Doklady Physics, Vol. 49, No. 10, 2004, pp. 549–552. Translated from Doklady Akademii Nauk, Vol. 398, No. 5, 2004, pp. 611–614.

PHYSICS

A New Method of Producing Subnanosecond High-Current Electron Beams

S. B. Alekseev, V. P. Gubanov, V. M. Orlovskiĭ, V. S. Skakun, and V. F. Tarasenko

Presented by Academician S.D. Korovin March 30, 2004

Received April 8, 2004

The traditional method of producing subnanosecond electron beams is based on the formation of subnanosecond voltage pulses by means of vacuum diodes [1]. This method makes it possible to produce high-current electron beams with a duration of tenths of a nanosecond. However, it requires the creation of comparatively complex facilities with several transmission lines and high-pressure dischargers. In the late 1960s, in studies of pulse discharges at atmospheric pressure in air [2] and in helium [3], the appearance of X-ray radiation was discovered that was related to the deceleration of runaway electrons at the discharger anode. However, the amplitudes of current pulses of runaway electrons under atmospheric pressure in air attained only fractions of an ampere (see review [4]). These parameters were much smaller than those obtained by the traditional method of producing subnanosecond electron beams [1] and were of no interest for practical applications.

In 2002–2003, at the Institute of High-Current Electronics, Siberian Division, Russian Academy of Sciences, a series of experimental studies [5–10] was performed devoted to the formation of electron beams in gas diodes filled with molecular and atomic gases, as well as with mixtures thereof. In the course of these experiments, a significant current amplitude in the beam beyond the discharger output-window foil was registered (~70 and ~200 A in air and helium, respectively). The full width at the half-maximum (fwhm) of the current-pulse duration was ~0.3 ns. In [8], it was proposed to refer to a subnanosecond electron beam in air with an amplitude of several tens of amperes as a SAEB (subnanosecond avalanche electron) beam. The value of the parameter $E/p = U/dp$ averaged over the diode gap was ~0.1 kV cm^{-1} torr^{-1}, where U is the voltage across the gap, d is the length of the interelectrode gap, and p is the gas pressure. This parameter was much smaller than the critical value E_{cr}/p required to form runaway electrons [11]. In order to explain these results, it was assumed in [5, 7, 8] that the critical field is attained in the near-anode region upon expanding the plasma cloud from the cathode to the anode. However, the potentialities of this method of forming subnanosecond electron beams in gas diodes remained unclear. Therefore, additional investigations were needed, including those with a recording-system time resolutions on the order of ~0.1 ns.

In this study, we demonstrate the promising potentialities of a new method for producing subnanosecond electron beams. We also investigate the possibility of forming an electron beam at gas-diode pressures exceeding atmospheric pressure and clarify conditions under which the beam amplitude becomes maximal.

In carrying out our investigations, we used three pulse generators forming voltage pulses with a duration of several nanoseconds and with a 0.5-ns (or less) front duration, as well as gas diodes of different designs. The gas diodes had a flat anode and a small-size cathode that provided additional enhancement of the electric field near the cathode. The SINUS pulse generator *1* [12] had an additional built-in transmission line with a wave resistance of 40 Ohms. The possibility of increasing the pressure in the gas diode up to 6 atm also existed. At the 40-Ohm matched load, this modernized generator formed pulses with an amplitude of ~180 kV and a fwhm duration of ~1.5 ns with a ~0.5-ns pulse-front. The first cathode consisted of three cylinders embedded into each other. Cylinders 12, 22, and 30 mm in diameter were made of 50-μm Ti foil [7].

The RADAN-220 generator (*2*) [13] had a wave resistance of 20 Ohms and produced (in the discharge gap) pulses of an amplitude up to 220 kV and of ~2-ns fwhm duration (with a pulse front of ~0.3 ns). The second cathode was made of a steel tube 6 mm in diameter and the wall thickness of 50 μm. The third cathode was made of steel, ball-shaped, and had a diameter of 9.5 mm.

Generator *3* was similar to generator *2* but had a smaller pulse duration of ~1 ns and, under the given conditions, its gas diode had minimal sizes (Fig. 1). The experiments with generator *3* were performed with the second cathode. The anode for all three generators was

Institute of High-Current Electronics, Siberian Division, Russian Academy of Sciences, pr. Akademicheskiĭ 4, Tomsk, 634055 Russia

1028-3358/04/4910-0549$26.00 © 2004 MAIK “Nauka/Interperiodica”

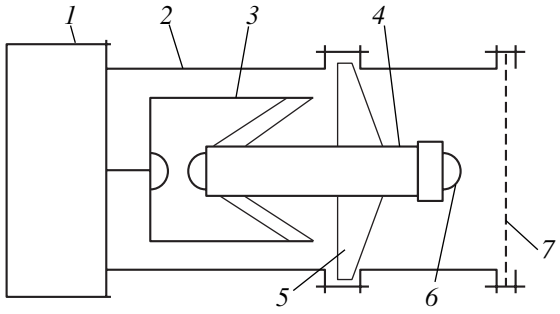


Fig. 1. Block diagram of an electron accelerator with a gas diode: (*1*) high-voltage pulse generator, (*2*) mainframe, (*3*) spark gap, (*4*) high-voltage output, (*5*) insulator, (*6*) cathode, and (*7*) anode.

made of 40-μm AlBe foil. The length of the gas gap varied from 4 to 28 mm.

Signals from the capacitive voltage divider, T collectors, and shunts were recorded by a TDS-7405 oscilloscope (4-GHz band) having 20 GS/s (20 dots per ns) or by a TDS-334 oscilloscope (0.3-GHz band) having 2.5 GS/s (2.5 dots per ns). The discharge glow was photographed by a digital photographic camera.

As has been shown in [5–10], in an inhomogeneous electric field with a small-size cathode (in the case of a short front of the voltage across the gap in different gases), an electron beam with an amplitude of several tens or hundreds of amperes is formed. In this case, volume discharge in the form of a cone or several cone-shaped streams with bright spots on the cathode is observed in the gap.

In this study, the beam-current amplitude in air was elevated by a factor of three and the range of pressures under which the electron beam had been formed was considerably extended. In particular, for the first time, an electron beam was obtained at a pressure exceeding atmospheric pressure. Figure 2 shows the pressure dependences of the voltage-pulse amplitude across the gap, the beam-current density, and the fwhm duration of the beam-current pulse obtained for helium and nitrogen. As is seen, the amplitude of the beam-current pulse and its fwhm duration remain virtually unchanged with the rise of helium pressure up to 6 atm. A decrease in the beam current in nitrogen at a pressure of 2–4 atm is related to the decrease in the region occupied by the volume discharge and to the onset of the discharge contraction at a pressure of ~4 atm. The fwhm durations of the beam-current pulse, which were measured with a resolution down to 0.1 ns M in both helium and nitrogen, differ insignificantly and were ~0.2 ns (Figs. 2 and 3, respectively). It is established that the measured duration of the beam-current pulse depends on the collector diameter and is at its smallest if the collector has a small size. It seems likely that the plasma expanding from the cathode approaches the anode at different points at different times. Accordingly, the time needed to reach the critical field has a spread T and the pulse duration of the total beam current from the entire anode area is greater than that from a portion of the anode area. In the optimal regime, the beam electrons have an average energy of ~60% of the energy corresponding to the maximum voltage across the gap. All these data confirm the assumption made in [5, 7, 8] that the formation of the electron beam occurs between the front of the plasma expanding from the cathode and the anode after the critical field has been attained.

It was found that the beam-current amplitude beyond the foil sharply depends on the gas-diode size (to be more precise, on its inductance). For generator *2* with the third cathode, the beam-current amplitude was ~170 A. The experiments were performed in air at atmospheric pressure and for the minimum foil–insula-

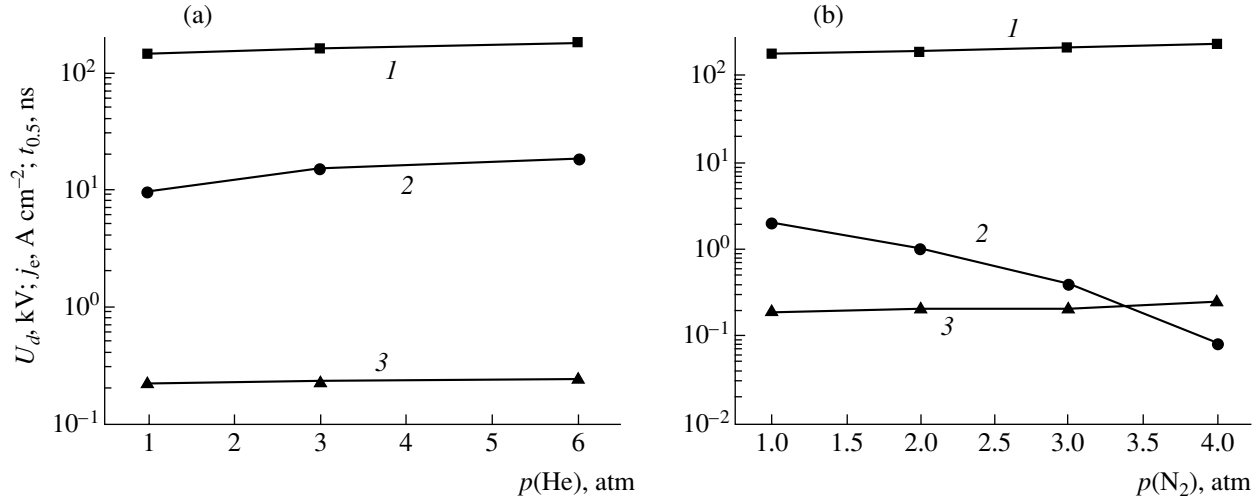


Fig. 2. Pressure dependences of (*1*) voltage-pulse amplitude, (*2*) electron-beam current density, and (*3*) fwhm duration of beam current for (a) helium and (b) nitrogen (curves are obtained with generator *1*).

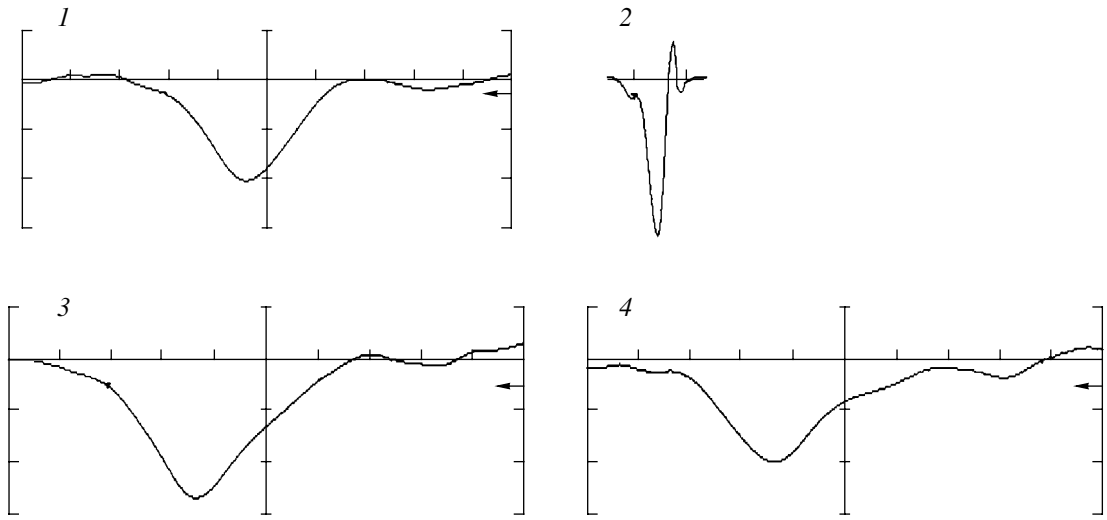


Fig. 3. Oscillograms of beam-current pulses recorded from a collector 1 cm^2 in area: (*1*) and (*2*) nitrogen at a pressure of 1 atm; (*3*) and (*4*) helium at pressures of 1 and 3 atm, respectively. Scale in the horizontal direction: (*1*), (*3*), and (*4*) 0.1 ns per division and (*2*) 0.5 ns per division. Scale in the vertical direction: (*1*) 1 A per division; (*2*) 14 A; (*3*) 3.7 A; (*4*) 7.8 A; (*1*), (*3*), and (*4*) generator *1* and (*2*) generator *3*.

tor distance. The design of the gas diode is described in detail in [10]. After moving off the foil from the isolator by 15 mm with the help of an additional metallic cylinder and correspondingly increasing the length of the cathode holder (providing the retained cathode–anode distance), the current amplitude in the electron beam decreased by a factor of two. With further increasing the size (inductance) of the gas diode, the amplitude of the beam current continued to rapidly decrease.

In the case of generator *3* (Fig. 1), which had minimum inductance of the discharge circuit, gas diode, and sharpening discharger, the beam-current amplitude beyond the foil and the fwhm pulse duration were 240 A and ~0.2 ns, respectively (Fig. 3, oscillogram *2*). The spot of the beam on the external surface of the foil was ~4 cm in diameter. At the central part of the spot (1.5 cm in diameter), the beam-current density was ~40 A cm^{-2}. It depended on the cathode design and the diameter of the volume-discharge region near the anode. For generator *1*, the average electron energy in the beam was on the order of 70 keV at a helium pressure of 1–6 atm.

We believe that very small pulse amplitudes of the beam current, have been obtained before (see [4]), are stipulated by a nonoptimal design of the gas diodes used. In the present study, the amplitude of the beam current beyond the foil was increased by several orders of magnitude compared to the amplitudes in previous investigations [4]. By comparing current amplitudes in subnanosecond beams produced by both the traditional [1] and proposed methods, we can see that, at a similar initial voltage, they differ insignificantly. However, the accelerators with gas diodes are much more simpler and form subnanosecond beams in diverse gaseous media at various pressures.

We may interpret the results obtained in the following manner. On applying a voltage pulse with a front-duration of 1 ns or less to a discharge gap, the electric field on a cathode with a small curvature radius increases. Initially, this occurs because the field increases by the mechanism of cold emission, which can be amplified by a positive ion charge, and, later, the delivery of electrons into the gap occurs by explosive electron emission. In this case, the field enhancement near the cathode is sufficient to form fast electrons. They provide preionization of the gap, formation of volume discharge, and expansion of plasma to the anode at a rate of ~10^9 cm s^{-1} and higher. In this case, some of the electrons at the boundary of expanding plasma are accelerated because of both positive voltage on the anode and pushing of electrons from the electron cloud. This mechanism leads to the formation of a subnanosecond electron beam. The leading front of the current pulse is determined by that of the voltage pulse with a subnanosecond duration. Furthermore, due to decreasing the plasma–anode distance and additionally increasing the electric field at the voltage-pulse front between the plasma and anode, the duration of the leading front of the beam-current pulse turns out to be shorter than that of the voltage pulse. The subnanosecond duration of the trailing front of the beam-current pulse is specified by two factors: a high rate of plasma propagation after attaining the critical field near the anode and equalizing the electric field in the gap after plasma has arrived at the anode.

Thus, the investigations performed have demonstrated that the new method of the formation of subnanosecond electron beams in gas diodes (SAEB beams) has attractive potentialities. In realizing this method, it is sufficient to use a single fast switch con-

nected to a cathode under the condition of a small inductance. To obtain the maximum beam current, we need to provide volume discharge in a gas diode and have a smooth anode surface. In addition, the voltage rise in the spark gap must be terminated before the maximum beam current has been attained. It is also very important to minimize the size of the gas diode. If helium or nitrogen pressures increase to 6 and 3 atm, respectively, the conditions under which the subnanosecond electron beams are formed are retained and the above pressures are not ultimate. This is stipulated by the formation of an electron beam on attaining the critical field between the anode and plasma expanding from the cathode. By optimizing the design of both the gas diode and high-voltage pulse generator, an amplitude of the subnanosecond beam current exceeding 240 A (with a beam current density of ~40 A cm^{-2}) was obtained at atmospheric pressure in air.

We assume that subnanosecond high-current electron beams produced in gas diodes will be widely applied in various fields of science and technology. In particular, we used the SAEB beam to form volume discharge in CO_2 lasers [14].

REFERENCES

1. K. A. Zheltov, *Picosecond High-Current Electron Accelerators* (Énergoatomizdat, Moscow, 1991).
2. Yu. É. Stankevich and V. G. Kalinin, Dokl. Akad. Nauk SSSR **177**, 72 (1967) [Sov. Phys. Dokl. **12**, 1042 (1967)].
3. R. C. Noggle, E. P. Krider, and J. R. Wayland, J. Appl. Phys. **39**, 4746 (1968).
4. L. P. Babich, T. V. Loĭko, and V. A. Tsukerman, Usp. Fiz. Nauk **160**, 49 (1990) [Sov. Phys. Usp. **33**, 521 (1990)].
5. S. B. Alekseev, V. M. Orlovskiĭ, and V. F. Tarasenko, Pis'ma Zh. Tekh. Fiz. **29** (10), 29 (2003) [Tech. Phys. Lett. **29** (5), 411 (2003)].
6. S. B. Alekseev, V. M. Orlovskiĭ, V. F. Tarasenko, *et al.*, Pis'ma Zh. Tekh. Fiz. **29** (16), 45 (2003) [Tech. Phys. Lett. **29** (8), 679 (2003)].
7. S. B. Alekseev, V. P. Gubanov, V. M. Orlovskiĭ, *et al.*, Prib. Tekh. Éksp., No. 4, 81 (2003) [Instrum. Exp. Tech. **46**, 505 (2003)].
8. V. F. Tarasenko, V. M. Orlovskiĭ, and S. A. Shunaĭlov, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 3, 94 (2003).
9. V. F. Tarasenko, S. I. Yakovlenko, V. M. Orlovskiĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 737 (2003) [JETP Lett. **77**, 611 (2003)].
10. V. F. Tarasenko, V. G. Shpak, S. A. Shunaĭlov, *et al.*, Pis'ma Zh. Tekh. Fiz. **29** (21), 1 (2003) [Tech. Phys. Lett. **29**, 879 (2003)].
11. A. N. Tkachev and S. I. Yakovlenko, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 264 (2003) [JETP Lett. **77**, 221 (2003)].
12. V. P. Gubanov, S. D. Korovin, I. V. Pegel', *et al.*, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 12, 110 (1996).
13. M. I. Yalandin and V. G. Shpak, Prib. Tekh. Éksp., No. 4, 5 (2001) [Instrum. Exp. Tech. **44** (3), 285 (2001)].
14. S. B. Alekseev, V. M. Orlovskiĭ, and V. F. Tarasenko, Kvantovaya Élektron. (Moscow) **33**, 1059 (2003) [Quant. Electron. **33**, 1059 (2003)].

Translated by Yu. Vishnyakov

Doklady Physics, Vol. 49, No. 10, 2004, pp. 553–555. Translated from Doklady Akademii Nauk, Vol. 398, No. 6, 2004, pp. 753–755.

PHYSICS

Manifestation of Optical Quadratic Nonlinearity in Gas Mixtures

V. P. Torchigin and A. V. Torchigin

Presented by Academician A.F. Andreev May 19, 2004

Received April 12, 2004

In optical media, particularly in gases, there exist several types of quadratic optical nonlinearity. This phenomenon is manifested in a nonlinear (quadratic) dependence of the refractive index n of the optical medium on the light wave amplitude. The best known is the Kerr optical nonlinearity, for which the increase in the refractive index is associated with orienting molecules in the light wave field. The other example is the phenomenon of electrostriction, in which the light wave compresses the medium and elevates its density, thereby increasing n [1]. It turns out that, in gas mixtures, there exists one more mechanism of manifesting optical nonlinearity. We are implying the fact that the molecules of a gas with the highest n are attracted to the region of the most intense light. As a result, the refractive index of the mixture is higher compared to the equilibrium mixture in the absence of optical radiation. Thus, quantity n inside the radiation field is dependent on the radiation intensity. In the case of increasing n, the work spent to separate molecules in a gas mixture can be smaller than that spent for the same increase in n by means of compression of the same gas mixture by the electrostriction mechanism. Therefore, the effect observed can be even stronger than in the case of electrostriction.

As is well known, the separation of a mixture is associated with an energy consumption determined by the expression $W = T\Delta S$, where T is the temperature of the gas mixture and ΔS is the rise of the entropy when the gases are mixed [2]. If this process is adiabatic, the increment of entropy is zero. Therefore, when the heat inflow to the system under consideration is negligible, the work needed for separation of the gas mixture is close to zero, and the nonlinearity coefficient of the gas mixture can be rather high. However, the time of occurring transition processes accompanied by the suction of molecules of a gas with the highest n is also relatively long, because it is associated with transport phenomena. Therefore, in the case of instantaneous appearance of radiation, e.g., in a gas discharge, there is no domain from which molecules can be sucked, since the radiation is present in the entire volume of the mixture. However, this does not imply that the effect of the indicated mechanism is not manifested.

In the same manner as the condensation of saturated vapor can be accompanied by the appearance of a fog consisting of small droplets of liquid, the molecules of a gas with the highest n are concentrated by intense radiation in a local domain. In contrast to droplets of liquid—where molecules occupy the entire droplet volume—in a gas mixture, molecules are concentrated in a thin spherical layer (TSL) having an increased refractive index. The TSL plays the role of a bent planar light guide directing the light that circulates in it over all possible paths. Intense light, in turn, provides the concentration in the TSL of molecules with the highest n. Thus, in a gas mixture, intense light condenses as though into light fog consisting of a set of different-size TSLs.

