and $\frac{a_1^4}{\omega^2}$ do not contribute to dispersion equation (13).

In the three-dimensional formulation for $\sigma \gg 1$, dispersion equation (13) gives $\omega = \omega_1^0 + i\omega_2^0$, where

$$\omega_1^0 = \sqrt{\frac{h^2}{3}b^2k^4\xi - \frac{2a_1^2k^2b^2}{a^2\xi}}, \quad \omega_2^0 = -\frac{a_1^2k^2}{h\omega_1^0\sqrt{2\frac{\omega_1^0}{\nu_m}}}. \quad (16)$$

For large finite σ values, Eq. (13) gives

$$\omega^2 = \frac{h^2}{3}b^2k^4\xi - \frac{2a_1^2k^2b^2}{a^2\xi} - a_1^2k^2\frac{2\tanh\nu_3\frac{h}{2}}{\nu_3 h}. \quad (17)$$

In the case $\sigma \gg 1$, we have $\tanh\nu_3\frac{h}{2} \approx 1$ and, since

$$i\frac{\nu_3^2}{\omega}\nu_m = \frac{a_1^2}{b^2} + 1, \quad (18)$$

Eqs. (16) follow from Eq. (17).

In the case of finite θ values, we have $a_1^2 = \frac{H_0^2}{4\pi\rho}$, and Eq. (17) provides

$$\omega^2 = \frac{h^2}{3}b^2k^4\xi - a_1^2k^2\frac{1 + \frac{b^2}{a^2}}{\xi}. \quad (19)$$

The direct substitution of Eq. (14) into Eq. (13) leads to the same value.

For comparison, we consider the averaged approach, where [1–3]

$$u_z = U(x,t), \quad u_x = -z\frac{\partial u}{\partial x}, \qquad U = \frac{1}{2}(A\exp(i\tau) + \text{c.c.}). \quad (20)$$

By solving the equations for the plates and the dielectric, we obtain

$$Dk^4 - \rho h\omega^2 = -i\frac{H_0^2}{4\pi} \times k^2\left(\frac{k^2h^3}{12} + 2\frac{\lambda_1^2 - k^2}{\lambda_1^3}\frac{\lambda_1\frac{h}{2}\cosh\lambda_1\frac{h}{2} - \sinh\lambda_1\frac{h}{2}}{\cosh\lambda_1\frac{h}{2} + \frac{k}{\lambda_1}\sinh\lambda_1\frac{h}{2}}\right)\frac{\omega}{i\omega - k^2\nu_m}. \quad (21)$$

Here,

$$Dk^4 = \frac{h^2}{3}b^2k^4\xi, \quad \lambda_1 = \left(k^2 - \frac{i\omega}{\nu_m}\right)^{1/2}.$$

For $\sigma \gg 1$, we have $\lambda_1\frac{h}{2} \gg 1$ and Eq. (21) has the solutions

$$(\omega_1^0)^2 = \frac{1}{\rho h}\left(Dk^4 + \frac{H_0^2}{4\pi}k^2h\right), \qquad \omega_2^0 = -\frac{H_0^2k^2}{4\pi\sqrt{2\nu^{-1}}\rho h(\omega_1^0)^{3/2}}. \quad (22)$$

The ω_2^0 value coincides with the second of Eqs. (16), but the ω_1^0 value differs qualitatively and quantitatively from the first of Eqs. (16) for the three-dimensional approach. Taking into account that $\frac{\lambda_1 h}{2}$ is small, Eq. (21) for finite θ value leads to

$$Dk^4 - \rho h\omega^2 = \rho a_1^2k^2h^4(\nu_m)^{-1}. \quad (23)$$

In the averaged formulation, Eq. (23) for finite θ results in

$$(\omega_1^0)^2 = \frac{Dk^4}{\rho h} - \frac{3}{4}a_1^4k^4h^4(\nu_m)^{-1}. \quad (24)$$

By repeating the above calculations in the case of the longitudinal field $H_0 = H_{0x}$ in the three-dimensional and averaged formulations for $\sigma \gg 1$, we obtain [7]

$$(\omega_1^0)^2 = \frac{h^2}{3}b^2k^4\xi + \frac{2a_1^2k}{h}. \quad (25)$$

Thus, unlike the case of the transverse field $H_0 = H_{0z}$, the three-dimensional and averaged approaches provide identical results for large σ. In the case of the

longitudinal field and finite σ, the three-dimensional approach provides

$$(\omega_1^0)^2 = \frac{h^2}{3}b^2k^4\xi + a_1^2k^2\left(1 + \frac{2b^2}{a^2\xi}\right), \quad (26)$$

whereas the averaged approach gives

$$\omega^2 = \frac{Dk^4}{\rho h} - \frac{2a_1^2\omega k\nu_m^{-1}i}{2k - i\omega\nu_m^{-1}h}. \quad (27)$$

As is seen, Eqs. (26) and (27) differ from each other.

To verify the theoretical results, we carried out experiments. Since the dispersion relation $\omega = \omega(k)$ is independent of the form of the boundary conditions at the plate edges [2], we used cantilever plates that were made of either aluminum or brass and were l = 34 cm long, 6 cm wide, and 1–2 mm thick. The longitudinal and transverse magnetic-field intensities were 0.05 and 0.5 T, respectively.

The second harmonic of a frequency of 600 Hz in the elastic problem was resonant. In all the experiments, the transverse field decreases the frequency by 3–5%, whereas the longitudinal field increases the frequency by 4–5%. These results agree with Eqs. (11), (15), and (17), which were obtained for the three-dimensional formulation but are inconsistent with Eqs. (14) and (16) for the averaged approach.

ACKNOWLEDGMENTS

I thank A.G. Bagdoev, Corresponding Member of the Academy of Sciences of Armenia, S.G. Saakyan, and A.A. Vantsayn for valuable advice.

REFERENCES

1. S. A. Ambartsumyan, G. E. Bagdasaryan, and M. V. Belubekyan, *Magnetoelasticity of Thin Shells and Plates* (Nauka, Moscow, 1977).
2. S. A. Ambartsumyan and G. E. Bagdasaryan, *Current-Conducting Plates and Shells in Magnetic Field* (Nauka, Moscow, 1996).
3. S. A. Ambartsumyan and M. V. Belubekyan, *Vibration and Stability of Current-Carrying Elastic Plates* (Izdat. Nat. Akad. Nauk Armenii, Erevan, 1992).
4. A. G. Bagdoev and L. A. Movsisyan, Izv. Akad. Nauk Arm. SSR, Mekh. **35**, 16 (1982).
5. A. G. Bagdoev and L. A. Movsisyan, Izv. Akad. Nauk Arm., Mekh. **52**, 25 (1999).
6. W. Nowacki, *Theory of Elasticity* (PWN, Warsawa, 1983; Mir, Moscow, 1975).
7. A. G. Bagdoev and S. G. Sahakyan, Informats. Tekhnol. Upravlenie, No. 2, 95 (1999).
8. A. G. Bagdoev and A. A. Vantsyan, Int. J. Solids Struct. (2001).
9. A. G. Bagdoev, A. A. Vantsyan, and Yu. S. Safaryan, Informats. Tekhnol. Upravlenie, No. 1, 155 (2001).

Translated by Yu. Vishnyakov