Figure 1 illustrates the stages of the appearance of these formations. We assume that, in a certain domain, a fluctuation of the density and/or of the concentration has occurred. In this case, the boundary of the domain with an elevated refractive index is convex. Then, in accordance with the eikonal equation, a beam propagating along the tangent to the boundary is bent toward the maximum refractive index. The radius of curvature is determined by the expression $R^{-1} = dn/dr$, where r is the distance along the straight line perpendicular to the beam. For example, if n varies by 0.25×10^{-4}, at the distance of 1 μm, then $R = 4$ cm. In this case, the bent beam shown in Fig. 1a propagates in the boundary domain for a longer time and for a grater distance compared to the rectilinear beam. Thereby, the conditions needed for separating a gas mixture are more favorable than in other spatial domains. This results in the fact that, at the boundary, the quantity $\left|\frac{dn}{dr}\right|$ increases, which, in turn, leads to a further rise of the beam rotation angle as is shown in Fig. 1b. As long as this process is continued, the beam rotation angle increases and reaches 360°. For performing further analysis, it is reasonable to pass

Institute of Informatics Problems,
Nakhimovskiĭ pr. 36/1, Moscow, 119208 Russia
e-mail: v.torchigin@mail.ru

1028-3358/04/4910-0553$26.00

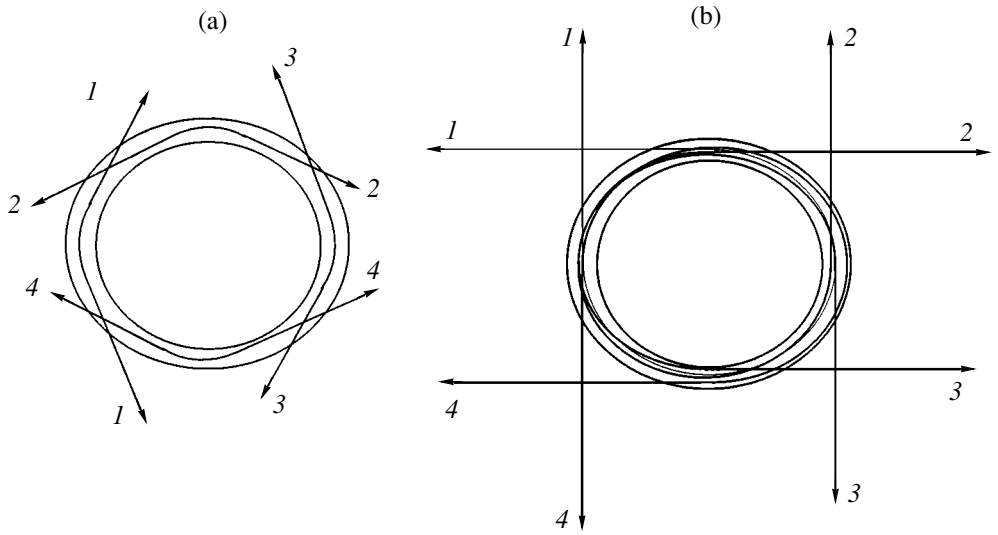


Fig. 1. Trajectory of light beams at different stages of the formation of a thin spherical layer: (a) initial stage, (b) intermediate stage. The onset and the end of the same beam are denoted by equal numbers.

from a beam approach to a wave one. In so doing, we may state that a TSL has been formed in which a wave of the whispering-gallery type circulates [3]. The insufficiently large difference in refractive index n between the TSL and the environment results in a large radiation loss in the TSL. However, as follows from the reciprocity principle, here, we are dealing with a reasonably high coupling factor between plane waves in the environment and the whispering-gallery wave in the TSL. In other words, light omnidirectionally propagating in a gas mixture excites whispering-gallery waves in the TSL rather efficiently. When, as a result of the excitation, the light intensity in the TSL increases, the difference in n between the TSL and the environment enhances. In this case, the radiation loss and the coupling factor between the whispering-gallery waves and the environment decrease. Ultimately, omnidirectionally circulating radiation is established in the TSL. The intensity of this radiation is equal to the intensity of light in the environment. (It should be noted that the TSL remains transparent for beams incident on its surface at arbitrary angles apart from those tangent to the surface.) At a reasonably high intensity of light circulating in the TSL, self-compression of the TSL thickness is possible [4]. This effect is similar to the well-known effect of spatial-soliton formation in a nonlinear optical medium [5] and results in a noticeable increase in pressure and refractive index n inside the TSL. As was shown in [6], in this case, the light scattering decreases while the light lifetime increases, the latter attaining dozens of seconds. In the case of the disappearance of light in the environment (e.g., in the case of ceasing the electric current in a gas discharge), light presenting in the gas mixture leaves it at the light velocity. However, light accumulated in the TSL continues to circulate in it.

There is no need to perform specified additional experiments to confirm the above-presented pattern of manifesting the nonlinear effect under consideration. The more than two-year history of studies of unusual autonomous objects arising as a result of electric discharge in gases presents a large number of experimental data that, unfortunately, cannot be comprehensively discussed in the framework of a single paper. A sufficiently complete review of these studies is given in [7, 8]. Here, we comment only on results of some recent investigations, which testify to the correctness of the presented pattern.

The spectral composition of autonomous objects obtained as a result of high-frequency discharge in different gases at atmospheric pressure is given in [9]. The following features of the spectra are noted. In all the gases, the visible spectrum of the discharge was emitted by admixtures and did not correspond to the spectrum of the gas in which autonomous objects had arisen. The spectrum of visible light basically corresponded to the CO_2-spectrum and involved several spectral lines of the electrode metal. The discharge color also indicated the formation of nitrogen dioxide; however, it was impossible to distinguish its spectrum against the CO_2-spectrum background. It is worth noting that n_{CO_2} = 1.0004197 and n_{NO_2} = 1.000515, whereas the refractive indices of air, nitrogen, and oxygen for which the investigations were performed lay within the range 1.00025–1.00028. Thus, radiation of the components possessing higher refractive indices n was present in the spectrum.

The explanation of properties of the so-called power-consuming plasma formations presented in [10] is rather simple. These objects are formed in a gas discharge. They have a density close to that of the sur-

rounding air, low gas temperature, weak radiation intensity, and high energy density and are distinguished by selective action on materials (they are capable of burning through metallic foil, but do not act on paper). They also strive to preserve their integrity upon meeting an obstacle, and their lifetime is anomalously long compared to ideal plasma. In fact, these power-consuming plasma formations are instances of miniature ball lightning, the parameters and behavior in the inhomogeneous atmosphere of which are considered in [4, 6], where the hypothesis is substantiated that the ball lightning is a TSL in which intense light circulates. The similarity in the behavior of the power-consuming plasma formations and of ball lightning is also noted in [10]. However, the assumption that these formations are to a certain extent associated with long-living plasma did not allow the authors of [10] to explain the characteristic features of these objects.

The experiments carried out in [11] on interaction of anomalous objects with liquid nitrogen turned out to be rather impressive. The anomalous objects arising as a result of electric discharge move toward the liquid-nitrogen surface. Upon finishing the discharge, blue luminescence of the entire volume of liquid nitrogen was observed over 5 s. Anomalous sphere-shaped sharp-boundary objects 0.5–4 mm in diameter were found on the vessel bottom, on the surface of the liquid nitrogen, and in the entire volume of the liquid nitrogen. The brightness of these objects considerably exceeded the total light background of the liquid nitrogen. The luminescence of the autonomous objects had a spectrum lying within the range 400–500 nm and lasted for 10–30 s, until smooth quenching occurred.

The penetration of anomalous objects into liquid nitrogen can be explained by the following reasons. TSLs move in the direction of the refractive-index gradient of the environment. The refractive index of gaseous nitrogen is $n_1 = 1.000277$ and practically coincides with that of air. Refractive index n_2 of cold gaseous nitrogen in the vicinity of the liquid-nitrogen surface is much higher: $n_2 \approx 1.000800$. Therefore, TSLs move toward the liquid-nitrogen surface. Approaching this surface, the TSLs evaporate nitrogen. In this case, the temperature of the interlayer between an anomalous object and liquid nitrogen is close to that of liquid nitrogen, and the refractive index of evaporated nitrogen considerably exceeds that of air at normal conditions. As a result, anomalous objects, tending to move into a domain with the maximum refractive index, i.e., into the coldest domain, form a depression on the liquid-nitrogen surface with a gradually increasing depression depth. Eventually, an anomalous object completely penetrates deep into the liquid nitrogen and forms a gaseous-nitrogen layer around itself. Apparently, the luminescence of the entire nitrogen is explained by the luminescence of a set of anomalous objects of rather small diameters. Such small anomalous objects have a large radiation loss (the loss increases with decreasing the sphere diameter) and, hence, a shorter lifetime (5 s) compared to large ones (30 s).

Thus, in numerous experiments, instances of ball lightning were actually obtained. Their short lifetime is explained by the short lifetime of the light circulating inside them. In order to increase the lifetime, it is necessary to elevate the radiation intensity in a gas mixture at the last instant prior to cutting off the radiation. The radiation duration should be sufficient to considerably change the mixture composition in a thin spherical layer during this time.

REFERENCES

1. G. S. Landsberg, *Optics* (Nauka, Moscow, 1976).
2. V. V. Nashchokin, *Technical Thermodynamics and Heat Transfer* (Vysshaya Shkola, Moscow, 1969).
3. A. N. Oraevskiĭ, Kvantovaya Élektronika **32**, 377 (2002).
4. V. P. Torchigin and A. V. Torchigin, Phys. Scr. **68**, 388 (2003).
5. V. E. Zakharov and A. B. Shabat, Zh. Éksp. Teor. Fiz. **61**, 118 (1971).
6. V. P. Torchigin, Dokl. Akad. Nauk **389** (1), 41 (2003) [Dokl. Phys. **48**, 108 (2003)].
7. S. Singer, *Nature of Ball Lightning* (Plenum Press, New York, 1971; Mir, Moscow, 1973).
8. J. D. Barry, *Ball Lightening and Bead Lightening: Extreme Forms of Atmospheric Electricity* (Plenum Press, New York, 1980; Mir, Moscow, 1983).
9. J. R. Powel, M. S. Zucher, J. F. Manwaring, *et al.*, Bull. Am. Phys. Soc. **12**, 751 (1967).
10. R. F. Avramenko, V. I. Nikolaeva, and L. P. Poskacheeva, in *Ball Lightning* (Khimiya, Moscow, 1994), pp. 15–86.
11. S. K. Dimitrov, S. K. Zhdanov, D. L. Kirko, *et al.*, in *Ball Lightning* (Khimiya, Moscow, 1994), pp. 79–87.

Translated by G. Merzon

Doklady Physics, Vol. 49, No. 10, 2004, pp. 556–558. Translated from Doklady Akademii Nauk, Vol. 398, No. 4, 2004, pp. 469–471.

TECHNICAL PHYSICS

Characteristics of Shock-Wave Loading of AMg6M Aluminum Alloy under Spall Conditions

V. I. Vovchenko, I. K. Krasyuk,
Corresponding Member of the RAS P. P. Pashinin, and A. Yu. Semenov

Received March 30, 2004

Previous dynamic-strength investigations of AMg6M aluminum alloy [1–5] showed that the spallation phenomenon [6, 7] can occur with various regimes of dynamic stress. They can be conditionally called (i) dynamic, (ii) quasi-stationary [4], and (iii) intermediate-type stress regimes.

These regimes can be classified by using the parameter $\xi = \frac{t_{sp}}{t_m}$. Here, t_{sp} and t_m are the times when the tensile stress σ reaches the spall value σ_{sp} and the peak value, respectively (Fig. 1). In the first case, the material breaks down at $\xi \le 1$ for stress increasing linearly with time. In the second case, the material is rapidly stretched and breaks down at almost constant stress and $\xi \ge 2$. The third, intermediate, case is a combination of two stress stages lasting comparable times. At the first stage, stress increases linearly with time, and at the second stage, stress is almost constant up to the time of the separation of a spall layer from the target at $1 < \xi < 2$. A regime in which spallation occurs earlier under given conditions of the shock-wave loading of a sample is realized in an experiment.

In this paper, we present the new results of processing measurements of the characteristics of the dynamic-stress regime at $\xi \le 1$. These results were obtained by numerical simulation using the experimental data taken from [1–4]. In order to find the stress form in the spall section, the numerical hydrodynamic code [8, 9] including the actual wide-range equation of state for aluminum was applied. The initial density of the substance was assumed to be 2.61 g cm^{-3}, i.e., equal to that of AMg6M aluminum alloy under normal conditions. A more comprehensive description of applying the numerical code to the given problem can be found in [5].

Prokhorov Institute of General Physics,
Russian Academy of Sciences,
ul. Vavilova 38, Moscow, 119991 Russia
e-mail: krasyuk@kapella.gpi.ru, semenov99@rambler.ru

Using the numerical code, the mean duration Δt of stress $\sigma(t)$ acting in the spall plane was calculated from experimental data according to the formula

$$\Delta t = \frac{1}{\sigma_{sp}} \int_0^{t_{sp}} \sigma(t)\,dt.$$

Figure 2 shows the calculated mean duration Δt of stress as a function of the volume-strain rate $\frac{\dot V}{V_0} = -\frac{1}{\rho_0}\frac{d\rho}{dt}$. Here, $V = \frac{1}{\rho}$ and ρ are the specific volume and material density, respectively. The results indicate that the dependence of the mean duration of stress on the volume-strain rate is nonmonotonic and complicated. For volume-strain rates $(0.1–1.1) \times 10^7$ s^{-1}, the stress duration decreases from 11.6 to 1.6 ns in accordance with the empirical expression

$$\Delta t, \text{ ns} = 2\left(\frac{\dot V}{V_0}, 10^7\ 1/\text{s}\right)^{-0.85}.$$

Then, the stress duration increases sharply up to 10.5 ns. Within the strain-rate range $(1.5–5.7) \times 10^7$ s^{-1}, where the spall strength of the material under study reaches the maximum value of 80 kbar and remains

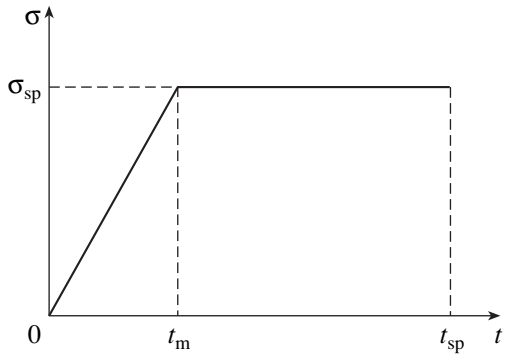


Fig. 1. Time dependence of the tensile stress in the spall plane.

1028-3358/04/4910-0556$26.00 © 2004 MAIK “Nauka/Interperiodica”

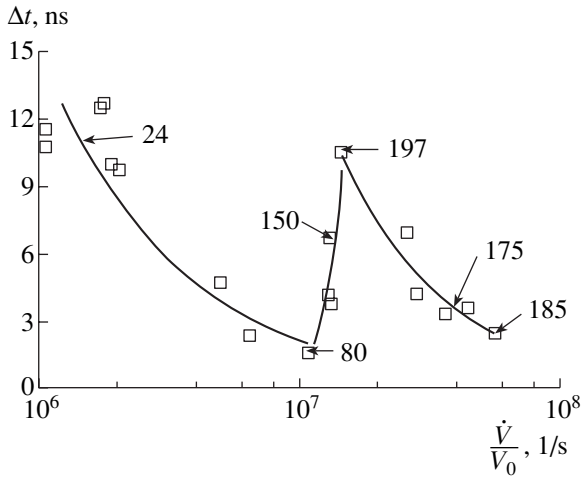


Fig. 2. Open squares: duration of the stress resulting in spallation as a function of the volume-strain rate in the dynamic regime for $\xi \le 1$. The curves represent experimental data smoothed by the least-squares method. The numbers correspond to pressure amplitudes in the spall plane at the points indicated by the arrows.

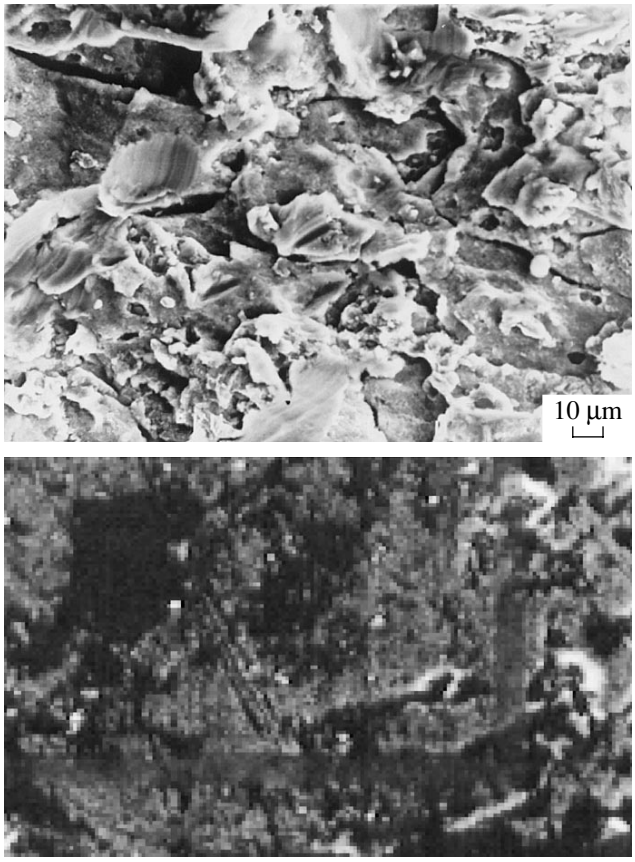


Fig. 3. Spall surface: the upper part corresponds to the coarse-grained structure observed at a volume-strain rate of about 10^6 s^{-1}, and the lower part corresponds to a fine-grained spall structure close to a mirror surface, which is observed at volume-strain rates higher than 10^7 s^{-1}.

almost constant (see also [3–5]), the stress duration decreases again down to 2.5 ns in accordance with the empirical expression

$$\Delta t, \text{ ns} = 16.2\left(\frac{\dot{V}}{V_0}, 10^7\ 1/\text{s}\right)^{-1.07}.$$

Thus, we have established the dependence of the duration of fracture stress on the volume-strain rate. This dependence characterizes the features of the kinetics of the growth and development of defects in the material under study. One of the possible causes of this dependence may be a phenomenon previously observed in [1–4]. This phenomenon is the hardening of the material under study in the region where the duration of fracture stress increases sharply and the material acquires new strength properties compared to the initial one. The material hardening is associated with the specific features of the dynamic method used to investigate strength properties. Indeed, tensile stresses responsible for the fracture of the material under study appear after the action of intense compressing loading. Therefore, we can expect variation of the number of initial defects in the material and their effect on the kinetics of the fracture process. The pressure amplitudes in the spall region, which were obtained by the numerical method with allowance for experimental conditions, are given in Fig. 2 for the corresponding points of the plot. These data show that pulsed pressure increases from 80 to 200 kbar in the volume-strain rate range $(1–1.5) \times 10^7$ s^{-1}.

Metallographic investigations using an electron microscope showed that the material structure in the spall plane depends on the material volume-strain rate. For volume-strain rates lower than 10^7 s^{-1}, the material has a granular structure in the fracture zone with specific grain sizes of 5–30 μm (upper part of Fig. 3). For volume-strain rates exceeding 1.5×10^7 s^{-1}, when the spall strength is close to the maximum value of 80 kbar, the view of the spallation zone is close to a mirror surface without any apparent structure (lower part of Fig. 3). In this case, the process of material fracture depends mainly on the intrinsic properties of the material rather than on its structure and various initial defects existing in it.

The above results indicate that the mechanism responsible for the fracture of the material for high volume-strain rates is complicated, which, in particular, should be taken into account in developing and refining theoretical models of dynamic fracture of materials.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation of Basic Research, project no. 03-02-16627.

REFERENCES

1. D. Batani, I. K. Krasyuk, P. P. Pashinin, *et al.*, in *Proceedings of the International Quantum Electronics Conference Collocated with Conference on Lasers, Applications and Technologies (IQEC/LAT), Moscow,* Rep. LWE5, p. 140.
2. I. K. Krasyuk, D. Batani, V. E. Fortov, *et al.*, in *Proceedings of the 27th European Conference on Laser Interaction with Matter (ECLIM), Moscow, 2002,* Rep. Th-CO3/O2, p. 182.
3. V. I. Fortov, D. Batani, A. V. Kilpio, *et al.*, Laser Part. Beams **20**, 317 (2002).
4. D. Batani, V. I. Vovchenko, A. V. Kil'pio, *et al.*, Dokl. Akad. Nauk **389**, 328 (2003) [Dokl. Phys. **48**, 123 (2003)].
5. I. K. Krasyuk, P. P. Pashinin, A. Yu. Semenov, and V. E. Fortov, Kvantovaya Élektron. (Moscow) **33**, 593 (2003).
6. Ya. B. Zel'dovich and Yu. P. Raĭzer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Nauka, Moscow, 1966; Academic Press, New York, 1967).
7. G. I. Kanel', S. V. Razorenov, A. V. Utkin, and V. E. Fortov, *Shock-Wave Phenomena in Condensed Media* (Yanus-K, Moscow, 1996).
8. A. Yu. Semenov, Zh. Vychisl. Mat. Mat. Fiz. **37**, 1376 (1997) [Comp. Maths. Math. Phys. **37**, 1334 (1997)].
9. A. G. Kulikovskiĭ, N. V. Pogorelov, and A. Yu. Semenov, *Mathematical Aspects of Numerical Solution of Hyperbolic Systems*, Vol. 118 of *Monographs and Surveys in Pure and Applied Mathematics* (Fizmatlit, Moscow, 2001; Chapman and Hall, Boca Raton, 2001).

Translated by G. Merzon

Doklady Physics, Vol. 49, No. 10, 2004, pp. 559–563. Translated from Doklady Akademii Nauk, Vol. 398, No. 4, 2004, pp. 472–476.

TECHNICAL PHYSICS

Pulse-Shape Synthesis Using the Best-Filtration Criterion for Signals Emitted by Super-Broadband Antennas

M. Ya. Izrailovich* and Corresponding Member of the RAS L. D. Bakhrakh**

Received June 1, 2004

INTRODUCTION

In the case when radio signals emitted by super-broadband antennas pass through a medium, noticeable pulse-shape distortions are often observed. Therefore, determination of the time stricture of initial signals, which requires the minimization of the distortions, i.e., provides the best filtration, is extremely urgent. At the same time, the problem of a sufficiently high intensity of pulses both emitted by an antenna and exiting the medium should be solved.

In this paper, we describe methods of pulse-shape synthesis for signals emitted by super-broadband antennas in both the single-pulse emission and packet-operation regimes. In the latter case, we deal with the emission of sequences of periodically repeated pulses. These methods must provide optimal pulse filtration at a sufficiently high intensity level when the radiation passes through a medium. In this case, the dynamic properties of the medium can be adequately simulated in the form of time-independent linear dynamical systems.

A SINGLE-PULSE EMISSION REGIME

We assume that the shape of a super-broadband single pulse is described by the function $x(t)$. In addition, we can assume that, with sufficient accuracy, $x(t) = 0$ as $t > T_1$. In other words, we practically ignore any residual radiation as $t > T_1$. In order to reduce the time T_1, we employ the methods described in [1, 2]. In the process of passing signals through a medium, the pulse shape $x(t)$ is subjected to a certain dynamic transformation. As a result, at the output of the medium, the signal $y(t)$ appears. For a wide class of linear media with properties independent of time, the aforementioned dynamic transformation can be represented as

$$\begin{aligned} y(t) &= \int_0^t h(t-\tau)x(\tau)d\tau, \quad t \in (0; T_1), \\ y(t) &= \int_0^{T_1} h(t-\tau)x(\tau)d\tau, \quad t \in [T_1; \infty). \end{aligned} \tag{1}$$

Here, $h(t-\tau)$ corresponds to the pulsed transition function of a medium.

In order to solve the problem of optimal filtration, i.e., of minimization of dynamic distortions arising in the process of passage of a signal through a medium, we introduce the functional $I(y - x)$ depending on the difference between the initial and transformed signals. The simplest form of this functional is the integral quadratic functional

$$I = \int_0^T [x(t) - y(t)]^2 dt,$$

where T is the time of observation of the signal $y(t)$ ($T > T_1$). However, in solving the formulated problem, i.e., finding the optimum law $x_*(t)$, the use of this functional turns out to be insufficient. Indeed, the functional satisfies only one requirement, namely, minimization of the divergence between $y(t)$ and $x(t)$. In this case, no conditions concerning the signal-intensity level are imposed. By virtue of this fact, the absolute minimum of the functional, which is equal to zero, is attained for the trivial solution $x(t) \equiv y(t) \equiv 0$. Therefore, to ensure a reasonably high intensity level for signals $x(t)$ and $y(t)$, we should introduce the extended functional

$$I(x, \gamma) = \left\{ \int_0^T [y(t) - x(t)] + \gamma[x_0(t) - x(t)]^2 \right\} dt, \tag{2}$$

where $x_0(t)$ is the given time-dependent function that characterizes the necessary intensity of the signal $x(t)$. In a particular case, $x_0(t) = x$, $t \in (0; T_1)$, $x_0(t) = 0$, $t \in [T_1; \infty)$, where X is a sufficiently large constant, γ is a weight factor determining the relation between require-

* *Blagonravov Institute of Engineering Science, Russian Academy of Sciences, ul. Bardina 4, Moscow, 117334 Russia*

** *Moscow Research Institute of Instrument Building, Kutuzovskiĭ pr. 34, Moscow, 121170 Russia*

1028-3358/04/4910-0559$26.00

ments to the minimum distortion of the signal $x(t)$ and to its sufficiently high intensity.

Theorem 1. *The function $x_*(t, \gamma)$ corresponding to the solution of the problem on the absolute level of the functional $I(x, \gamma)$ (2) exhibits the following properties*: (*a*) $x_*(t, 0) = 0$ *and* (*b*) $\lim x_*(t, \gamma) = x_0(t)$ *as* $\gamma \longrightarrow \infty$.

The most efficient procedure for the determination of the function $x_*(t, \gamma)$ is based on the employment of the abstract theory of linear operators in the $L^2(0; T)$ space and on the search for conditions determining the time-independent behavior of the functional, which uses the method of calculating its Gateaux derivative [3]. In this case, the operator $y(x)$ (1) and the function (2) can be written in the form

$$y = Ax, \tag{3}$$

and

$$I(x, \gamma) = (Ax - x, Ax - x) + \gamma(x_0 - x, x_0 - x), \tag{4}$$

respectively.

As a result of determining conditions for the time-independent behavior of the functional $I(x, \gamma)$ (4) (in correspondence with the procedure described in [3]), we arrive at the following operator equation with respect to the desired function:

$$A^*Ax - A^*x + \gamma x = \gamma x_0, \tag{5}$$

where A^* is a linear operator conjugate to the operator A.

Equation (5) is a linear operator equation with the solution

$$x_*(t, \gamma) = R(\gamma)\gamma x_0, \tag{6}$$

where

$$R(\gamma) = (A^*A - A^* + \gamma)^{-1}.$$

Since Eq. (5), being written in a nonabstract form, belongs to the class of Volterra integral equations of the second kind, the operator $R(\gamma)$ entering into Eq. (6) is its resolvent.

A PACKET-EMISSION REGIME

We now consider the emission by super-broadband antennas of a packet of signals, i.e., a sequence of n periodically repeated (with a period T) signals, each having an identical pulse shape (with allowance for the delay by the time of the period T: $x_i(t) = x_{i-1}(t - T)$, $i = 2, 3, \ldots, n$). In this case, the problem of the determination of the optimum pulse shape $x_{i*}(t)$ becomes considerably more complicated. This is caused by the fact that the response of a medium to a sequence of periodically repeated signals is not a periodic function of time. (The only exclusion is the case in which the number n is sufficiently large and the process of attaining a periodic regime can be ignored.) By virtue of this fact, solving the problem of the determination of the pulse shape $x_*(t)$, which is based on the best-filtration criterion, is reduced to finding a T-periodic function $x_*(t)$, i.e., a function defined for $t \in [0; T]$, whereas the functional $I(x, \gamma)$ is defined for $t \in [0; nT]$. In this case, according to the concept described above, functional (2) to be minimized takes the form

$$I_n(x, \gamma) = \left\{\int_0^{nT} [y(t) - x(t)]^2 + \gamma[x_0(t) - x(t)]^2\right\}dt, \tag{7}$$

where $x(t)$ is a T-periodic function, $y(t)$ is an aperiodic function, and $x_0(t)$ is given in the same form as the T-periodic function (in a particular case, it is constant).

When the desired function $x_*(t)$ is T-periodic, the problem of minimizing functional $I_n(x, \gamma)$ (7) cannot be solved immediately in the analytical form on the basis of the method developed in [3] using the above-described procedure for the case of the single-pulse emission. The problem can also be solved on the basis of other methods neither well known in the modern calculus of variations nor in the optimum-control theory.

In order to solve this problem, we should preliminary apply procedures and transformations similar to those presented in [4]. They relate to the construction of periodic correcting input signals in the case of antennas operating in the packet regime of the signal emission.

With this aim, we initially find a sequence of Green's functions describing the response of a medium to periodically repeated emission signals and then use the transformations given in [5, 6].

By analogy with transformations (1), the response of a medium at the time of the action of the first signal is determined by the formulas

$$\begin{aligned} y(t) &= \int_0^t h(t - \tau)x(\tau)d\tau, \quad t \in (0; T_1], \\ y(t) &= \int_0^{T_1} h(t - \tau)x(\tau)d\tau, \quad t \in (T_1; T]. \end{aligned} \tag{8}$$

In accordance with [4], the response of a medium is determined by summing the responses to each of the sequentially repeated signals. As a result of transforming integration variables, we can write out the function $y(t)$ in the form

$$y(t) = \int_0^{T_1} \varphi_l(t, \tau, T_1, T)x(\tau)d\tau, \tag{9}$$

where the sequence of the Green's functions $\varphi_l(t, \tau, T_1, T)$ is realized in the following manner:

$$\varphi_0(t, \tau, T_1, T) = \begin{cases} h(t-\tau), & \begin{cases}\tau \in (0; t), & t \in (0; T_1],\\ \tau \in (0; T_1], & t \in (T_1; T],\end{cases}\\ 0, \quad \tau \in (t; T_1], & t \in (0; T], \end{cases}$$

$$\varphi_l(t, \tau, T_1, T) = \begin{cases} g_l(t-\tau), & \begin{cases}\tau \in (0; t-lT], & t \in (lT; lT+T_1],\\ \tau \in (0; T_1], & t \in (lT+T_1; (l+1)T],\end{cases}\\ g_{l-1}(t-\tau), & \tau \in (t-lT; T_1], \quad t \in (lT; lT+T_1], \end{cases} \tag{10}$$

$$l = 1, 2, \ldots, n-2,$$

$$\varphi_{n-1}(t, \tau, T_1, T) = g_{n-1}(t-\tau), \quad \tau \in (0; T_1], \quad t \in ((n-1)T; \infty),$$

$$g_l(t-\tau) = \sum_{i=0}^{l} h(t-\tau-iT).$$

As was indicated above, we had assumed that the signal $x(t)$ emitted by an antenna on passing through a medium is subjected to a linear time-independent dynamic transformation. In the sufficiently general case of such a transformation, this results in the fact that the signal $y(t)$ at the output of a medium and the input signal are connected with each other by the linear differential equation

$$d(p)y = b(p)x, \tag{11}$$

where $d(p)$ and $b(p)$ are constant-coefficient polynomials in terms of the differentiation operator. The polynomials have the degrees m and s, respectively. We assume in this case that, first, $m > s$, and, second, the equation $d(p) = 0$ has m simple roots p_k, $k = 1, 2, \ldots, m$, the real-valued parts of which are negative. By virtue of this fact, the expressions for the functions $h(t-\tau)$, $g_i(t-\tau)$ entering into formulas (1) and (10) can be rewritten as

$$h(t-\tau) = \sum_{k=1}^{m} \frac{b(p_k)}{d'(p_k)} e^{p_k(t-\tau)},$$

$$g_l(t-\tau) = \sum_{k=1}^{m}\sum_{i=0}^{l} \frac{b(p_k)}{d'(p_k)} e^{p_k(t-\tau-iT)} \tag{12}$$

$$= \sum_{k=1}^{m} \frac{b(p_k)}{d'(p_k)} e^{p_k(t-\tau)} (1-e^{p_kT})^{-1} (e^{-p_klT} - e^{p_kT}).$$

The function $y(t)$ (9) determines the response of a medium to a sequence of periodically repeated signals, which is expressed in terms of the pulse shape $x(t)$ of a single signal for $t \in (0, T_1)$.

Furthermore, similarly to (4), the functional $I_n(x, \gamma)$ (7) transforms into the form

$$I_n(x, \gamma) = \sum_{l=0}^{n-1} \int_{lT}^{(l+1)T} \{[y_l(t) - x(t)]^2 + \gamma[x_0(t) - x(t)]^2\} dt. \tag{13}$$

Here, $y_l(t)$ denotes values of the function $y(t)$, which it takes within the semisegments $t \in [lT; (l+1)T]$, $l = 0, 1, \ldots, (n-1)$. As a result, the functional $I_n(x, \gamma)$ (13) takes the form

$$I_n(x, \gamma) = \sum_{l=0}^{n-1} \int_0^T \{[y_l(t_l) - x(t_l)]^2 + \gamma[x_0(t_l) - x(t_l)]^2\} dt_l. \tag{14}$$

As far as in each of the integration terms, the integration variables t_l vary within the same limits $[0; T]$; furthermore, instead of t_l, we introduce the integration variables θ identical for each term. As a result, relation (14) transforms into the form

$$I_n(x, \gamma) = \sum_{l=0}^{n-1} \int_0^T \{[y_l(\theta) - x(\theta)]^2 + \gamma[x_0(\theta) - x(\theta)]^2\} dt. \tag{15}$$

It should be noted that the function $y_l(\theta)$ in functional (15) is determined by replacing the argument $t - lT$ by θ in the Green's functions $\varphi_l(t, \tau, T_1, T)$ entering into formulas (10).

As a result, the functional $y_l(\theta)$ can be written as

$$y_l(\theta) = \int_0^{T_1} \varphi_l(\theta, \tau, T_1, T) x(\theta) d\theta, \tag{16}$$

where, upon the above-indicated replacement, the sequence of the Green's functions $\varphi_l(\theta, \tau, T_1, T)$ with allowance made for (10) can be determined in the form

$$\varphi_0(\theta, \tau, T_1, T) = \begin{cases} h(\theta - \tau), & \begin{cases} \tau \in (0; \theta], & \theta \in (0; T_1], \\ \tau \in (0; T_1], & \theta \in (T_1; T], \end{cases} \\ 0, \quad \tau \in (\theta; T_1], & \theta \in (0; T_1], \end{cases}$$

$$\rho_l(\theta, \tau, T_1, T) = \begin{cases} g_l(\theta + lT - \tau), & \begin{cases} \tau \in (0; \theta], & \theta \in (0; T_1], \\ \tau \in (0; T_1], & \theta \in (T_1; T], \end{cases} \\ g_{l-1}(\theta + lT - \tau), & \tau \in (\theta; T_1], \quad \theta \in (0; T_1], \end{cases} \tag{17}$$

$$l = 1, 2, \ldots, n-1.$$

With due regard to expression for $g_l(t - \tau)$ (12), the function $g_l(\theta + lT - \tau)$ entering into relationships (17) is represented as

$$g_l(\theta + lT - \tau) = \sum_{k=1}^{m} \frac{b(p_k)}{d'(p_k)}(1 - e^{p_k T})^{-1} \times [e^{p_k(\theta - \tau)} - e^{p_k(\theta - \tau + (l+1)T)}]. \tag{18}$$

To minimize the functional $I_n(x, \gamma)$ (13), we apply a modified procedure based on that described in [3]. In this case, expression (16) is written in the abstract form

$$y_l = A_l x, \tag{19}$$

where the linear operators A_l and functional (15) with expression (19) taken into account are determined by Green's functions (17) and by the formula

$$I_n(x, \gamma) = \sum_{l=0}^{n-1} (A_l x - x, A_l x - x) + \gamma n (x_0 - x, x_0 - x), \tag{20}$$

respectively. In accordance with [3], the condition of the time-independent behavior of functional (15) is of the form

$$\left(\sum_{l=0}^{n-1} A_l^* A_l - \sum_{l=0}^{n} A_l^* + \gamma n \right) x = \gamma n x_0, \tag{21}$$

where operators conjugate to operators A_l (19) are denoted as A_l^*.

Similarly to the case of a single pulse, Eq. (21) is a linear operator equation with respect to the desired function $x(t, \gamma)$. The solution to this equation is

$$x_*(t, \gamma, n) = R(\gamma, n)\gamma n x_0,$$

where

$$R(\gamma, n) = \left(\sum_{l=0}^{n-1} A_l^* A_l - \sum_{l=0}^{n} A_l^* + \gamma n \right)^{-1}$$

is the resolvent of the Volterra integral equation of the second kind.

A STABLE PERIODIC REGIME

In the case in which the number of pulses n is sufficiently large, it seems reasonable to consider a periodic stable-emission regime. In this case, we assume that, like the signal $x(t)$ emitted by an antenna, the signal $y(t)$ at the output of a medium is also a T-periodic function of time.

It is worth noting that the description of stable system response (11) to an arbitrary periodic input signal $x(t)$ is possible on the basis of the formalism of pulse–frequency characteristics, which was developed in [7]. However, a specific feature of the problem is that $x(t) = 0$ as $t \in (T_1, T]$. This makes it possible to employ a simpler method to find the corresponding Green's function, provided that we use expression (12) for $g_l(t - \tau)$ and the sequence of the Green's functions for transient regime (10). It should be remembered that $lT \longrightarrow \infty$ as $t \longrightarrow \infty$. At the same time, the difference $t - lT$ remains finite and corresponds to the current time of a stable periodic regime.

With allowance made for this fact, as was indicated above, $\mathrm{Re} p_k < 0$, and we obtain from (12)

$$\lim_{t_1 \to \infty} g_l(t - \tau) = \lim_{t \to \infty} \sum_{k=1}^{m} \frac{b(p_k)}{d'(p_k)}(1 - e^{p_k T})^{-1} \times [e^{p_k(t - lT - \tau)} - e^{p_k(t - \tau + T)}] = \sum_{k=1}^{m} \frac{b(p_k)}{d'(p_k)}(1 - e^{p_k T})^{-1} e^{p_k(\theta - \tau)} = g(\theta - \tau). \tag{22}$$

In a similar manner, we can prove that

$$\lim_{t\to\infty} g_{l-1}(t-\tau) = \sum_{k=1}^{m} \frac{b(p_k)}{d'(p_k)}(1-e^{p_k T})^{-1} e^{p_k(\theta-\tau+T)} = g(\theta-\tau+T). \quad (23)$$

Based on formulas (22) and (23), we can find from (10) the stable response at the output of a medium as $t \longrightarrow \infty$, $n \longrightarrow \infty$:

$$y(\theta) = \int_0^{T_1} \varphi(\theta-\tau, T_1, T)x(\tau)d\tau, \quad (24)$$

where

$$\varphi(\theta-\tau, T_1, T) = \begin{cases} g(\theta-\tau), & \theta \in (0; T_1], \\ g(\theta-\tau+T), & \theta \in (T_1, T]. \end{cases}$$

In the case of a stable periodic regime, the functional to be minimized is

$$I(x,\gamma) = \int_0^{T} [y(\theta)-x(\theta)]^2 + \gamma[x_0(\theta)-x(\theta)]d\theta. \quad (25)$$

In the abstract form, operator y (24) and functional (25) have a form similar to (3) and (4), respectively. The minimizing function $x_*(\theta, \gamma)$ is determined in form (6) in which the operator A and the conjugate operator A^* are determined from (24).

The theorem given above for the case of a single pulse also remains valid for unstable and periodic regimes of a packet emission.

REFERENCES

1. L. D. Bakhrakh and M. Ya. Izrailovich, Dokl. Akad. Nauk **379**, 325 (2001) [Dokl. Phys. **46**, 491 (2001)].
2. L. D. Bakhrakh and M. Ya. Izrailovich, Dokl. Akad. Nauk **380**, 338 (2001) [Dokl. Phys. **46**, 647 (2001)].
3. R. Bellman, I. Gliksberg, and O. Gtoss, *Certain Problems of Mathematical Theory of Control Processes* (Izd. Inostr. Lit., Moscow, 1962).
4. L. D. Bakhrakh and M. Ya. Izrailovich, Dokl. Akad. Nauk **395**, 330 (2004) [Dokl. Phys. **49**, 158 (2004)].
5. M. Ya. Izrailovich, Mashinoved., No. 1, 34 (1976).
6. M. Ya. Izrailovich, Problemy Mashinostr. Nadezhn. Mashin, No. 6, 10 (1995).
7. E. N. Rozenvasser, *Nonlinear-System Oscillations. Method of Integral Equations* (Nauka, Moscow, 1969).

Translated by G. Merzon

Doklady Physics, Vol. 49, No. 10, 2004, pp. 564–569. Translated from Doklady Akademii Nauk, Vol. 398, No. 4, 2004, pp. 477–481.

TECHNICAL PHYSICS

Orientation Dependence of γ–ε–α' Martensitic Transformations in Single Crystals of Austenitic Stainless Steel with a Low Stacking-Fault Energy

I. V. Kireeva*, Yu. I. Chumlyakov*, V. A. Kirillov*, and Corresponding Member of the RAS S. P. Efimenko**

Received May 27, 2004

Austenitic stainless steels are important structural materials that are widely applied in practice [1–3]. Depending on the concentration of Cr, Ni, Mn, and N, as well as on test temperatures, these steels are deformed by slipping and mechanical twinning and undergo martensitic transformations from a high-temperature γ phase to a face-centered close-packed ε phase and body-centered tetragonal α' martensite [2, 3]. Deformation mechanisms in metastable Fe–Cr(18%)–Ni(8–14%) steels have been experimentally studied for polycrystals [2, 4, 5]. It was shown that types of the arising dislocation structure are planar and cellular and that the deformation mechanisms are translational slipping and mechanical twinning. The ordering of γ–ε–α' and γ–α' martensitic transformations turns out to be dependent on both the crystallographic orientation of grains with respect to an applied load and the character of deformation (by tension or compression) [2, 4, 6].

In order to develop the theory of γ–ε–α' martensitic transformations and to elucidate the physical mechanisms underlying formation of ε and α' phases under plastic deformation, it is necessary to systematically investigate single crystals of these steels, which makes it possible to directly examine an effect of the crystal orientation on the transformations.

In studies of Fe–Cr(17%)–Ni(12%)–Mn(2%)–Si(0.75%) (wt %) steels containing single crystals, we investigated an effect of crystal orientation on the γ–ε–α' martensitic transformation.

Single crystals of austenitic stainless steel chosen for the experiments are characterized by low stacking-fault energies γ_0 equal to 0.025 and 0.015 J m^{-2} at T = 300 K and 77 K, respectively [6]. The low values of γ_0, in combination with a high level of deforming stresses caused by deformation at T = 77 K, promote loss of stability of the perfect $\frac{a}{2}\langle 110\rangle$ dislocation with respect to its splitting into partial $\frac{a}{6}\langle 211\rangle$ Shockley dislocations [3, 6]. The ultimate case of such a splitting results in the appearance of an intrinsic stacking fault in the slip plane. This stacking fault can be considered as a nucleus of the ε phase [5, 7]. In the process of motion of the partial $\frac{a}{6}\langle 211\rangle$ dislocations through a single close-packed (111) plane, ε martensite forms, whereas a mechanical twin in the face-centered cubic lattice is produced by shifting the $\frac{a}{6}\langle 211\rangle$ dislocations in each (111) plane [7].

Analysis of the forces that act on a perfect $\frac{a}{2}[-110]$ dislocation split into two $\mathbf{b}_1 = \frac{a}{6}[-211]$ and $\mathbf{b}_2 = \frac{a}{6}[-12-1]$ partial Shockley dislocations makes it possible to calculate splitting d as a function of applied-stresses σ, of their orientation, of their sign (tension/compression), and of stacking-fault energy γ_0:

$$d = \frac{Gb^2}{8\pi\gamma_{\text{eff}}}, \tag{1}$$

$$\gamma_{\text{eff}} = \gamma_0 \pm \frac{m_2 - m_1}{2}\sigma b_1. \tag{2}$$

Here, γ_0 is the stacking-fault energy, which depends on both the alloy composition and the deformation temperature [6, 8]; σ corresponds to axial stresses applied to the crystal; $\mathbf{b}_1$ is the Bürgers vector of the $\frac{a}{6}\langle 211\rangle$ partial Shockley dislocation; m_1 and m_2 are the Schmid factors for the leading twinning Shockley dislocation b_1 (which produces an intrinsic stacking fault in the process of its motion in the slip plane) and for trailing dis-

* *Kuznetsov Siberian Physicotechnical Institute, Tomsk State University, Novosobornaya pl. 1, Tomsk, 634050 Russia*

** *Baĭkov Institute of Metallurgy, Russian Academy of Sciences, Leninskiĭ pr. 49, Moscow, 117911 Russia*

1028-3358/04/4910-0564$26.00 © 2004 MAIK "Nauka/Interperiodica"

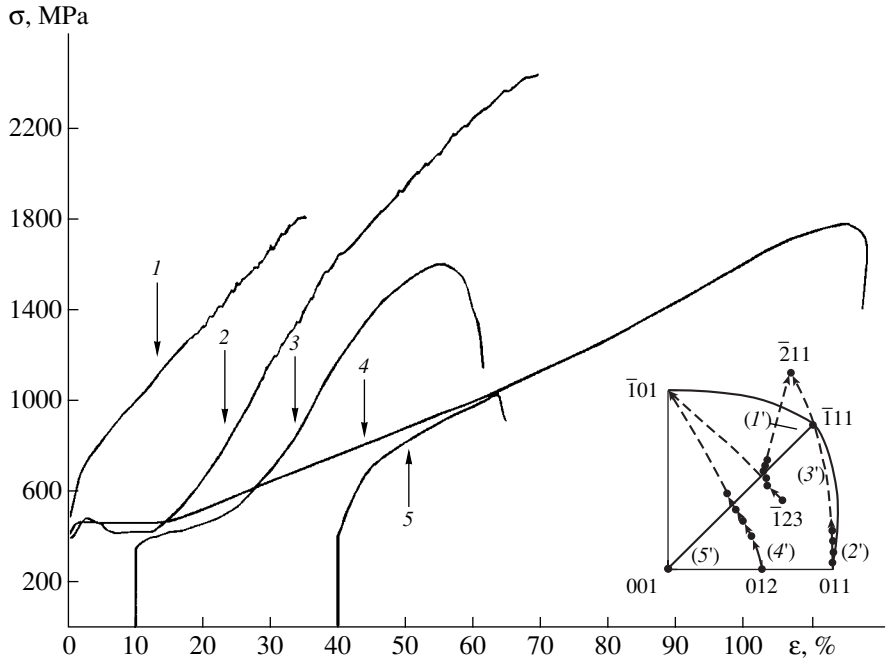


Fig. 1. Curves for flow and precession of crystal axes under deformation by tension of single crystals in Fe–Cr(17%)–Ni(12%)–Mn(2%)–Si(0.75%) steel: (*1*, *1*'), (*2*, *2*'), (*3*, *3*'), (*4*, *4*'), and (*5*, *5*') are the axes of [$\bar{1}$11], [011], [$\bar{1}$23], [012], and [001] crystals, respectively.

location b_2 (which regains a regular atomic packing in the slip plane), respectively; and the "plus" and "minus" signs correspond to deformations by tension or compression, respectively [3, 6, 9]. Equations (1), (2) and the table show that varying the crystal orientation by tension, it is possible to control the value of splitting d and, thereby, the ability to develop the γ–ε martensitic transformation.

The single crystals of Fe–Cr(17%)–Ni(12%)–Mn(2%)–Si(0.75%) austenitic stainless steel were grown by the Bridgman method using seeds placed into alundum crucibles in ambient helium. The method of preparing test samples is described in [6]. To determine the deformation mechanism acting in crystals oriented for a single shift, we investigated a variation in the orientation of the crystal axis under deformation [10]. The phase composition of deformed crystals was examined by both the electron-microscopy and X-ray diffraction methods using techniques described in [11, 12].

In single crystals of steel taken at $T \geq 77$ K, a martensite phase formed by cooling is absent. At T = 300 K, plastic deformation occurs because of slipping up to fracturing of the crystals. The γ–ε–α' martensitic transformation occurs only after deformation by slipping at T = 77–177 K. In this study, we present the results of investigations at T = 77 K.

It is found that the values of plastic deformation ε_{sl}, preceding the γ–ε–α' martensitic transformation, and critical cleavage stresses $\tau_{cr}^{\gamma-\varepsilon}$ and $\tau_{cr}^{\varepsilon-\alpha'}$ (for the γ–ε and ε–α' martensitic transformations, respectively) depend on crystal orientations (see the table and Fig. 1).

In [011] crystals, the γ–ε martensitic transformation occurs after a small deformation by slipping, ε_{sl} = 3% (Fig. 1, curve *2*). The plastic flow in the initial and conjugate systems is realized at ε < 3%, whereas, at ε > 3%, a γ–ε martensitic transformation develops via the Lüders band in only the [$\bar{2}$11](111) system. This conclusion is based on investigation of the crystal-axis precession in the γ–ε martensitic transformation (Fig. 1, curve *2*'). Using the electron-microscopy method, we

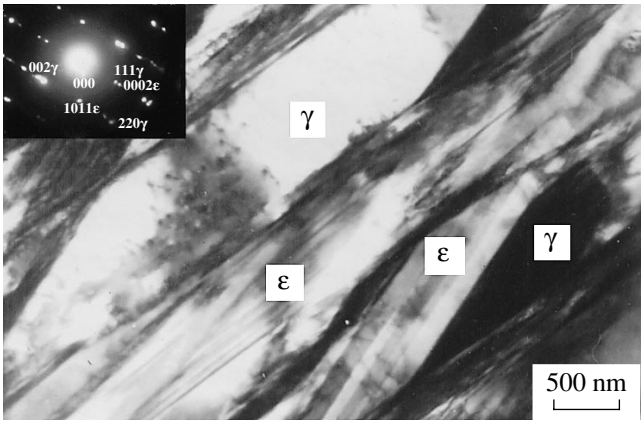


Fig. 2. Dislocation structure of [011] single crystals in Fe–Cr(17%)–Ni(12%)–Mn(2%)–Si(0.75%) steel under deformation by tension at 77 K (ε = 24%). The foil plane is 110_γ.

Basic characteristics of γ–ε–α' martensitic transformations in single crystals of Fe–Cr(17%)–Ni(12%)–Mn(2%)–Si(0.75%) steel under deformation by tension at 77 K

Parameters	Crystal orientations under tension				
	[001]	[011]	$[\bar{1}\,11]$	$[\bar{1}\,23]$	[012]
m_{sl}	0.41	0.41	0.27	0.45	0.5
m_1	0.236	0.471	0.314	0.45	0.4
m_2	0.47	0.24	0.16	0.35	0.4
$Q = \frac{(m_2 - m_1)}{2}$	0.12	−0.12	−0.08	−0.05	0
ε_{sl} before the onset of the γ–ε martensitic transformation, %	15	3	5	18	70
σ_{cr} for the onset of the γ–ε martensitic transformation, MPa	900	480	770	640	1120
τ_{cr} for the γ–ε martensitic transformation, MPa	212	230	240	290	450
γ_{eff}, J m^{-2}	0.021	0.0078	0.0097	0.012	0.015
ε before the onset of the ε–α' martensitic transformation, %	20	28	10	18	80
σ_{cr} for the onset of the ε–α' martensitic transformation, MPa	1000	1080	980	640	1280
τ_{cr} for the ε–α' martensitic transformation, MPa	300	460	280	280	570
$\frac{U}{\sigma}$ [4]	0.0597	0.0237	0.05	0.0665	0.0609
$\frac{d\sigma}{d\varepsilon}\frac{1}{G} \times 10^{-3}$	20	60	40	60	14
$\frac{d\tau}{d\gamma}\frac{1}{G} \times 10^{-3}$		10.2		12	3.4

Note: m_{sl} is the Schmid factor for slipping in the initial slipping $[\bar{1}\,01](111)$ system, m_1 is the Schmid factor of the leading $[\bar{2}11]$ Shockley dislocation, m_2 is the Schmid factor of the trailing $[\bar{1}\,\bar{1}\,2]$ Shockley dislocation in the $[\bar{2}11](111)$ system, and Q is the orientation factor [9]. The quantities τ_{cr} of the γ–α' martensitic transformations were calculated for the $[\bar{1}\,01](\bar{1}\,21)$ system [4].

observed plates of the ε phase in the same $[\bar{2}11](111)$ system in combination with a high density of intrinsic stacking faults in the initial (111) slipping plane. The stage characterized by a linear hardening at ε > 12% is associated with the simultaneous development of the γ–ε martensitic transformations in both systems (Fig. 2). The precession of the crystal axis occurs toward the $[\bar{2}11]$ pole, Therefore, the linear-hardening stage at 12% < ε < 24% is associated with developing the ε phase, preferably in the initial slipping system. At ε > 24%, α' martensite is detected by both the electron-microscopy and X-ray diffraction methods.

In $[\bar{1}11]$ crystals, the γ–ε–α' martensitic transformation is observed for ε > 5% (see Fig. 1, curve *1*; Fig. 3; and the table). At the instant of time when the ε martensite appears, the dislocation structure is characterized by a high density of intrinsic stacking faults and the $\frac{a}{2}\langle 110\rangle$ perfect dislocations split into $\frac{a}{6}\langle 211\rangle$ partial Shockley dislocations. The ε martensite develops in several systems simultaneously, and α' martensite is observed in areas in which ε-martensite plates intersect each other.

Thus, in [011] and $[\bar{1}11]$ crystals, conditions for the γ–ε martensitic transformation are attained at minor plastic deformations by slipping (ε < 3–5%); the values of τ_{cr} for these transformations are equal to 230–240 MPa. γ–α' martensitic transformations in $[\bar{1}11]$

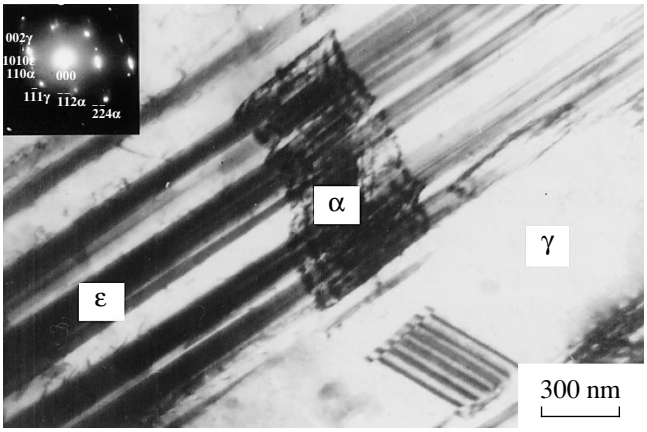


Fig. 3. Dislocation structure of [–111] single crystals in Fe–Cr(17%)–Ni(12%)–Mn(2%)–Si(0.75%) steel under deformation by tension at $T = 77$ K ($\varepsilon = 10\%$). The face-centered close-packed phase, the body-centered tetragonal phase, and the face-centered cubic phase are ε martensite, α' martensite, and the γ phase, respectively. The foil plane is 110_γ.

crystals are observed for $\varepsilon = 10\%$ and $\tau_{cr} = 280$ MPa and, in [011] crystals, for $\varepsilon = 28\%$ and $\tau_{cr} = 460$ MPa (see the table). A high work-hardening coefficient Θ_{II} at the linear stage of hardening in [011] and $[\bar{1}11]$ crystals is associated with the interaction between ε and α' martensites. The ratio $\frac{\Theta_{II}}{G}$ (G is the shear modulus for steel, which equals 85×10^3 MPa at 77 K) exceeds by a factor of three to five those usually observed for slipping in pure face-centered cubic metals and in their substitution alloys with close values of γ_0 (see the table) [10].

In $[\bar{1}23]$ and [012] crystals, γ–ε martensitic transformations take place after plastic deformations of 18 and 70% by slipping, respectively; i.e., when the crystal axis in the process of plastic flow attains the [001]–$[\bar{1}11]$ symmetral (a symmetral is a perpendicular passing through the midpoint of a segment) upon precession of the crystal axis toward the $[\bar{1}01]$ pole (Fig. 1, curves *3*, *3*', *4*, and *4*'). The critical cleavage stresses τ_{cr} for the γ–ε martensitic transformations in [012] and $[\bar{1}23]$ crystals are equal to 440 and 290 MPa, respectively. We should note that the γ–α' and γ–ε martensitic transformations in both $[\bar{1}23]$ and [012] crystals start virtually simultaneously. This is confirmed by both investigation of the dislocation structure and phase analysis (Fig. 4). The stresses τ_{cr} for γ–ε and γ–α' martensitic transformations in $[\bar{1}23]$ single crystals lie within the range 280–290 MPa. For [012] single crystals, the critical cleavage stresses τ_{cr} for γ–ε and γ–α' martensitic transformations are equal to 450 and 570 MPa, respectively.

In [001] crystals, no ε and α' phases are observed by the X-ray diffraction method (Fig. 4). At $\varepsilon > 15\%$, the electron-microscopy method makes it possible to observe ε martensite, twins, and extrinsic stacking faults. The volume fraction of α' martensite turns out to be close to 2%. The formation of α martensite occurs at $\varepsilon > 20\%$, and its volume fraction does not exceed 1%. Thus, at 77 K, the basic mechanism of deformation in

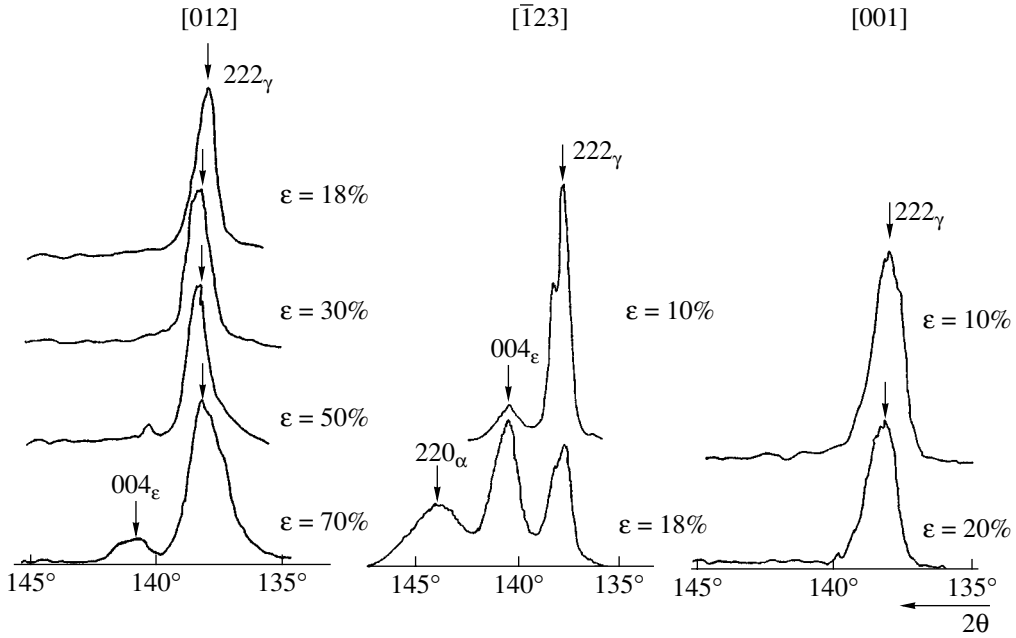


Fig. 4. Phase composition of single crystals in Fe–Cr(17%)–Ni(12%)–Mn(2%)–Si(0.75%) steel as a function of the degree of plastic deformation by tension at $T = 77$ K.

[001] crystals is slipping. The ratio $\frac{\Theta_{II}}{G}$ is proved to be close to that usually observed in pure face-centered cubic crystals oriented for multiple slipping (see the table) [10].

The physical reason for a high orientation dependence of developing the γ–ε martensitic transformation under plastic deformation of single crystals in Fe–Cr(17%)–Ni(12%)–Mn(2%)–Si(0.75%) steel is associated with the effect of the external-stress field σ on the splitting value d according to relations (1) and (2).

Under deformation by tension of [011] and $[\bar{1}11]$ crystals, the Schmid factors for the partial Shockley dislocations b_1 turn out to be larger than those (m_{sl}) for deformation by slipping. The orientation factor is $Q < 0$ and attains the maximum values in these orientations (see the table). In combination, both these factors promote nucleation and growth of the intrinsic stacking faults, i.e., the development of γ–ε martensitic transformations under minor deformations by slipping. Nucleation of ε martensite can be caused by the action of a so-called slipping source when a perfect $\frac{a}{2}\langle 110\rangle$ dislocation is split in its slipping plane as $\gamma_{eff} \longrightarrow 0$ (see relations (1) and (2) and the table), forming a nucleus of the ε phase [13].

In the [012] and $[\bar{1}23]$ orientations, the stresses in the slipping and twinning systems turn out to be the same—$m_1 \approx m_{sl}$—and the modulus of the Q factor decreases compared to that in [011] and $[\bar{1}11]$ crystals (see the table). Therefore, in these orientations, the conditions under which the mechanism of deformation by slipping is replaced by the γ–ε martensitic transformation are attained at the final stages of plastic flow. It should be noted that the change of these mechanisms takes place at the moment the crystal axis attains the [001]–$[\bar{1}11]$ symmetral. There, the values of the Schmid factors for the initial and conjugate slipping systems are equal to each other. Hence, interaction between the two slipping systems can result in the formation of steps on the dislocations, which can act as pole sources of ε martensite [14].

In [001] crystals, where $m_1 < m_{sl}$, the orientation factor Q becomes positive. As a result, γ_{eff} increases, whereas d decreases, and formation of intrinsic stacking faults and of the γ–ε martensitic transformation does not occur.

The orientation dependence of the ε–α' martensitic transformation of single crystals in steel is specified by two reasons. Firstly, α' martensite in austenitic stainless steels with low γ_0 is formed in the lamellas of the ε phase. Therefore, the ε phase is considered to be an intermediate phase for the γ–α' martensitic transformation [2, 4]. In $[\bar{1}23]$ and [012] crystals, α' martensite is formed after strong deformation by slipping due to hampering of the γ–ε martensitic transformation. In [001] crystals, α' martensite is absent because of suppression of the formation of ε martensite. Finally, in $[\bar{1}11]$ and [011] crystals, the γ–ε martensitic transformation occurs in the case of ε ≥ 3–5%. However, in $[\bar{1}11]$ and [011] crystals, α'martensite is observed even for deformations ε ≥ 5%, and ε ≥ 26%, respectively. Secondly, to explain the differences between the ε–α' martensitic transformations in $[\bar{1}11]$ and [011] crystals, it is necessary to take into account a factor describing the orientation dependence of the work U required to form α'-martensite crystals [4]:

$$\frac{U}{\sigma} = \omega_0 \cos\Theta\cos\lambda \pm 0.5(1 + \cos 2\Theta)\varepsilon_0. \qquad (3)$$

Here, $\omega_0 = 0.192$ is the shear component of the shape variation and $\varepsilon_0 = 0.089$ is the shape-variation component normal to the habit plane (dilatation component) of the ⟨110⟩ (112) shift system, both of which lead to the formation of α' martensite; λ is the angle between the shear direction and the crystal axis; and Θ is the angle between the normal to the habit plane and the crystal axis [2, 4].

As is seen from the table, the ratio $\frac{U}{\sigma}$ attains maximum values for the $[\bar{1}11]$, $[\bar{1}23]$, and [012] orientations. These are orientations in which α' martensite nuclei appear in plates of the ε phase immediately after the γ–ε martensitic transformation. In [011] crystals, the ratio $\frac{U}{\sigma}$ has its minimal value. Therefore, the required level of stresses needed for nucleation and growth of α' martensite is attained by cold hardening in the process of the development of the γ–ε martensitic transformation in two intersecting systems and in the process of the crystal-axis reorientation under precession [2, 4, 5].

Thus, in single crystals of austenitic stainless steels, we have found experimentally the orientation dependence for the development of the γ–ε–α' martensitic transformation under plastic deformation. It is shown that the field of external stresses promotes the appearance of a nucleus of the ε phase, which is caused by splitting the perfect $\frac{a}{2}\langle 110\rangle$ dislocations into partial $\frac{a}{6}\langle 211\rangle$ Shockley dislocations. These results can be used for developing models describing γ–ε–α' martensitic transformations in polycrystals.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 02-03-32013); the Basic-Research Support Foundation established by OAO MMK, ITTs Ausferr, and FniO Intels (grant

no. 04-02-02); and by the Ministry of Education of Russian Federation (grant no. A03-3.17-462).

REFERENCES

1. S. P. Efimenko, Izv. Vuzov: Metally, No. 1, 119 (1992).
2. M. A. Filippov, V. S. Litvinov, and Yu. R. Nemirovsliĭ, *Steels with Metastable Austenite* (Metallurgiya, Moscow, 1988).
3. Yu. I. Chumlyakov, I. V. Kireeva, S. P. Efimenko, *et al.*, Dokl. Akad. Nauk **340**, 486 (1995) [Phys. Dokl. **40**, 65 (1995)].
4. D. Goodchild, W. T. Roberts, and D. V. Wilson, Acta Metall. **18**, 1137 (1970).
5. R. Lagneborg, Acta Metall. **12**, 823 (1964).
6. I. V. Kireeva, Yu. I. Chumlyakov, and N. V. Luzginova, Fiz. Met. Metallogr. **94** (5), 92 (2002) [Phys. Met. Metallogr. **94**, 508 (2002)].
7. J. P. Hirth and J. Lothe, *Theory of Dislocations* (McGraw-Hill, New York, 1967; Atomizdat, Moscow, 1972).
8. P. G. J. Gallacher, Metal. Trans. **1**, 2429 (1970).
9. S. M. Copley and B. H. Kear, Acta Metall. **16**, 227 (1968).
10. R. Berner and H. Kronmüller, *Plastic Deformation in Single Crystals* (Springer, Berlin, 1965; Mir, Moscow, 1969).
11. P. B. Hirsh, A. Howie, R. B. Nicholson, *et al.*, *Electron Microscopy of Thin Crystals* (Butterworths, London, 1965; Mir, Moscow, 1968).
12. S. S. Gorelik, Yu. A. Skakov, and L. N. Rastorguev, *X-ray Diffraction and Electron Optical Analyses* (MISIS, Moscow, 1994).
13. J. A. Venables, Phil. Mag. **6**, 379 (1963).
14. A. Seeger, Z. Metallk. **47**, 653 (1956).

Translated by Yu. Vishnyakov

Doklady Physics, Vol. 49, No. 10, 2004, pp. 570–572. Translated from Doklady Akademii Nauk, Vol. 398, No. 6, 2004, pp. 756–758.

TECHNICAL PHYSICS

Relaxation Mechanisms and Different Paths of Defect-Structure Evolution under Severe Plastic Deformations

A. M. Glezer and V. A. Pozdnyakov

Presented by Academician Yu.K. Koveristyĭ March 31, 2004

Received April 29, 2004

The evolution of the structure and properties, as well as the possibility of the formation of nanostructural states in metals and alloys under severe plastic deformations, have been investigated by numerous authors [1–7]. Analysis of the data available in the literature shows the existence of a wide spectrum of structural states formed at different deformation stages and significant discrepancies in both the characteristics of the observed structural states and the description of possible paths of structural evolution with an increasing degree of deformations.

In [1], the conception of an ultimate deformation structure was proposed. According to this conception, upon attaining a certain degree of plastic deformation, mutual misorientation of mesoscopic-scale domains and formation of boundaries of microscopic-scale grains occur. The fundamental role of rotational plasticity modes at the stage of developed plastic deformation was shown and a quantitative description of mesodefects arising at grain boundaries was given. It was indicated that, in this case, fragment sizes cannot be lower than 0.2 μm [1]. Therefore, the formation of a nanostructural state under severe deformation remains outside the framework of this approach. Electron-microscopy studies have proven that the evolution of the metal structure in the plastic-deformation process exhibits a clearly pronounced stage character [8]. At initial deformation stages, formation of a dislocation cellular structure occurs. In the deformation process, further passage to the fragmentation structure, as a rule, occurs via an intermediate state, namely, via formation of a band structure, i.e., a set of almost parallel reorientation bands [4, 7, 8].

On the basis of experimental data obtained, theoretical analysis was performed of different scenarios on the formation of the fragmented structure [1–8]. However, by now, there have been no clear physical conceptions why different materials demonstrate the unequal character of the evolving defect structure and how to explain the transformations of some of them into the nanocrystalline and even amorphous states with increasing degree of deformation.

In the present study, we show the possibility of realizing several variants of cyclic development of the defect structure at the stage of developed plastic deformation at a low and moderate homological temperature, $T/T_{melt} < 0.3$.

At an arbitrary stage of plastic deformation of a material, both processes resulting in the elevation of the accumulated elastic energy and relaxation processes leading to its lowering are simultaneously realized. Macroscopically nonequilibrium systems exist in the state of local equilibrium provided that the rate of variation of their macroscopic state is considerably lower than the rate of each elementary process determining this microscopic state. The assumption that, in the deformation process, a structure appears that corresponds to the energy minimum represents the facts if characteristic times of accommodation processes proceed much faster than the formation of this structure. In order to realize these accommodation structural processes, dislocations must possess sufficiently high mobility.

In analyzing deformation structures, an approach based on the concepts on low-energy dislocation structures (LEDSs) [9] has been successfully applied. One of the authors of the present study (V.A. Pozdnyakov) has extended this approach in [4] to dislocation-disclination structures. In this case, the formation of band structures and fragmented structures is considered as a relaxation process of the division of the system into so-called plastic domains. This process lowers the elastic deformation energy as much as the division of ferroplastics or martensite crystals into elastic domains [10].

With increasing misorientation angle θ and decreasing band thickness *d* in the process of plastic deforma-

Bardin Central Research Institute of Ferrous Metallurgy, Vtoraya Baumanskaya ul. 9/23, Moscow, 105005 Russia
e-mail: glezer@imph.msk.ru

1028-3358/04/4910-0570$26.00

tion, the band structure becomes less stable. However, under validity of the condition [4]

$$d^* < d^*_{\rm cr} = \frac{2\gamma}{e_m}, \qquad (1)$$

where γ is the specific energy of band boundaries, and e_m is elastic deformation, the fragmented structure becomes more profitable from the energy standpoint: the passage to misoriented approximately equilibrium domains takes place, i.e., the fragmentation occurs [1].

Thus, in the process of evolution of the defect structure in the course of plastic deformation, new basis structural states appear that are characterized by the minimum free energy. The lower the dislocation mobility is, the lower the ability to relax arising stresses, the higher the degree of nonequilibrium of the defect structure, and the stronger the differences between the actual and basis quasi-equilibrium structures.

Different variants of the further evolution of fragmented structure are possible. In the case of a sufficiently high dislocation mobility and of realization of various mechanisms of dynamic recovery, the average sizes and misorientations of fragments of the regular structure being formed attain a saturation level. If the dislocation mobility is low, the process of the deformation-structure development can be unstable. In [3], as a result of studies of the copper substructure after severe deformation by torsion (up to 4000%), processes of dynamic recrystallization are observed. These processes occur at room temperature and exhibit a collective character of reorientation (by means of a certain structural explosion). Under further deformation, the fragmentation arises again but, in this case, the size of the fragments being formed is considerably smaller than that of the initial ones. In the course of subsequent deformation, reduction of internal elastic stresses occurs due to the development of recovery processes. The nucleation of fresh grains is possible at low temperatures in local strongly cold-hardened domains.

In [11], an effect of periodic variation of the defect structure and microhardness of a material being deformed with $\varepsilon = 0.3$–0.6 and with the characteristic deformation period $\delta\varepsilon$ has been discovered and investigated (nonthermal softening phenomenon). In [12], a model of periodic variation of the dislocation structure and deformation behavior has been proposed. The model is based on the concepts of rotational instability of plastic deformation. Nonmonotonic deformation behavior under severe plastic deformation is, apparently, of a more general nature. This behavior can be associated with variation of the type of the regular defect structure, e.g., with the realization of local low-temperature recrystallization processes. It can also be associated with the activation of a new structure-scale level of plastic deformation, namely, fundamental relaxation phenomena in the process of severe plastic deformation. As a result of the relaxation of high local

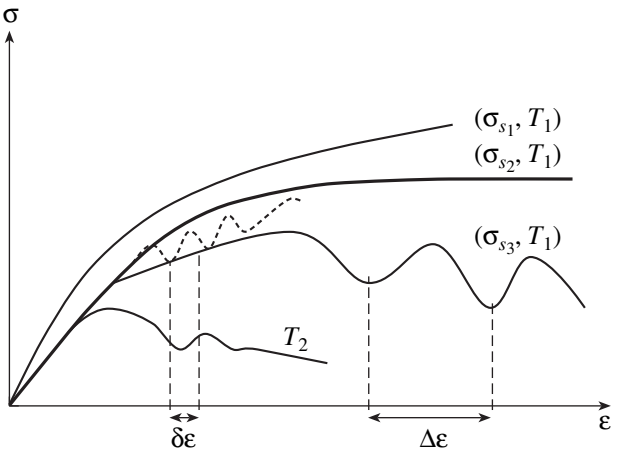


Fig 1. Schematic deformation curves for high degrees of deformation in materials with different degrees of dislocation mobility (with yield stresses $\sigma_{s1} > \sigma_{s2} > \sigma_{s3}$) at temperatures T_1 and T_2, $T_2 \gg T_1 \approx 300$ K.

stresses and their moments, nanocrystals can be formed. If the relaxation rate $\left(\frac{d\sigma}{d\varepsilon}\right)_{\rm rr}$ of internal stresses in a fundamental relaxation process exceeds the deformation-hardening rate $\left(\frac{d\sigma}{d\varepsilon}\right)_{\rm dhr}$, i.e.,

$$\left|\left(\frac{d\sigma}{d\varepsilon}\right)_{\rm rr}\right| \geq \left(\frac{d\sigma}{d\varepsilon}\right)_{\rm dhr}, \qquad (2)$$

then softening of the material and nonmonotonic deformation behavior at super-high levels of plastic deformation with a certain deformation period $\Delta\varepsilon \gg \Delta\varepsilon$ are possible (Fig. 1).

At higher temperatures, continuous recrystallization can proceed [13]. If the free energy of a defect crystal is higher than that of the corresponding amorphous state, then amorphization of the crystal in the deformation process is probable [14].

Various variants of further development of the fragmented structure in the course of continuing deformation are possible, depending on the material's nature, as well as on the temperature, rate, and character of the active deformation under condition of suppressing fracture processes (Fig. 2).

(1) The average sizes and misorientation of fragments attain the saturation level. In this case, the size of fragments free of dislocations does not noticeably change with increasing the degree of deformation.

(2) Cyclic variations of the size and morphology of structural elements take place under the appearance of low-temperature recrystallization. In this case, formation of fresh defect-free grains occurs in the process of fragmented-structure deformation as a result of low-

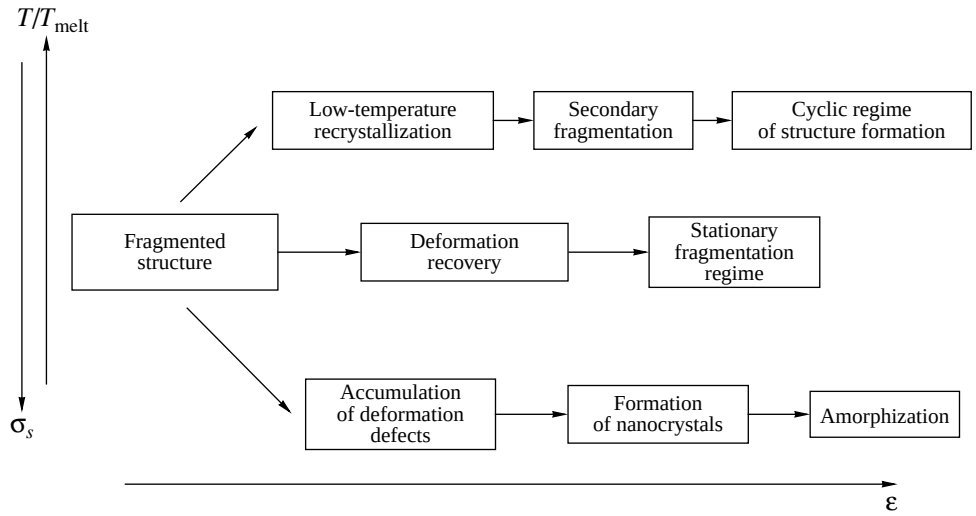


Fig 2. Variants of developing deformation structure under continuing severe deformation of materials with different dislocation mobilities.

temperature recrystallization. Furthermore, the second cycle of sequential variation of structural states is realized up to the next stage of low-temperature recrystallization.

(3) A continuous decrease in the fragment sizes and formation of nanostructures with a possible further transition into the amorphous structural state occur.

As a rule, these processes proceed nonuniformly over the volume of the deformed material. It is evident that with decreasing the dislocation mobility and/or with lowering the deformation temperature, the probability of realizing a certain sequence of events increases in the direction from the first variant toward the third one (Fig. 2) Thus, the variety of structural states observed in various experiments on severe plastic deformations is stipulated, first, by different above-considered variants of defect-structure evolution and, second, by different stages of realizing the structural state in each of these variants, which have managed to proceed under the given deformation conditions. Thus, the wide spectrum of observed structures, including fragmented, nanocrystalline, and amorphous ones, satisfies the conception developed in this study.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 03-02-17296.

REFERENCES

1. V. V. Rybin, *Severe Plastic Deformations and Fracture of Metals* (Metallurgiya, Moscow, 1986).
2. N. Hansen and D. J. Jensen, Philos. Trans. R. Soc. London **357A**, 1447 (1999).
3. V. M. Bykov, V. A. Likhachev, Yu. A. Nikonov, *et al.*, Fiz. Met. Metalloved. **45** (1), 163 (1978).
4. V. I. Izotov, V. V. Rusanenko, V. I. Kopylov, *et al.*, Fiz. Met. Metalloved. **81** (9), 123 (1986).
5. A. N. Tyumentsev, M. V. Tret'yak, Yu. P. Pinzhin, *et al.*, Fiz. Met. Metalloved. **90** (5), 44 (2000).
6. V. I. Vladimirov and A. E. Romanov, *Disclinations in Crystals* (Nauka, Leningrad, 1986).
7. É. V. Kozlov, N. A. Koneva, and L. I. Trishkina, *Disclinations and Rotational Deformation of Solids* (Izd. FTI, Leningrad, 1990).
8. *Structural Levels of Plastic Deformation and Fracture* (Nauka, Novosibirsk, 1990), p. 255.
9. D. Kuhlmann-Wilsdorf, Acta Mater. **47**, 1697 (1999).
10. A. L. Roĭtburd, Usp. Fiz. Nauk **113** (1), 69 (1974).
11. Ya. D. Vishnyakov and S. A. Vladimirov, Dokl. Akad. Nauk SSSR **206**, 584 (1972).
12. B. K. Barakhtin, V. I. Vladimirov, S. A. Ivanov, *et al.*, Fiz. Tverd. Tela **28**, 2250 (1986).
13. O. Sh. Sitdikov, R. O. Kaĭbyshev, I. M. Safarov, *et al.*, Fiz. Met. Metalloved. **92** (3), 65 (2001).
14. E. V. Tat'yanin, V. G. Kurdyumov, and V. B. Fedorov, Fiz. Met. Metalloved. **62** (1), 133 (1986).

Translated by G. Merzon

Doklady Physics, Vol. 49, No. 10, 2004, pp. 573–577. Translated from Doklady Akademii Nauk, Vol. 398, No. 4, 2004, pp. 482–486.

ASTRONOMY, ASTROPHYSICS, COSMOLOGY

Natural Phenomena and the Substance in the Ablation Trail of the Vitim Meteoroid (September 25, 2002)

V. S. Antipin*, S. A. Yazev, Academician M. I. Kuz'min*, A. B. Perepelov*, S. V. Efremov*, M. A. Mitichkin*, and A. V. Ivanov*****

Received May 25, 2004

On the night of September 25, 2002, the inhabitants of the Mama–Chuĭsk and Bodaĭbo districts of Irkutsk oblast observed the flight of a bright fireball accompanied by light and sound effects along with an intense airwave that extended to great distances. Weather conditions were unfavorable for observation of the meteoroid. Nevertheless, many eyewitnesses of this event noted the gradual appearance of a uniform white bright luminescence to the southwest that occupied the entire sky and, then, shifted from the Vitim River valley in a northeastern direction. According to the testimony of some eyewitnesses, the color of this luminescence changed from white to blue and reddish-claret. Local inhabitants who observed the flight stated that it had been accompanied by a dummy boom and finished with an impact and shaking of the earth.

The Vitim meteoroid caused a great deal of acoustic effects. The eyewitnesses noted that acoustic phenomena preceded and accompanied the luminescence (a rustle, crackling, and boom as if from an aircraft). These effects may be associated with the production of an electromagnetic wave in the atmosphere.

The appearance of an alternating electric field explains the fact that at this time, incandescent lamps in houses in the settlement of Mama became slightly luminescent, although, at that time, there was no electric power in the area. These phenomena allow classification of the Vitim meteoroid as an electrophone event.

There was almost no objective recording of seismic phenomena associated with the fall of the Vitim meteoroid. According to data of the Institute of Geosphere Dynamics, Siberian Division, Russian Academy of Sciences, obtained by O.N. Popova, the records of the Chara, Nelyaty, and Peleduĭ seismic stations, located in the northern part of Irkutsk oblast, exhibited a weak signal that corresponded to a local event occurring at that time. The flight of the meteoroid was detected by the surveillance system of a US satellite. Bright luminescence was recorded by the satellite at an altitude of 62 km at a point with coordinates of 57.21° N and 112.90° E and was traced further to a point with an altitude of 30 km and coordinates of 58.21° N and 113.46° E. Based on these data, the trajectory of the meteoroid fall was reconstructed. This allowed its inclination angle (approximately 34°) with respect to the horizon to be estimated.

Three expeditions with the participation of scientists from the Institute of Solar and Terrestrial Physics (Siberian Division, RAS), Institute of Geochemistry (IGC, Siberian Division, RAS), and Institute of the Earth's Crust (IEC, Siberian Division, RAS) conducted work in the region of the probable meteoroid fall over 2002–2003. In one of these expeditions, scientists from the Meteoritic Committee of the RAS also took part. At the same time, groups of scientists and students from Yekaterinburg and Krasnoyarsk participated in the search for the meteoritic substance. None of the groups managed to find craters or meteoroid fragments. However, multiple instances of damage to forests were observed near the meteoroid flight route (Fig. 1). Numerous broken-off tops of foliate and coniferous trees were observed, as well as broken branches and fallen trunks 30–40 cm in base diameter, sometimes along with roots. According to the testimony of local inhabitants, that date of this forest damage, which had been based on a number of apparent attributes, corresponded to the autumn of 2002. Data of the Weather Forecast Service indicate the absence of hurricane winds in this region during the time period under consideration.

In spite of certain confusion, the fact of mechanical action on the ambient nature of the shockwave caused by the meteoroid flying through the lower atmosphere cannot be doubted. However, the amplitude of the wave incident from an altitude of 20–30 km must have been

* *Vinogradov Institute of Geochemistry, Siberian Division, Russian Academy of Sciences, ul. Favorskogo 1A, Irkutsk, 664033 Russia*

** *Astronomical Observatory, Irkutsk State University, ul. Sovetskaya 119A, Irkutsk, 664009 Russia*

*** *Institute of the Earth Crust, Siberian Division, Russian Academy of Sciences, ul. Lermontova 128, Irkutsk, 664033 Russia*

e-mail: antipin@igc.irk.ru; uustar@star.isu.ru; dir@igc.irk.ru; region@igc.irk.ru; esv@igc.irk.ru; antipin@igc.irk.ru; aivanov@crust.irk.ru

1028-3358/04/4910-0573$26.00 © 2004 MAIK "Nauka/Interperiodica"

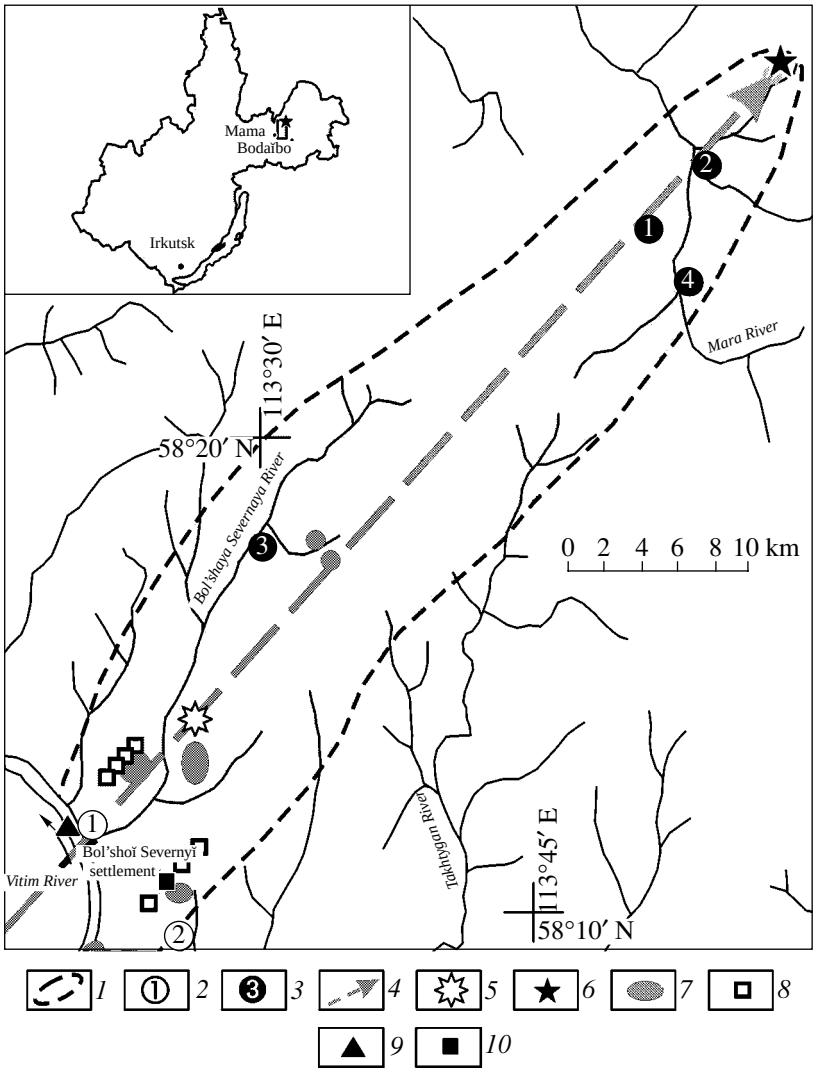


Fig. 1. Map of the region of the expedition studies and snow-blanket analysis near the point of fracture and supposed fall of the Vitim meteoroid. The following conventional denotations are used: (*1*) region of expedition studies of 2002–2003, (*2*) winter stations of the expedition of 2003 [(1) Bol'shoĭ Severnyĭ settlement, (2) mouth of the Berezovyĭ stream], (*3*) summer stations of the expedition of 2003 [(1) upper part of the Mara river left tributary, (2) mouth of the Mustag stream, (3) mouth of the Upornyĭ stream, (4) right bank of the Mara River upper flow], (*4*) projection of the calculated part of the meteoroid-flight trajectory, (*5*) projection of the flash point recorded by a US satellite, (*6*) suggested point of the meteoroid fall, (*7*) region of the observed forest damage and inrush, (*8*) points of sampling of the blanket of snow, (*9*) points of sampling snow on the Vitim River ice, and (*10*) points of finding spherules in samples of the blanket of snow.

insignificant, which contradicts the pattern of numerous fallen and damaged trees.

The basic goal of the second Vitim expedition in April 2003 was the selection of snow samples from bald-peak parts of the mountain ridge along the meteoroid flight trajectory and at the probable fall point of the meteoroid. According to the Weather Forecast Service, 2–5 cm of snow already lay at that time on the mountain peaks. Therefore, particles of a material of cosmic origin might be retained in the lowest snow layer. At the moment of sampling, the blanket-of-snow thickness exceeded 1.5 m. The lowest snow layer (10–15 cm), covering the frozen ground and taken from cleaned areas each of 1–2 m^2, was chosen for samples. In total, 13 snow samples at the tops of cone-shaped hills with altitudes of 1000 m and higher (including a background snow sample taken from ice of the Vitim River) were selected for analysis.

The projection of the Vitim meteoroid route falls within the territory of the Mama–Chuĭsk district, where rocks of carbonate-terrigenous strata of the Mama series are scattered over a recent erosion cut. These rock complexes were metamorphosed in conditions of the amphibole phase. Then, they underwent recrystallization and anatexis, accompanied by the formation of numerous granite cores and mica-bearing pegmatites [1, 2].

Then, diagnostics and investigation of mineral particles extracted as a result of filtration of thawed water from the snow samples were carried out. Among fragments of mineral granules found in rocks of the given region, hollow spherules with a size usually not exceeding 100–200 μm were found on some filters. The fragments of these brittle spherical formations have brown or dark-brown color tints. X-ray diffraction analysis performed by Z.F. Uschapovskaya (IEC) made it possible to identify oxide (spinel, cristobalite) and silicate (enstatite, nepheline) minerals in the material of the spherules. All these mineral phases are encountered in meteorites—in particular, in chondrites [3].

The composition of particles found in the blanket of snow were analyzed by L.A. Pavlova (IGC) and N.S. Karmanov (Institute of Geology), having used, respectively, a Joel Superprobe-733 electronic microanalyzer and an LEO 1430VP electron microscope (manufactured by the LEO Company) that had been equipped with an Inca Energy 300 energy-dispersion analyzer manufactured by Oxford Instruments Company (see Table 1). The data obtained testify to the existence of both sulfide and metallic phases among the microparticles.

A grain of Ni-contained pyrite and particles of ferrous compounds, i.e., oxidized and metallic iron, as well as nickelous iron, were identified. A pyrite particle 3–4 μm in size is associated with a particle of metallic Mn-contained iron among sooty material. A nickelous-iron particle 6 μm × 8 μm in size is associated with a quartz microparticle (Fig. 2a). The particle is inhomogeneous and contains inclusions of metallic iron with admixtures of Ni and nickelous iron. The found sulfide and metallic phases are usual for meteorites; however, they are also found in terrestrial rocks of various origin. At the same time, the ratio of Ni and Fe in the nickelous-iron particle corresponds to taenite and kamacite minerals characteristic of meteoritic substance [4]. The comparison of the data obtained with those for iron and nickel in the metallic particles of some meteorites (Table 2) shows the consistency of the concentrations observed.

Thus, spherules having the characteristic porous structure are present in the material under study, which was obtained from the blanket of snow within the region damaged by the Vitim meteoroid. The fact, in itself, of their presence and similarity to micrometeorites from other regions of the world [5, 6] (Fig. 2d) testifies to the possible meteoritic nature of these formations. The shapes of the spherules and their structural features do not contradict the version of their possible origin in the meteoroid ablation trail. This conclusion is also confirmed by the analysis of mineral associations found in the spherule compositions but not characteristic of basic rocks in the Mama–Chuĭsk district.

Prismatic-shape compositions of Fe–Ti-minerals (up to 4 μm × 7 μm) were observed in the porous-structure spherule (Fig. 2c). The calculation of crystal-chemistry formulas for these minerals with high concentrations of Mn, Fe, and Ti and low concentrations of Si and Al shows the stoichiometric noncorrespondence to oxide phases such as ilmenite or spinel. At the same time, this mineral exhibits the stoichiometry in the case of calculation by the formula X_3O_5 characteristic of minerals of the pseudobrookite–armalcolite group. The material composition of the spherule matrix is calculated on the basis of formulas for clay minerals and hydromicas. This material can be a mixture of illite and montmorillonite components. The matrix of a fragment

Table 1. Representative compositions of mineral particles in the ablation trail of the Vitim meteoroid (blanket of snow)

Component	Fe–Mn	Pyrite	Taenite	Kamacite
	1	2	3	4
Fe	77.85	31.52	28.3–75.0	91.0–96.5
Ni	nf	1.21	21.5–70.6	2.5–8.4
Mn	0.40	nf	nf	nf
S	nf	50.90	nf	nf
Sum	78.25	83.63	100.00	100.00
Component	Illite	Ti–Fe	Amph	Fe–Ti
	5	6	7	8
SiO_2	32.60	3.67	36.83	3.46
TiO_2	1.41	46.31	5.18	20.31
Al_2O_3	26.90	3.10	10.56	2.41
Fe_2O_3	21.89	18.63	17.2	17.15
FeO	8.88	4.54	2.73	44.68
MnO	nf	11.85	nf	0.37
MgO	1.85	0.67	1.59	1.90
CaO	0.54	nf	8.03	0.64
Na_2O	0.55	nf	5.28	0.67
K_2O	2.70	nf	1.80	0.28
P_2O_5	nf	nf	2.45	nf
Sum	97.32	88.77	91.65	91.88
Σ K*	14	3	15	3
Σ A*	20	5	22	4
Balance (±)	–4	0	–2	0

Note: Mineral phases: 1. Mn-containing iron; 2. Ni-containing pyrite; 3. Taenite; 4. Kamacite (the metal concentration range is indicated for compositions normalized to 100%); 5. Spherule matrix with a porous structure (illite?); 6. Ti–Fe mineral inclusions in the illite matrix (stoichiometry of minerals of the pseudobrookite–armalcolite group); 7. Matrix of the amphibole-composition spherule; 8. Inclusions in the Fe–Ti mineral matrix (stoichiometry of spinellides). The ratio Fe_2O_3/FeO is calculated according to stoichiometry. ΣK* is the sum of cations; ΣA* is the sum of anions. Balance (±) implies anion balance. The abbreviation nf indicates that this component has not been found.

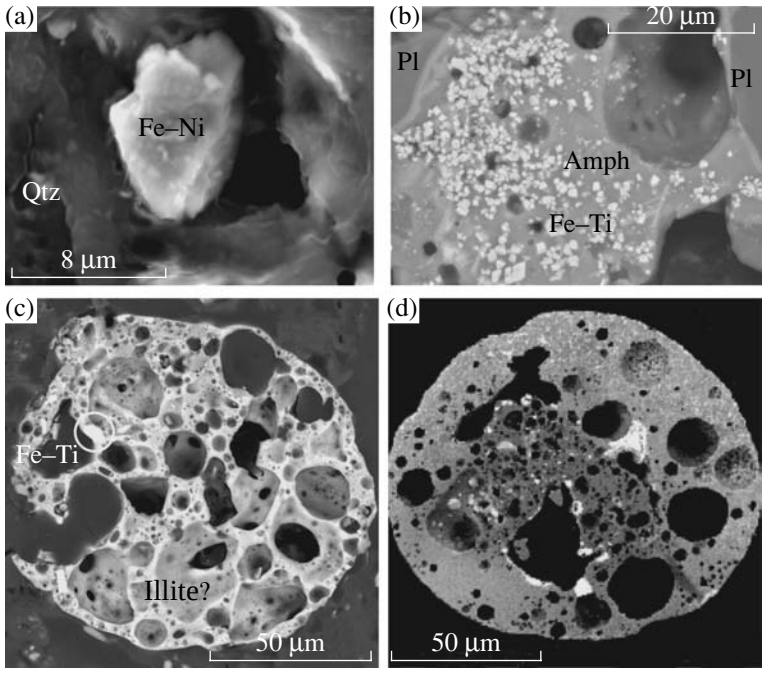


Fig. 2. Particles extracted out of the blanket of snow in the region of the Vitim meteoroid flight route. Particle images are obtained by the method of electron backscattering: (a) particle of nickelous iron (Fe–Ni) and of quartz (Qtz), (b) spherule fragment with the matrix of the amphibole composition (Amph) and inclusions of Fe–Ti minerals (spinellides), (c) spherule with the illite matrix and inclusions of Fe–Ti minerals with the pseudobrookite stoichiometry, and (d) micrometeorite found in the Antarctic blanket of snow [5].

of another spherule with the size exceeding 100 μm × 100 μm (Fig. 2b) is represented by a rather homogenous mineral compound containing very fine crystalline precipitations of Fe–Ti minerals (spinellides) and concretion of needle-shaped crystallites of Ba-contained silicates (up to 15% of Ba). In the edge zones of the spherule fragment under study, which has reaction-character contacts, plagioclase crystals with a significant iron content were found. The observed silicate composition of the second-spherule matrix and the results of the calculation for this composition on the basis of the amphibole crystal-chemistry formula show that the substance can correspond to a mineral mixture of titanous amphibole (kaersutite) and Na–Ca amphiboles. As is well known, alkaline amphiboles—i.e., arfvedsonites [7] and kaersutites [8]—have been found before in meteoric material, as well as in regolith breccias. In some cases, Ba-containing glasses were also found in meteorites [9].

Thus, the diagnostics carried out of particles found in snow samples and studies of their mineral and chem-

Table 2. Composition (wt %) of nickelous iron in samples of some meteorites [4] and in particles found in the blanket of snow within the area of the Vitim meteoroid fracture

Meteorite	Kamacite		Taenite	
	Fe	Ni	Fe	Ni
Kaali	91.6–92.7	6.1–6.5	64.6–71.4	27.5–35.5
Chinge	95.1	4.9	79.3	20.0
Pilistvere	92.2–93.0	6.9–7.8	80.1–89.2	9.6–19.5
Bakhmut	91.7–93.3	5.9–6.0	64.2–72.3	27.6–35.9
Krymka	92.5–94.9	3.9–6.0	48.8–58.4	40.3–49.4
Zhigalovka	92.0–92.5	5.5–6.5	47.7–71.1	28.4–51.1
Aleksandovskiĭ farm	92.2–92.6	7.4–7.5	46.1–67.6	31.7–52.5
Vitim meteoroid*	91.0–96.5	2.5–8.4	28.3–75.0	21.5–70.6

* Compositions of particles found in the blanket of snow are normalized to 100%.

ical composition showed that the particles contain a material that occurs in meteorites and is unlikely to be classifiable as rock from the Mama–Chuĭsk district. In the absence of fragments of the Vitim meteoroid, which have not yet been found, the data reported in this paper are, today, the only weighty evidence of the composition of the meteoroid material.

The problem concerning the rather intense shock-wave that accompanied the meteoroid fall, which was responsible for a number of physical phenomena, including damage to trees, is of no less importance.

It is also worth noting that the accuracy of the satellite determination of the meteoroid fall coordinates should not be overestimated (e.g., the Moravka meteorite was found 20 km to one side of the projection of its fall trajectory as determined from satellite data). It is not improbable that the basic mass of Vitim meteoroid fragments fell out farther along the trajectory or in a sideways direction from the calculated trajectory (in the case of erroneous determination of the altitudes of two points of the fall route or their coordinates, respectively).

ACKNOWLEDGMENTS

The authors are grateful to the heads of the Mama–Chuĭsk district of Irkutsk oblast, as well as to the expedition guides and reliable friends V.G. Andreev, A.P. Petruchenko, and A.I. Skibitskiĭ for their assistance and inestimable help in performing fieldwork in the north of Irkutsk oblast.

These studies were supported by the Siberian Division of the Russian Academy of Sciences.

REFERENCES

1. A. N. Neelov, in *Muscovite Pegmatites of the USSR* (Nauka, Leningrad, 1975), pp. 168–174.
2. V. M. Makagon, *Pegmatite Granites in Zones of Regional High-Pressure Metamorphism* (Nauka, Novosibirsk, 1972).
3. R. T. Dood, *Meteorites. A Petrologic-Geochemical Syntheses* (Cambridge Univ. Press, Cambridge, 1981; Mir, Moscow, 1986).
4. É. V. Sobotovich and V. P. Semenenko, *Meteoritic Substance* (Naukova Dumka, Kiev, 1984).
5. S. Taylor, D. E. Brownlee, R. P. Harvey, *et al.*, US Army Cold Regions Res. Eng. Lab. Rep. **97** (1), 37 (1997).
6. S. Taylor and D. E. Brownlee, Meteoritics **26**, 203 (1991).
7. A. Ivanov and M. Zolensky, Lunar and Planet. Sci. **32** (2001), CD-ROM, Abstr. No. 1386.
8. A. Treiman, *Amphibole in Martian Meteorite EET 79001LX Annual Meteoritic Society Meeting, Hawai, 1997*; www.lpi.usra.edu/meetings/metsopc97/pdf/5134.pdf.
9. A. Taylor, K. C. Misra, S. I. Demidova, *et al.*, Meteoritics and Planet. Sci. **38** (4), 485 (2003).

Translated by G. Merzon

Doklady Physics, Vol. 49, No. 10, 2004, pp. 578–582. Translated from Doklady Akademii Nauk, Vol. 398, No. 6, 2004, pp. 759–763.

ASTRONOMY, ASTROPHYSICS, COSMOLOGY

A Model for Scattering of Small-Size Cosmic-Body Fragments in the Planetary Atmosphere

A. G. Ivanov and V. A. Ryzhanskiĭ

Presented by Academician V.M. Titov February 18, 2004

Received April 6, 2004

In the present paper, we analyze disintegration of a small-size cosmic body (SSCB) in interactions with the planetary atmosphere in the framework of the concept of a two-stage process. The first and second stages of this process are, respectively, the fragmentation of an SSCB and the scattering of the fragments produced. The theoretical model of fragmentation has been presented by the authors previously (see [1]) and attracted the attention of a number of investigators [2, 3]. In the present paper, we analyze the disintegration of an SSCB and consider the stage of scattering of its fragments accompanied by the formation of the primary crater field on a planet's surface.

The model developed in [1] used the integral approach in the energy concept of modern fracture mechanics [4]. This resulted in the description of SSCB fragmentation as a discrete process in which the events of sequential refinement of the SSCB are separated in space and time. In this case, the fracture of both an SSCB and its fragments occurs as the result of separation by a brittle crack into two equal parts geometrically similar to the parent body. It is of importance to note that, in [5], for a spherical SSCB or a rectangular parallelepiped, it was shown that, even in these extreme cases, the body's shape only slightly affects the results of fragmentation calculations.

In accordance with the two-stage representation of the SSCB disintegration, its analysis is divided into two steps. In the first step, the fragmentation is calculated and the results obtained are used as the initial data at the second step. The latter consists in calculating the fragment scattering after each fragmentation event and ceasing the fragmentation prior to the fall of fragments onto the ground.

In order to calculate the fragmentation, the procedure developed in [5] is employed. According to this procedure, an SSCB is considered to be brittle[1] and has the shape of a rectangular parallelepiped with a mass M_0 and a ratio of edges equal to $2^{-1/3}L_0 : L_0 : 2^{1/3}L_0$, where L_0 is the basis size. Penetrating into the atmosphere with cosmic velocity V_0 at angle α to the horizon, an SSCB overcomes the atmosphere drag by virtue of inertia, with the drag maximum corresponding to the center of the frontal surface. The SSCB center of mass (point C_0) moves along the trajectory ξ axis at velocity $V = \frac{d\xi}{dt} = \dot{\xi}$ and acceleration $\ddot{\xi} = -\frac{d^2\xi}{dt^2}$. Center of mass C_0 is associated with the origin of the mobile orthogonal coordinate system ξ'–η–ζ, the ξ and ξ' axes being collinear. Under the action of both aerodynamic drag[2] and inertial forces, a tension domain arises,[3] from which a brittle crack originates. As a result, the SSCB breaks up into two parts. It is considered that the fracture occurs when the frontal surface attains its maximum area ($2^{1/3}L_0^2$), whereas the crack surface has the minimum value ($2^{-1/3}L_0^2$). Such an orientation is more favorable for fracturing and more stable, because the center of mass maximally approaches the frontal (basis) surface undergoing the action of the aerodynamic pressure, which is consistent with the principle of minimum expenditures of energy. Thus, for the indicated ratios of SSCB size, the fragments obtained are geometrically similar to the initial SSCB and break up again, but each time in deeper layers of the atmosphere. Indeed, by virtue of the effect of the scaling factor [4], the fragments become much stronger. The basis size of a fragment formed as a result of the kth fracture act is $L_k = 2^{-k/3} L_0$, its mass is M_k with the center at the point

Russian Federal Nuclear Center,
All-Russia Research Institute of Experimental Physics,
Sarov, Nizhni Novgorod oblast, 607188 Russia
e-mail: root@gdd.vniief.ru

[1] The brittleness is provided not only by the properties of the SSCB, but also by the rather low (cosmic) temperature. It is suggested to be virtually invariable due to the short time of penetration and ablation.

[2] The aerodynamic-drag coefficient for the given SSCB shape is taken as equal to $C_\xi = 1.5$.

[3] As was shown in [6], a similar domain also arises in a spherical SSCB.

1028-3358/04/4910-0578$26.00

C_k, and the fracture-surface area is $2^{-1/3}L_{k-1}^2$ ($1 \le k \le n$, n being the ordering number of the last break-up event). As a result of the calculation, quantity L_k, altitude z_k above the planet's surface, trajectory velocity V_k, and the trajectory-segment length $\xi_k = \frac{z_{k-1} - z_k}{\sin\alpha}$ between the fragmentation acts are determined. After fragmentation has finished, the quantities n, z_n, and V_n; the trajectory segment length $\xi_n = \frac{z_n}{\sin\alpha}$ between the last fragmentation point and the ground, and the velocity of motion in this segment,

$$V(x) = V_n \exp\{-2^{n/3}A[\exp(-x) - \exp(-x_n)]\}, \quad (1)$$

are calculated. Here, $A = \frac{0.509 C_\xi \rho_0 H}{\rho_b L_0 \sin\alpha}$ (with allowance made for the body's rotation), ρ_0 is the atmosphere density at the altitude $z = 0$, ρ_b is the body's density, H is the thickness of the standard atmosphere, and $x = \frac{z}{H}$.

The scattering of the fragments proceeds from the instant ($t = 0$) of the SSCB disintegration. For each fragment, the process occurs according to the same scheme. For example, kth fragments are formed by a $(k-1)$th fragment (Fig. 1). In this case, as in [1, 5], ablation is not taken into account.[4] The process of fragment scattering seems to have two phases. The first phase corresponds to the rotation of fragments about their centers of mass C_k by angle φ at angular velocity $\dot\varphi = \frac{d\varphi}{dt}$ and at angular acceleration $\ddot\varphi = \frac{d^2\varphi}{dt^2}$ under preservation of the mutual contact along the line of edge tangency on the frontal surface. The second phase is the transverse expansion of fragments as independent bodies from the symmetry plane ξ'–ζ after the contact has ceased (the ζ axis is perpendicular to the plane of the plot). In this case, fragments perform complicated motions symmetric with respect to the trajectory C_{k-1} of motion, translational motion by virtue of inertia, and rotation about their centers C_k of mass.

In the first phase, the basic cause of fragment rotation is the aerodynamic-drag force. This force is assumed to be applied at the center-of-mass point C_{k-1} (since the total fragment middle area remains constant) and equals

$$F_{k-1} = 0.5 C_\xi \rho_{ak} \dot\xi^2 S_m \exp\frac{\xi \sin\alpha}{H},$$

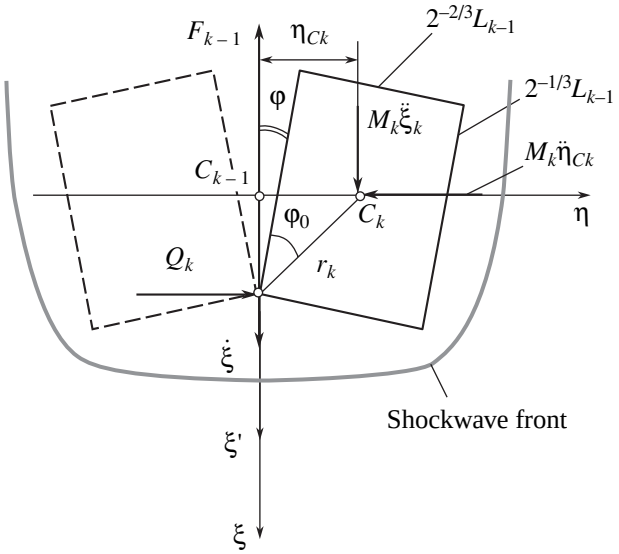


Fig. 1. Diagram of the first scattering phase for a $(k-1)$th fragment of a small-size cosmic body (SSCB).

where $\rho_{ak} = \rho_0 \exp\left(-\frac{z_k}{H}\right)$ is the atmospheric density at the altitude z_k, $S_m = 4 r_k L_{k-1} \sin(\varphi_0 + \varphi)$ is the middle-section area of the fragments, $r_k = 2^{-4/3}\sqrt{1 + 2^{-2/3}}\, L_{k-1}$, and $\varphi_0 \approx 0.671$ rad ($\sim 38°26'$) is the initial inclination angle of the vector r_k to the ξ axis. The inertia force applied to the fragment center of mass (in Fig. 1, it is shown in projections onto the ξ' and η axes) and reaction force Q_k of a neighboring fragment are stipulated by the action of force F_{k-1}. In this case, the point C_k shifts along the η axis from the position C_{k-1} by the distance $\eta_{Ck} = r_k \sin(\varphi_0 + \varphi)$ at the velocity $\dot\eta_{Ck} = r_k \dot\varphi \cos(\varphi_0 + \varphi)$ and at the acceleration $\ddot\eta_{Ck} = r_k[\ddot\varphi \cos(\varphi_0 + \varphi) - \varphi^2 \sin(\varphi_0 + \varphi)]$.

The kinetostatic equations for a kth fragment are of the form

$$M_k \ddot\xi_k = 0.5 F_{k-1}, \quad M_k \ddot\eta_{Ck} = Q_k,$$
$$I_{Ck} \ddot\varphi = 0.5 F_{k-1} \eta_{Ck} - Q_k r_k \cos(\varphi_0 + \varphi).$$

With allowance made for what has been said above, these equations are reduced to the set of equations

$$\ddot\xi_k = a r_k^{-1} \frac{\rho_{ak}}{\rho_b} \sin(\varphi_0 + \varphi) \exp\left(\frac{\xi \sin\alpha}{H}\right) \dot\xi^2, \quad (2)$$

$$Q_k = M_k r_k [\ddot\varphi \cos(\varphi_0 + \varphi) - \dot\varphi^2 \sin(\varphi_0 + \varphi)], \quad (3)$$

$$\left[5\frac{I_{Ck}}{M_k r_k^2} + \cos 2(\varphi_0 + \varphi)\right]\ddot\varphi - \dot\varphi^2 \sin 2(\varphi_0 + \varphi) = 2 a r_k^{-2} \left(\frac{\rho_{ak}}{\rho_b}\right) \exp\left(\frac{\xi}{H}\sin\alpha\right) \dot\xi^2 \sin^2(\varphi_0 + \varphi). \quad (4)$$

[4] It is shown in sections 2 of [7, 8] that the effect of this phenomenon is not defining.

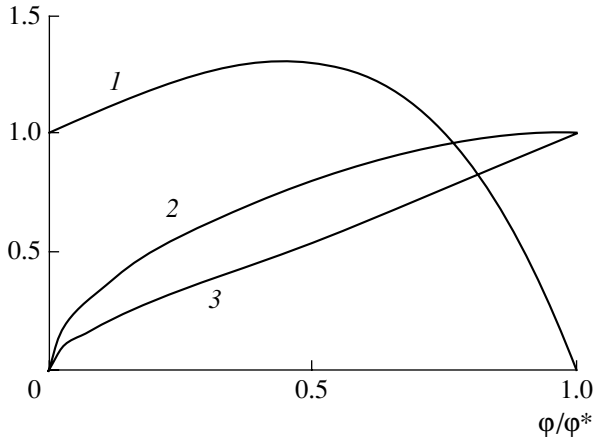


Fig. 2. Typical interrelation between the reaction of a neighboring fragment and the angular and transverse velocities of a *k*th fragment (according to the calculation data for the first scattering phase): (*1*) $\frac{Q_k(\varphi)}{Q_k(0)}$; (*2*) $\frac{\dot\varphi_k(\varphi)}{\dot\varphi_k^*}$; (*3*) $\frac{\dot\eta_k(\varphi)}{\dot\eta_k^*}$.

Here, $a = 2^{-5/3}(1 + 2^{-2/3})C_\xi \approx 0.77$ and $I_{Ck} = \frac{M_k r_k^2}{3}$ is the fragment moment of inertia with respect to C_k. The initial conditions (for $t = 0$) are $\varphi = 0$, $\dot\varphi = 0$, $\xi = 0$, and $\dot\xi = V_k$.

In the general case, solving the set of Eqs. (2)–(4) in quadratures is impossible, because the variables ξ and φ are not separated. For their separation, we can use, e.g., the approximation

$$\dot\xi(\xi, \varphi) = V_k \exp\left[-\frac{\xi \sin\alpha}{2H}\right]\psi(\varphi),$$

where $\psi(\varphi)$ is an empiric function determined numerically. In particular, for $\psi(\varphi) = 1$,

$$\dot\xi = V_k \exp\left[-\frac{\xi \sin\alpha}{2H}\right] \tag{5}$$

and the solution to the set of Eqs. (2)–(4) is of the form

$$\dot\varphi_k = \sqrt{a}\frac{V_k}{r_k}\sqrt{\frac{\rho_{ak}}{\rho_b}} \times \sqrt{\frac{2(\varphi_0 + \varphi) - \sin 2(\varphi_0 + \varphi) - (2\varphi_0 - \sin 2\varphi_0)}{\frac{5I_{Ck}}{M_k r_k^2} + \cos 2(\varphi_0 + \varphi)}}. \tag{6}$$

With due regard to this expression, Eq. (4) yields the following formula for angular acceleration:

$$\ddot\varphi_k = \frac{\dot\varphi^2 \sin 2(\varphi + \varphi_0) + aV_k^2 r_k^{-2}\frac{\rho_{ak}}{\rho_b}[1 - \cos 2(\varphi + \varphi_0)]}{\frac{5I_{Ck}}{M_k r_k^2} + \cos 2(\varphi_0 + \varphi)}. \tag{7}$$

As follows from (3), with allowance made for (6) and (7), with the rotation of a fragment, the quantity $Q_k(\varphi)$ rises from $Q_k(0) \approx \frac{3\rho_{ak} V_k^2 r_k^2 (1 - \cos 2\varphi_0)}{5/3 + \cos 2\varphi_0}$ to its maximum value, after which, at a certain critical instant at $t = t_k^*$ and $\varphi = \varphi^*$, it vanishes (Fig. 2). This implies cessation of the contact between the fragments. For an arbitrary k, $\varphi^* = \text{const}$ and depends only on the shape of the fragment, the basis parameter of which is r_k (in the given case, $r_k \approx 0.507 L_{k-1}$, $\varphi^* \approx 27°$, and $\varphi_0 + \varphi^* \approx 65°20'$).[5] For $0 \le \varphi \le \varphi^*$, the values of $\dot\varphi$ and $\dot\eta_{Ck}$ rise to

$$\dot\varphi_k^* \approx 0.94\frac{V_k}{r_k}\sqrt{\frac{\rho_{ak}}{\rho_b}},$$

and

$$\dot\eta_{Ck}^* = r_k \dot\varphi_k^* \cos(\varphi_0 + \varphi^*) \approx 0.4 V_k \sqrt{\frac{\rho_{ak}}{\rho_b}}$$

(Fig. 2). It is evident that $\dot\varphi_k^*$ has a strong inverse dependence on both the body's density and size, whereas $\dot\eta_{Ck}^*$ is dependent only on the body's density. This implies that fragments of denser and larger bodies, all other factors being the same, rotate and fly out more slowly than in the opposite case. Integrating (6) for $0 \le \varphi \le \varphi^*$ and (5) for $0 \le t \le t_k^*$, we obtain the fragment rotation time

$$t_k^* = \int_0^{\varphi^*} \dot\varphi^{-1} d\varphi \approx \frac{1.46 r_k}{V_k}\sqrt{\frac{\rho_b}{\rho_{ak}}}$$

and the corresponding displacement C_{k-1} along the trajectory

$$\xi_k^* = \int_0^{t_k^*} \dot\xi^{-1} d\xi = \frac{2H}{\sin\alpha}\ln\left(1 + \frac{V_k t_k^* \sin\alpha}{2H}\right)$$

with the velocity at the end of the displacement, according to (5), being

$$V_k^* = \dot\xi(\xi_k^*) = V_k \exp\left[-\frac{\xi_k^* \sin\alpha}{2H}\right].$$

[5] In the case of a sphere, $r_k \approx 0.534 D_{k-1}$ and $\varphi^* \approx 45°$, but $\varphi_0 + \varphi^* \approx 65°33'$, which is close to the case of a parallelepiped.

At $t > t_k^*$, the motion of kth fragments corresponds to the second phase of scattering and divergence with respect to the trajectory of motion C_{k-1}. After each fragmentation event, reorientation of the fracture surface and, correspondingly, of the vector $\dot{\eta}_{Ck}^*$ takes place. Therefore, after the first disintegration, the fragment transverse velocity is $V_{\eta 1} = \dot{\eta}_{C1}^*$. However, further on (for $k > 1$), the transverse velocity of new fragments is estimated as $V_{\eta k} \approx \sqrt{V_{\eta k-1}^2 + \dot{\eta}_{Ck}^{*2}}$. This velocity is considered to be constant until the next fragmentation event occurs, since, usually, in accordance with the results of calculations, $V_{\eta k} \ll V_k^*$. In this case, the trajectories C_k deviate from the trajectory C_{k-1} by the angle $\arctan\frac{V_{\eta k}}{V_k^*}$, and, prior to the next fragmentation, they fly apart over new trajectories at the velocity

$$V(x) = \sqrt{V_k^{*2} + V_{\eta k}^2}\exp\{-2^{k/3}A \times [\exp(-x) - \exp(x_k^* - x_k)]\},$$

where $x_k^* = \frac{\xi_k^* \sin\alpha}{H}$. The time and radius of their scattering are

$$t_{fk} \approx H\int_{x_{k+1}}^{x_k^*} \frac{dx}{V(x)},$$

and

$$R_{\eta k} \approx R_{\eta k-1} + r_k \sin(\varphi_0 + \varphi^*) + V_{\eta k-1} t_k^* + V_{\eta k} t_{fk},$$

respectively.

Upon completion of the fragmentation process and rotation of the fragments by the angle $\varphi = \varphi^*$ (for $t = t_n^*$), the altitudes $z_n^* = z_n - \xi_n^* \sin\alpha$ above the planet's surface and the length of the final trajectory segment $\xi_{\text{fin}} = \frac{z_n^*}{\sin\alpha}$ preceding the fall onto the ground are determined. Here, at altitude $z_{\exp} \approx H\ln(2^{1+n/3}A)$ [5], the peak intensity of the fragment-energy release (i.e., of the energy transfer to the atmosphere) is attained, which results in the effect of the SSCB explosion. Then, using (1), where V_n should be replaced by V_n^*, we can find scattering time $t_{\text{fin}} = H\int_0^{x_n^*} \frac{dx}{V(x)}$ of the fragments (here, $x_n^* = \frac{z_n^*}{H}$), and the scattering-spot radius at the point of crossing the ground by the trajectory is $R_{\text{imp}} \approx R_{\eta n} + V_{\eta n} t_{\text{fin}}$. In this case, the fragments fly apart in the space of the frustum of a cone with its axis coinciding with the initial trajectory and the elliptic base on the ground, which is the primary scattering field. The width (X) and length (Y) of this field are estimated as $X \approx Y\sin\alpha$ and

$$Y \approx 2R_{\text{imp}}\left[1 + \frac{\cos^2\alpha}{\sin(\alpha+\beta)\sin(\alpha-\beta)}\right]\sin\alpha,$$

$$\beta = \arctan\frac{R_{\text{imp}} - R_{\eta n}}{\xi_{\text{fin}}}.$$

Calculated parameters of fragmentation and scattering for the Sikhoté-Alin' meteoroid (iron, M_0 = 100 t, L_0 = 2.34 m, V_0 = 20 km s^{-1})

Parameter	$k = 1$	$k = 2$	$k = 3$	$k = 4$
Fragmentation stage				
z_k, km [3]	17.07	15.90	14.54	12.73
V_k, km s^{-1} [3]	18.47	18.18	17.70	16.74
L_k, m [3]	1.86	1.47	1.17	0.93
ξ_k, km	–	1.83	2.12	2.81
Scattering stage				
$\dot{\varphi}_k^*$, s^{-1}	64.52	86.08	114.99	153.45
$\dot{\eta}_{Ck}^*$, m s^{-1}	31.92	33.81	35.83	37.95
$V_{\eta k}^*$, m s^{-1}	31.92	46.50	58.70	69.90
t_k^*, 10^{-2} s	2.12	1.59	1.19	0.89
ξ_k^*, km	0.39	0.29	0.21	0.15
$\frac{\xi_k^*}{\xi_{k+1}}$, %	21.20	13.60	7.40	0.75
V_k^*, km s^{-1}	18.19	17.97	17.55	16.64
z_k^*, km	16.82	15.71	14.40	12.63
t_{fk}, s	0.08	0.10	0.15	1.35
$R_{\eta k}$, m	3.61	9.74	19.88	115.51

As an example, we now demonstrate the given SSCB-scattering model for the Sikhoté-Alin' meteoroid (iron with nickel additions, ρ_b = 7800 kg m^{-3}). The size of the scattering field for the largest craters is known; it attains ≈0.3 × 0.5 km [9] or ≈0.4 × 0.7 km [10]. In accordance with [5], the fragmentation stage for this SSCB is completed after three- or four-fold fragmentation, which depends on the initial mass and velocity (300 t and 12 km s^{-1} or 100 t and 20 km s^{-1}, respectively). It is worth noting that these data are in satisfactory agreement with the previously published conclusions of the Meteoritic Commission of the Academy of Sciences of the Soviet Union [11]. For these data, three basic disintegration stages were selected, which had been proceeded at three different velocities of meteoroid motion. As the initial data, we took the results obtained in [5]. In particular, according to these

data, the altitude corresponding to the peak intensity of the energy release (explosion) of the SSCB after its fragmentation is ≈10 km (M_0 = 100 t) and ≈5 km (M_0 = 300 t).

Numerically solving the set of Eqs. (2), (4) has demonstrated the possibility of their analytical solution in form (6), since the absolute error of the approximation of $\dot{\xi}(\xi, \varphi)$ by formula (5) does not exceed 0.6%. In the table, examples of calculation results for scattering fragments of the SSCB (M_0 = 100 t) are presented. The calculated sizes of the fragment-scattering ellipsis on the ground were $X \times Y \approx 0.23 \times 0.36$ km and $\approx 0.14 \times 0.22$ km for the SSCB mass M_0 = 100 t and 300 t, respectively, i.e., ≈1.5 times smaller in the second case. We should note that similar calculations performed on the basis of initial data of [5] for scattering the spherical SSCB yields virtually the same sizes for the scattering ellipsis (differences do not exceed 6%). If the actual sizes of the crater field relate to the primary scattering field, then we can consider their agreement with the calculated results as satisfactory and it is better for M_0 = 100 t than for M_0 = 300 t.

It should be noted that the presented model of SSCB scattering is valid for $\frac{\xi_k^*}{\xi_{k+1}} < 1$, i.e., when the kth fragments manage to rotate by the angle φ^*. Otherwise, their scattering does not occur and, prior to the next fracture event, they move in the mutual contact. In the case of the Sikhoté-Alin' meteoroid, this condition is fulfilled (see table).

The pioneering study [12] belonging to previously published models of the SSCB scattering is based on the analysis of different crater fields, which results in a significant indeterminacy of the final results. In the model developed in [7, 8], the initial conditions were given arbitrarily, which has led to loss of the rotational component of motion for fragments flying apart. However, in the process of SSCB scattering, this component can play a noticeable role (see, e.g., [13]).

Thus, the proposed model of SSCB scattering as a final stage of SSCB disintegration, taken together with the model of the preceding fragmentation stage, makes it possible to thoroughly study the process of SSCB interaction with the atmosphere. In particular, this model allows us to estimate the altitudes at which the fragmentation occurs and the explosion effect arises. The model is also capable of estimating the number and final sizes of fragments, velocities of their trajectory motion, rotation, and scattering, as well as the sizes of the primary crater field. The satisfactory agreement with the measurement data and the consistency of the calculation results for the disintegration of a rectangular and spherical SSCBs also allow us to expect the same effect for intermediate-shape SSCBs.

ACKNOWLEDGMENTS

The authors are deeply grateful to Academician V.M. Titov for his interest in the present work, as well as in our previous study [1].

REFERENCES

1. A. G. Ivanov and V. A. Ryzhanskiĭ, Dokl. Akad. Nauk **353**, 334 (1997) [Phys. Dokl. **42**, 139 (1997)].
2. V. P. Stulov and L. Yu. Titova, Dokl. Akad. Nauk **376**, 53 (2001) [Dokl. Phys. **46**, 50 (2001)].
3. N. G. Barri and V. P. Stulov, Dokl. Akad. Nauk **385**, 186 (2002) [Dokl. Phys. **47**, 528 (2002)].
4. A. G. Ivanov, Dokl. Akad. Nauk SSSR **310**, 866 (1990) [Sov. Phys. Dokl. **35**, 170 (1990)].
5. A. G. Ivanov and V. A. Ryzhanskiĭ, Fiz. Goren. Vzryva **35**, 120 (1999).
6. Yu. I. Fadeenko, Fiz. Goren. Vzryva **3**, 276 (1967).
7. N. A. Artem'eva and V. V. Shuvalov, Shock Waves, No. 5, 359 (1996).
8. N. A. Artem'eva and V. V. Shuvalov, J. Geophys. Res. **106** (E2), 3297 (2001).
9. I. V. Nemchinov and O. P. Popova, Astron. Vestn. **31**, 458 (1997).
10. V. V. Svetsov, Astron. Vestn. **32**, 76 (1998).
11. E. Krinov, *Iron Rain* (Nauka, Moscow, 1981).
12. Q. R. Passey and H. J. Melosh, Icarus **42**, 211 (1980).
13. E. Asphaug and W. Benz, Nature **370**, 120 (1994).

Translated by G. Merzon

Doklady Physics, Vol. 49, No. 10, 2004, pp. 583–587. Translated from Doklady Akademii Nauk, Vol. 398, No. 4, 2004, pp. 487–491.

MECHANICS

Stability of Kolmogorov Flow in a Channel with Rigid Walls

E. I. Oparina and O. V. Troshkin

Presented by Academician O.M. Belotserkovskiĭ April 12, 2004

Received April 12, 2004

FORMULATION OF THE PROBLEM

In this paper, we analyze the time-dependent equation of the stream function ψ for plane-parallel flow of a viscous incompressible fluid in a channel with rigid walls:

$$-\frac{\partial}{\partial t}\Delta\psi + \varepsilon\Delta\Delta\psi - \frac{\partial\psi}{\partial y}\frac{\partial\Delta\psi}{\partial y} + \frac{\partial\psi}{\partial x}\frac{\partial\Delta\psi}{\partial y} = \varepsilon\cos y,$$

$$0 < x < \frac{2\pi}{\alpha}, \quad 0 < y < 2\pi, \quad \varepsilon, \alpha = \text{const} > 0, \qquad (1)$$

$$\Delta \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

The equation is characterized by the given period in the longitudinal x direction,

$$\psi\left(t, x + \frac{2\pi}{\alpha}, y\right) = \psi(t, x, y) \qquad (2)$$

and by the conditions of impenetrability and adherence at the boundary,

$$\psi|_{y=0} = \left.\frac{\partial\psi}{\partial y}\right|_{y=0} = \psi|_{y=2\pi} = \left.\frac{\partial\psi}{\partial y}\right|_{y=2\pi} = 0. \qquad (3)$$

The third equality in (3) determines the zero fluid-flow rate.

The replacement of given boundary conditions (3) by the condition of periodicity in the transverse y direction, i.e., $\psi(t, x, y + 2\pi) = \psi(t, x, y)$ leads to the well-known Kolmogorov problem [1], the solution of which

$$\psi = \cos y - 1 \qquad (4)$$

is stable for arbitrary Reynolds numbers

$$\text{Re} = \frac{1}{\varepsilon}, \quad 0 < \text{Re} < +\infty \qquad (5)$$

or is globally stable in a channel with a short period $\frac{2\pi}{\alpha}$, i.e., for reasonably large wave numbers α,

$$\frac{2\pi}{\alpha} \le 2\pi \quad \text{or} \quad \alpha \ge 1, \qquad (6)$$

or, which is the same thing, for limited wave numbers $\frac{2\pi}{\alpha}$ in the longitudinal direction [2].

We should note that, at the same time, flow (4) is the exact solution to Eq. (1) in the case of boundary conditions (3).

In the present paper, we prove that Kolmogorov flow (4) in channel (1), (2) with short period (6) also remains globally stable in the sense of (5) in the presence of rigid walls (3).

THEORETICAL ANALYSIS OF THE PROBLEM

The constructions being developed are based on the representation of boundary value problem (1)–(3), which was proposed in [3] in the form of an abstract model of an infinite-dimensional dissipative top, the general equation of which is written as

$$\frac{d}{dt}A\psi + \varepsilon B\psi + [\psi, A\psi] = \varepsilon f. \qquad (7)$$

In the case under study, the operations of inertia A and dissipation B, as well as the commutator $[\psi, \chi]$ entering into Eq. (7), are determined as narrowing of the Laplacian, of the biharmonic operator, and of the Poisson brackets, respectively under the assumption of the validity of conditions (2) and (3):

$$A \equiv -\Delta, \quad B \equiv \Delta\Delta, \quad [\psi, \chi] \equiv \frac{\partial\psi}{\partial y}\frac{\partial\chi}{dx} - \frac{\partial\psi}{\partial x}\frac{\partial\chi}{dy}. \qquad (8)$$

At the same time, steady-state flow (4) is represented by the principal rotation of the top, i.e., satisfies the conditions of the following eigenvalue problem:

$$B\psi = \lambda A\psi. \qquad (9)$$

The corresponding principal moment of inertia,

$$\lambda = 1, \qquad (10)$$

Institute of Design Automation,
Russian Academy of Sciences,
Vtoraya Brestskaya ul. 19/18, Moscow, 123056 Russia
e-mail: a.oparina@icad.org.ru

1028-3358/04/4910-0583$26.00

turns out to be the smallest eigenvalue of problem (9) for short channel (6). Indeed, as is easy to see, in the case under consideration, the solutions to problem (9) are represented as harmonics with separated variables

$$\begin{aligned} &\psi = \varphi(y)\cos(\alpha n x) \\ \text{or } &\psi = \varphi(y)\sin(\alpha n x), \quad n = 0, 1, \ldots, \end{aligned} \tag{11}$$

and with a common amplitude $\varphi(y)$ that satisfies the conditions

$$\begin{aligned} &(D^2 - \alpha^2 n^2)(D^2 - \alpha^2 n^2 + \lambda)\varphi = 0, \\ &0 < y < 2\pi, \quad D \equiv \frac{d}{dy}, \\ &\varphi(0) = \varphi'(0) = \varphi(2\pi) = \varphi'(2\pi) = 0. \end{aligned} \tag{12}$$

In this case, eigenvalue (10) turns out to be minimal for a series of solutions to problem (12), which corresponds to $n = 0$. This can be easily confirmed by direct verification.

It is also easy to verify that, in the case of $n > 0$, the eigenvalues of (12) are not lower than $\alpha^2 n^2$. Then, from condition (6), the desired estimate,

$$\lambda \geq \alpha^2 n^2 \geq n^2 \geq 1, \quad n = 1, 2, \ldots,$$

immediately follows.

Thus, in the case of validity of condition (6), for all n in expression (11) and for all eigenvalues λ of problem (9), we have

$$\lambda \geq 1. \tag{13}$$

ANALYTICAL CALCULATIONS

Using the aforementioned minimal value of moment of inertia (10), we perform necessary calculations with the goal of obtaining the desired estimates concerning stability of solution (4).

First of all, we would note that this solution turns out to be the simple rotation of the top, in so as far as it simultaneously satisfies two identities:

$$\begin{aligned} &[\psi, A\psi] = 0 \quad \text{and} \quad B\psi = f \\ &(B\psi = A\psi \quad \text{and} \quad f = A\psi). \end{aligned} \tag{14}$$

In addition, the difference $A\psi - \lambda\psi$ turns out to be orthogonal with respect to the scalar product

$$(\psi, \chi) \equiv \int_0^{2\pi/\alpha} \int_0^{2\pi} \psi\chi \, dx dy \tag{15}$$

within the range of the top-commutator values

$$(A\psi - \lambda\psi, [\chi, \xi]) = 0, \quad \lambda = 1. \tag{16}$$

Indeed, it is easy to verify that

$$\begin{aligned} (A\psi - \lambda\psi, [\chi, \xi]) &= (\xi, [A\psi - \lambda\psi, \chi]) \\ &= -\int_0^{2\pi/\alpha} \int_0^{2\pi} \xi(\psi'' + \psi)'\chi_x dx dy = 0, \end{aligned}$$

because $(\psi'' + \psi)' = 0$.

We now vary Eq. (7) for the top, namely,

$$\begin{aligned} &\frac{d}{dt}A(\psi + \chi) + \varepsilon B(\psi + \chi) \\ &+ [\psi + \chi, A(\psi + \chi)] = \varepsilon(f + h). \end{aligned}$$

Then, we obtain, as a corollary, the following equation in variations:

$$\frac{d}{dt}A\chi + \varepsilon B\chi + [\psi, A\chi] + [\chi, A\psi] + [\chi, A\chi] = \varepsilon h.$$

Taking sequentially the scalar products of this equation by χ and $A\chi$, with allowance made for both symmetry of the operator A and the relations

$$\begin{aligned} (\chi, [\psi, A\chi]) &= (\psi, [A(\chi, \chi)]), \\ (\chi, [\chi, A\psi]) &= (\chi, [\chi, A\chi]) = 0, \\ (A\chi, [\psi, A]) &= (A\chi, [\chi, A\chi]) = 0, \\ (A\chi, [\chi, A\psi]) &= (A\psi, [A\chi, \chi]), \end{aligned}$$

which can be immediately verified, we arrive at the identities

$$\begin{aligned} &\frac{1}{2}\frac{d}{dt}(A\chi, \chi) + \varepsilon(B\chi, \chi) + (\psi, [A\chi, \chi]) = \varepsilon(h, \chi), \\ &\frac{1}{2}\frac{d}{dt}(A\chi, A\chi) + \varepsilon(B\chi, A\chi) + (A\psi, [A\chi, \chi]) = \varepsilon(h, A\chi). \end{aligned} \tag{17}$$

Then, multiplying the first of these relationships by λ and subtracting the result from the second one, we obtain

$$\begin{aligned} &\frac{1}{2}\frac{d}{dt}J(\chi) + \varepsilon K(\chi) = \varepsilon(h, A\chi - \lambda x), \\ &J(\chi) \equiv (A\chi, A\chi - \lambda\chi), \quad K(\chi) \equiv (B\chi, A\chi - \lambda\chi). \end{aligned} \tag{18}$$

Further constructions are based on the completeness and orthogonality of the system of principal rotations (9) proved in [3] and also the discreteness (or countability) of the set of their principal moments λ for both the positive definite operator A (as is the case in the problem under study) and the fulfilled identity

$$(B\chi, \chi) = (A\chi, A\chi), \quad (A\chi, \chi) > 0, \quad (\chi, \chi) > 0. \tag{19}$$

We now decompose the perturbation χ into its projection χ_λ onto the invariant space of the moment λ and its orthogonal complement ξ:

$$\chi = \chi_\lambda \oplus \xi, \quad B\chi_\lambda = \lambda A\chi_\lambda, \quad (\chi_\lambda, A\xi) = 0. \tag{20}$$

In this case, we can be convinced (at least formally) of the validity of further identities and inequalities in

which $\lambda' > \lambda$ is the principal moment of problem (9). This moment follows the minimal moment $\lambda = 1$:

$$J(\chi) = J(\xi), \quad K(\chi) = K(\xi),$$
$$K(\xi) \geq \lambda J(\xi), \quad \lambda = \min_{\eta > 0}\{\eta: B\chi = \eta A\chi\} = 1, \quad (21)$$
$$J(\xi) \geq (\lambda' - \lambda)(A\xi, \xi),$$
$$\lambda' = \min_{\mu > 1}\{\mu: B\xi = \mu A\xi\} > \lambda.$$

Indeed, using (19) and (20), we find

$$J(\chi) = (A\chi, A\chi) - \lambda(A\chi, \chi) = (B\xi - \lambda A\chi, \chi)$$
$$= (B\xi - \lambda A\xi, \chi) = J(\xi) - (B\xi - \lambda A\xi, \chi_\lambda)$$
$$= J(\xi) - (\xi, B\chi_\lambda) = J(\xi) - \lambda(\xi, A\chi_\lambda) = J(\xi),$$

$$K(\chi) = (B\chi, A\chi) - \lambda(A\chi, A\chi) = (B\chi - \lambda A\chi, A\chi)$$
$$= (B\xi - \lambda A\xi, A\chi) = K(\xi) + (B\xi, A\chi_\lambda) - \lambda(A\xi, A\chi_\lambda),$$
$$(A\xi, A\chi_\lambda) = (\xi, B\chi_\lambda) = \lambda(\xi, A\chi_\lambda) = 0.$$

Furthermore, decomposing ξ into the principal rotations,

$$\xi = \sum_\mu \xi_\mu, \quad B\xi_\mu = \mu A\xi_\mu,$$

we arrive at

$$(B\xi, A\chi_\lambda) = \sum_\mu \mu(A\xi_\mu, A\chi_\lambda)$$
$$= \sum_\mu \lambda\mu(\xi_\mu, A\chi_\lambda) = 0,$$

which leads to the second identity in (21).

Then, again decomposing ξ into the principal rotations, we have

$$K(\xi) = (B\xi - \lambda A\xi, A\xi) = \sum_\mu \sum_\eta (\mu - \lambda)(A\xi_\mu, A\xi_\eta)$$
$$= \sum_\mu \sum_\eta (\mu - \lambda)\mu(A\xi_\mu, \xi_\eta) = \sum_\mu (\mu - \lambda)\mu(A\xi_\mu, \xi_\mu)$$
$$\geq \lambda \sum_\mu (\mu - \lambda)(A\xi_\mu, \xi_\mu) = \lambda J(\xi)$$
$$\geq \lambda(\lambda' - \lambda)\sum_\mu (A\xi_\mu, \xi_\mu) = \lambda(\lambda' - \lambda)(A\xi, \xi),$$

which completes the proof of the validity of (21).

From (18) and (21), it follows as a corollary that

$$\frac{1}{2}\frac{d}{dt}J(\xi) + \varepsilon\lambda J(\xi) \leq \varepsilon(h, A\chi - \lambda\chi)$$

or, upon multiplying by $2e^{2\varepsilon\lambda t}$ and integrating over t, we obtain

$$J(\xi) \leq J(\xi_0)e^{-2\varepsilon\lambda t} + 2\varepsilon e^{-2\varepsilon\lambda t}\int_0^t e^{2\varepsilon\lambda t'}(h, A\chi - \lambda\chi)dt',$$
$$\xi_0 = \xi|_{t=0}.$$

At $h = 0$ (the absence of perturbations of the exterior moment of force), the inequality obtained, taken together with the last inequality of (21), results in the estimate

$$(A\xi, \xi) \leq \frac{J(\xi_0)}{\lambda' - \lambda}e^{-2\varepsilon\lambda t}. \quad (22)$$

In order to finally estimate the total perturbation χ, we employ (17), for which, with the help of (16), we find

$$(A\psi, [A\chi, \chi]) = \lambda(\psi, [A\chi, \chi]) = \lambda(A\chi, [\chi, \psi])$$
$$= \lambda(A\chi - \lambda\chi, [\chi, \psi]) = \lambda(A\xi - \lambda\xi, [\chi, \psi])$$
$$= \lambda(A\xi - \lambda\xi, [\xi + \chi_\lambda, \psi]) = \lambda(A\xi - \lambda\xi, [\xi, \psi])$$
$$= \lambda(A\xi, [\xi, \psi]).$$

(Here, the additional identity $[\chi_\lambda, \psi] = 0$ is used, which, in the case under consideration, is provided by the absence of the dependence of both χ_λ and ψ on the variable x.)

Furthermore, by virtue of boundary conditions (2) and (3), we find

$$(A\xi, [A\xi, \psi]) = \int_0^{2\pi/\alpha}\int_0^{2\pi}(\xi_{xx} + \xi_{yy})\xi_x\psi' dxdy$$
$$= \int_0^{2\pi}\psi' dy \int_0^{2\pi/\alpha}\left(\frac{\xi_x^2}{2}\right)_x dx + \int_0^{2\pi/\alpha} dx\int_0^{2\pi}(\xi_x\xi_y\psi')_y dy$$
$$- \int_0^{2\pi}\psi' dy \int_0^{2\pi/\alpha}\left(\frac{\xi_y^2}{2}\right)_x dx - \int_0^{2\pi}\int_0^{2\pi/\alpha}\xi_x\xi_y\psi'' dxdy$$
$$= \int_0^{2\pi}\int_0^{2\pi/\alpha}\xi_x\xi_y \sin y dxdy$$
$$\leq \frac{1}{2}\int_0^{2\pi}\int_0^{2\pi/\alpha}(\xi_x^2 + \xi_y^2)dxdy = \frac{1}{2}(A\xi, \xi),$$

which, in combination with the estimate

$$(B\chi, A\chi) = \sum_\mu \mu^2(A\chi_\mu, \chi_\mu) \geq \lambda\sum_\mu \mu(A\chi_\mu, \chi_\mu)$$
$$= \lambda(B\chi, \chi) = \lambda(A\chi, \chi)$$

for $h = 0$, yields that, in the left-hand side of (17), we have

$$\frac{1}{2}\frac{d}{dt}(A\chi, A\chi) + \varepsilon\lambda(A\chi, A\chi) \le \frac{\lambda}{2}(A\xi, \xi). \qquad (23)$$

From (22) and (23), we obtain the desired estimate providing the asymptotic stability of stem stream (4) under validity of condition (6).

$$(A\chi, A\chi) \le \left((A\chi_0, A\chi_0) + \frac{(A\xi_0, A\xi_0) - \lambda(A\xi_0, A\xi_0)}{\lambda' - \lambda}t\right)e^{-2\varepsilon\lambda t}.$$

DISCUSSION

As in the case of the Kolmogorov problem, further analysis of the stability of flow (4) under validity of conditions (3) is based on the investigation of an alternative possibility of the appropriateness of a channel with a long period $\frac{2\pi}{\alpha} > 2\pi$ (or $\alpha < 1$). This analysis is also based on theorems on the bifurcation of the given solution into a new steady-state or self-oscillating flow regime in the case of increasing Reynolds number (5) and attaining a certain critical value with it. This value is similar to that already investigated analytically in [4, 5] (but locally, i.e., near the critical value of the Reynolds number) and numerically (but globally, i.e., in the segment of post-critical values of the Reynolds number) [6] in the case of conditions periodic with respect to *y*.

A rather complete review of the corresponding results is presented in [7]. Related experiments are described in [8].

The present-day development of the noted classical topic is characterized by passing to spatial periodic flows and directly numerically modeling on the basis of parallel-processing supercomputers capable of involving up to 8000^3 harmonics of the Fourier expansion (or, which is the same thing, points in the $2\pi^3$ period) [9].

For a perfect (inviscid and incompressible) fluid, the model of an infinite-dimensional top was proposed in [10]. There, in the framework of this model and on the basis of the above-noted analogy, a theorem on the stability of plane-parallel flow was proved for the first time. In [3], this approach was extended to a viscous medium (passage from the compact Lie groups to extensions of the corresponding algebras). In this case, the initial concept of the Euler–Poisson theorem on the stability of the rotation about the minimal and maximal inertia axes has remained valid in both [10, 3] and the present study.

As was already noted, the alternative possibility of rotation about a middle (not extremal) principal inertia axis (the case of $\alpha < 1$) results in the stability loss of the stem stream and in the appearance of a new regime (periodic or self-oscillating). It should be emphasized that this transformation involves a finite number of harmonics (or is reasonably well described by their limited number). However, this fact cannot be considered as the appearance of turbulence the description of which requires practically the entire Fourier series in order to provide the necessary statistic.

For two-dimensional flows, similar calculations were performed in [11], where an actual channel with a given pressure drop was studied. Then, the averaging of the arising pulsations yielded characteristic velocity profiles and correlations that had been experimentally studied in [12].

It is important that, in this case, the passage to the turbulence occurred for the entire spectrum in the form of a soft jump similar to bifurcation. The corresponding model for such a passage, which is based on schemes of second-order closure, was proposed and analyzed in [13, 14], where the above-indicated averaged profiles and the response curve turned out to be close to the experimental ones that had been measured in [12, 15].

These facts, certainly, do not reduce the interest in the aforementioned preturbulent regimes for which the turbulent statistic increases with the Reynolds number. However, it seems that the related calculations, independently of the rate of their progress [9], will be doomed to successive failure in understanding the nature of turbulence. They can be successful only being maintained by adequate physical concepts and, for the first turn, by involving the phenomena of compressibility and thermal conductivity in the Navier-type model under discussion.

The latter statement becomes evident if we address the well-known Joule experiments devoted to measurements of the heat mechanical equivalent. We can note here that, in this case, *the transformation of mechanical work into thermal motion is realized just via turbulent fluid mixing.*

Thus, if we accept the necessity of dealing with a large number of modes in passing to turbulence, then, in addition to the arising processes of diffusion, dissipation, and relaxation (directed toward equilibrium) of a randomly fluctuating medium, we must allow for the phenomena of compressibility and thermal conductivity of fluid.

ACKNOWLEDGMENTS

The authors are grateful to Academician O.M. Belotserkovskiĭ for his attention to this study.

The work was supported by the Complex Program “Mathematical Simulation and Intellectual Systems” of the Presidium of the Russian Academy of Sciences.

REFERENCES

1. V. I. Arnold and A. D. Meshalkin, Usp. Mat. Nauk **15**, 247 (1960).

2. A. D. Meshalkin and Ya. G. Sinaĭ, Prikl. Mat. Mekh. **25** (6), 1140 (1961).
3. O. V. Troshkin, *Nontraditional Methods in Mathematical Hydromechanics, Translation Mathematical Monographs* (Am. Math. Soc., Providence, 1995), Vol. 44.
4. V. I. Yudovich, Prikl. Mat. Mekh. **29** (3), 453 (1965).
5. V. I. Yudovich, Izv. Akad. Nauk SSSR, Ser. Mekh. Zhidk. Gaza, No. 1, 32 (1973).
6. S. O. Belotserkovskiĭ, Candidate's Dissertation in Mathematical Physics (Moscow, 1979).
7. O. M. Belotserkovskiĭ, *Numerical Experiments in Turbulence Problems: from Order to Chaos* (Nauka, Moscow, 2000).
8. E. B. Gledzer, F. V. Doltanskiĭ, and A. M. Obukhov, *Sets of Hydrodynamic-Equations and Their Application* (Nauka, Moscow, 1981).
9. Y. Kaneda, T. Ishihara, M. Yokokawa, *et al.*, Phys. Fluid **15**, 21 (2003).
10. V. Arnold, Ann. Instr. Fourier (Grenoble) **16**, 319 (1966).
11. B. L. Rozhdestvenskiĭ and I. N. Simakin, Zh. Vych. Mat. Mat. Fiz. **25** (1), 96 (1985).
12. J. Conte-Belloux, *Turbulent Flow in a Channel with Parallel Walls* (Mir, Moscow, 1968).
13. D. N. Zubarev, V. G. Morozov, and O. V. Troshkin, Dokl. Akad. Nauk SSSR **290**, 313 (1986) [Sov. Phys. Dokl. **31**, 712 (1986)].
14. D. N. Zubarev, V. G. Morozov, O. V. Troshkin, *et al.*, Preprint No. 63, MI AN SSSR (Mathematical Inst., Academy of Sciences of the USSR, 1988).
15. L. G. Loĭtsyanskiĭ, *Mechanics of Fluids and Gases* (Nauka, Moscow, 1987).

Translated by G. Merzon

Doklady Physics, Vol. 49, No. 10, 2004, pp. 588–592. Translated from Doklady Akademii Nauk, Vol. 398, No. 5, 2004, pp. 615–619.

MECHANICS

On the Rigidity of Nearly Periodic Paths in the Joukowski Sense and Properties of Ultimate Motions in Dynamic Systems

O. V. Druzhinina and A. A. Shestakov

Presented by Academician V.V. Rumyantsev April 13, 2004

Received April 28, 2004

INTRODUCTION

Starting from the seminal paper [1], the problems of the stability of motions in dynamical systems have been treated in numerous publications (see, e.g., [2] and references therein). The rigidity of paths in dynamical systems was discussed in [3–8] and other papers. The stability of recurrent and nearly periodic paths and the properties of ultimate motions for these paths were studied in [9–14].

In the present paper, we study the properties of ultimate motions for rigid and asymptotically rigid paths in the Joukowski sense and prove theorems concerning the existence of rigid and asymptotically rigid paths in the Joukowski sense for dynamical systems. The results obtained demonstrate that rigid motions for nearly periodic rigid paths are also rigid and nearly periodic.

AUXILIARY PROPOSITIONS

Let a mechanical system be specified with many degrees of freedom modeled by a vector steady-state equation in the form

$$\frac{dx}{dt} = g(x), \quad x \in R^n, \quad n \geq 2. \tag{1}$$

Here, the n-dimensional vector function $g(x)$ is continuous on the open set $G \subset R^n$. Equation (1) in space R^n determines mechanical flux $\varphi(t, p)$ in the Birkhoff sense [9]. We consider a positive semi-path $C_\varphi^+ = \{\varphi(t, p): t \in R^+\}$ and infinitely growing sequence $0 \leq t_1 < t_2 < \ldots < t_n < \ldots$, $t_n \longrightarrow +\infty$ as $n \longrightarrow +\infty$. The limiting point p^* of sequence $\varphi(t_1, p)$, $\varphi(t_2, p)$, …, $\varphi(t_n, p)$, … is referred to as the ω-limiting point of motion $\varphi(t, p)$. Similarly, each limiting point p^* of the negative semi-path $C_\varphi^- = \{\varphi(t, p): t \in R^-\}$ is referred to as the α-limiting point of motion $\varphi(t, p)$. Motion $\varphi(t, p^*)$ passing through the limiting point p^* is referred to as the ultimate motion corresponding to Eq. (1).

We denote as Ω_φ and A_φ the sets of all ω-limiting and α-limiting points of motion $\varphi(t, p)$, respectively. Sets $H_\varphi^+ = C_\varphi^+ \cup \Omega_\varphi$ and $H_\varphi^- = C_\varphi^- \cup \Omega_\varphi$ are called positive and negative semi-shells of motion $\varphi(t, p)$, respectively, whereas set $H_\varphi ::= H_\varphi^+ \cup H_\varphi^-$ is referred to as the shell of motion $\varphi(t, p)$.

Lemma 1 [10]. *Sets Ω_φ and A_φ are closed invariant sets.*

Motion $\varphi(t, p)$ is referred to as L^+-stable (or L^--stable, or L-stable) if set H_φ^+ (or H_φ^-, or H_φ) is a compact set. The L-stability is also referred to as stability according to Lagrange [10, 13]. From this definition, it follows that the ω-limiting set Ω_φ is not empty for the L^+-stable motion $\varphi(t, p)$ and the α-limiting set A_φ is not empty for the L^--stable motion.

Lemma 2 [10]. *If motion $\varphi(t, p)$ is L^+-stable, then the Ω_φ set is the connected set and $d(\varphi(t, p), \Omega_\varphi) = 0$ at $t \longrightarrow +\infty$, where $d(\cdot, \cdot)$ is the distance between the point and the set.*

Set $M \subset R^n$ is referred to as minimal if it is not empty, closed, and invariant and has no true subset with these three properties.

Lemma 3 [10]. *An invariant compact closed set always contains the minimal set.*

Motion $\varphi(t, p)$ of Eq. (1) is referred to as *recurrent*, if for any $\varepsilon > 0$, we can find number $T = T(\varepsilon) > 0$ such that an arbitrary arc of the path $C_\varphi = C_\varphi^+ \cup C_\varphi^-$ corresponding to the time length T approximates the C_φ path with an accuracy equal to ε.

Lemma 4 [10]. *Any path belonging to a minimal compact set is a recurrent path.*

Lemma 5 [10]. *The L-stable motion for Eq.* (1) *is recurrent if the following necessary and sufficient condition is met: for any $\varepsilon > 0$, the set of t values obeying the inequality $\rho(\varphi(t, p), p) < \varepsilon$ should be relatively dense.*

Russian State Open Technical University for Railway Transportation, Chasovaya ul. 22/2, Moscow, 125993 Russia

1028-3358/04/4910-0588$26.00

Motion $\varphi(t, p)$ of Eq. (1) is referred to as nearly periodic if, for any $\varepsilon > 0$, we can find number $L = L(\varepsilon)$ determining a relatively dense set $\{\tau_n\}$ such that $\rho(\varphi(t, p), \varphi(t + \tau_n, p)) < \varepsilon$ for $t \in R$. From this definition, it follows that any nearly periodic motion is recurrent motion.

DEFINITIONS FOR THE CONDITIONAL RIGIDITY OF THE PATH AND CONDITIONAL STABILITY OF THE PATH

We now introduce the following notation. Symbols $B(y, r)$ and $B[y, r]$ denote, respectively, an open and closed spherical sets of radius $r > 0$ with the center at the point $y \in R^n$.

Definition 1. Let $t_0 \in R$ be an arbitrary but fixed number. Path C_φ of motion $\varphi(t)$, $\varphi(t_0) = p$ for Eq. (1) is referred to as

(a) positive (negative, or double-side) *rigid in the Joukowski sense with respect to set* $Q \subseteq R^n$ if there exists a number $\mathcal{T} > 0$ such that, for an arbitrary $\varepsilon > 0$, we have $\delta = \delta(\varepsilon)$ with the following property: if $\psi(t) = \psi(t, q)$, $q \in Q$, is an inextensible solution to Eq. (1) and, if there exist numbers τ_1, τ_2, $|\tau_1 - \tau_2| \le \mathcal{T}$ such that for

$$\rho(\psi(\tau_1), \varphi(\tau_2)) < \delta, \tag{2}$$

we have, for all $t \in R^+$ (for all $t \in R^-$, or for all $t \in R$),

$$\rho(\psi(t + \tau_1), \varphi(t + \tau_2)) < \varepsilon, \tag{3}$$

where $\rho(\cdot, \cdot)$ is the distance between two points.

If, in addition to condition (3), the following condition is met:

$$\exists \tau_3 \lim_{t \to +\infty} \rho(\psi(t), \varphi(\tau_3 + t)) = 0, \tag{4}$$

then the path is referred to as positive (negative, or double-side) *asymptotically rigid* in the Joukowski sense with respect to set $Q \subseteq R^n$;

(b) the positive (negative, or double-side) *rigid in the Joukowski sense with respect to set* $Q \subseteq R^n$ according to the measure of set M if for an arbitrary $\varepsilon > 0$, there exists a number $\mathcal{S} > 0$ and the measurable set $M \subset B(p, \varepsilon)$ with a measure larger than zero such that, for $q \in Q \cap M$ and for all $t \in R^+$ (for all $t \in R^-$, or for all $t \in R$), we can find a number τ_1 such that, for $|t - \tau_1| \le \mathcal{T}$, the following inequality is met:

$$\rho(\psi(t), \varphi(\tau_1)) < \varepsilon, \quad \psi(t) ::= \varphi(t, q); \tag{5}$$

(c) the positive (negative, or double-side) *rigid in the Joukowski sense with respect to itself* if there exists a number τ_1, τ_2, and $\mathcal{T}$ such that, for $\rho(\varphi(\tau_1), \varphi(\tau_2)) < \delta$ and $|\tau_1 - \tau_2| \le \mathcal{T}$, we have $\rho(\varphi(t + \tau_1), \varphi(t + \tau_2)) < \varepsilon$ at all $t \in R^+$ (for all $t \in R^-$, or for all $t \in R$).

It is evident that Definition 1a is a particular case of Definition 1b. To verify this, it is sufficient to set $M = B(p, \delta)$ and $\mathcal{T} = 0$. The path double-side rigid with respect to Q (or with respect to itself), we also call rigid with respect to Q (or with respect to itself).

In [7, 8], the authors proposed a definition for the rigidity of the path according to which rigidity in the Joukowski sense corresponds to stability in the Lyapunov sense at an appropriate reparametrization $\tau_1 = \tau_1(t)$ for the unperturbed path C_φ and an appropriate reparametrization $\tau_2 = \tau_2(t)$ for each perturbed path. Here, $\tau_1(t)$ and $\tau_2(t)$ belong to set Σ of all one-one and continuous mappings $\tau(t)$: $R^+ \longrightarrow R^+$ such that $\tau(0) = 0$. In Definition 1a, mappings $\tau_1(t)$ and $\tau_2(t)$ are replaced by numbers τ_1 and τ_2, where $|\tau_1 - \tau_2| \le \mathcal{T}$, $\mathcal{T}$ being a given number.

Definition 2. Let $t_0 \in R$ be an arbitrary but fixed number. Path C_φ (or motion $\varphi(t) = \varphi(t, p)$, $\varphi(t_0) = p$) is referred to as

(a) the positive (negative, or double-side) *stable in the Lyapunov sense with respect to set* $Q \subseteq R^n$ if for each spherical set $B(p, \varepsilon)$ there exists a spherical set $B(p, \delta)$, $\delta < \varepsilon$ such that as soon as $q \in B(p, \delta)$, $q \in Q$, we have, for all $t \ge 0$ (for all $t \le 0$, or for all $t \in R$),

$$\rho(\psi(t), \varphi(t)) < \varepsilon, \tag{6}$$

where $\psi(t) = \varphi(t, q)$, $\varphi(t_0, q) \ne p$ is a solution to Eq. (1);

(b) the positive (negative, or double-side) *stable in the Lyapunov sense according to the measure* M *with respect to set* $Q \subseteq R^n$ if for each spherical set $B(p, \varepsilon)$ there exists a measurable set M with a nonzero measure such that as soon as $q \in M$, $q \in Q$, inequality (4) is met for all $t \ge 0$ (for all $t \le 0$, or for all $t \in R$).

For equilibrium states, the corresponding Definitions 1 and 2 are equivalent, but they are not equivalent for periodic solutions. In Definition 2, the requirement that number t_0 is an arbitrary fixed number could be replaced by the requirement that inequality $\rho(\varphi(t_0, p), \varphi(t_0, p)) < \delta$ should be met only up to a certain value of t_0. However, in the latter case, we have a stronger definition of stability in the Lyapunov sense.

From Definitions 1 and 2, as well as from the definition of orbital stability [15], it follows that the rigidity in the Joukowski sense is a concept intermediate between the concepts of orbital stability and stability in the Lyapunov sense. Note that these concepts are independent. The rigidity in the Joukowski sense follows from stability in the Lyapunov sense, whereas the orbital stability follows from the rigidity in the Joukowski sense. Thus, the class of all motions stable in the Lyapunov sense is a subclass of all motions rigid in the Joukowski sense, and the class of all motions rigid in the Joukowski sense is a subclass of all orbital-stable motions. These subclasses are the proper subclasses.

Theorem 1. *Path* C_φ *of any nearly periodic motion* $\varphi(t) = \varphi(t, p)$ *is rigid in the Joukowski sense with respect to itself: for any* $\varepsilon > 0$, *we can find* $\delta > 0$ *such that* $\rho(\varphi(\tau_1 + t), \varphi(\tau_2 + t, p)) < \varepsilon$ *for* $t \in R$ *as soon as*

$\rho(\varphi(\tau_1), \varphi(\tau_2)) < \delta$ *becomes to hold. The reverse statement is also true: if path* C_φ *of motion* $\varphi(t)$ *is rigid in the Joukowski sense, then it is nearly periodic.*

Theorem 1 follows from the definitions of nearly periodicity, recurrence, and rigidity in itself, as well as from Theorem 1 formulated in [11]. Indeed, it was demonstrated in [11] that, in order to be nearly periodic, the *L*-stable motion of Eq. (1) must meet the following necessary and sufficient conditions: this motion should be recurrent and have property (3) for all $t \in R$, if condition (2) is met. In this Franklin theorem, the requirement of recurrence is needless, since any *L*-stable motion is recurrent if it obeys conditions (2) and (3).

PROPERTIES OF ULTIMATE MOTIONS RIGID IN THE JOUKOWSKI SENSE

We now discuss characteristic features of ω-limiting motions for rigid and asymptotically rigid paths of Eq. (1). A similar problem for the paths stable in the Lyapunov sense was studied in [12] and [13].

Theorem 2. *If path* C_φ *of* L^+*-stable motion* $\varphi(t)$ *of Eq.* (1) *is rigid in the Joukowski sense, then all* ω*-limiting paths* C_π *in* Ω_φ *are also rigid in the Joukowski sense, i.e., if we have*

$$C_\pi \subset \Omega_\varphi, \tag{7}$$

then C_π *is rigid in the Joukowski sense.*

Proof. Since C_φ is L^+-stable

$$\exists \tau_0\ \exists r > 0 \quad \varphi(t) \in B(0, r) \quad \forall t \geq \tau_0, \tag{8}$$

and then $\pi(t)$ is defined for all values of *t*. Let $\tau_0 \geq \mathcal{T}$, then it follows from Eq. (7) that there exists a number $\tau \geq \mathcal{T}$ such that the following inequality is met:

$$\rho(\varphi(\tau), \pi(\tau_0)) < \frac{1}{2}\delta\left\{\min\left(\frac{1}{2}\delta, \frac{1}{2}\varepsilon\right)\right\}. \tag{9}$$

Suppose that $\psi(t)$ is a solution to Eq. (1) and there exist numbers $\tau_1, \tau_2 \geq \mathcal{T}$, for which

$$\rho(\psi(\tau_1), \pi(\tau_2)) < \frac{1}{2}\delta\left(\frac{1}{2}\varepsilon\right). \tag{10}$$

Since path C_φ is rigid, it follows from inequality (9) that, at $t \geq 0$,

$$\rho(\varphi(t + \tau), \pi(t + \tau_0)) < \min\left\{\delta\left(\frac{1}{2}\varepsilon\right), \frac{1}{2}\varepsilon\right\}. \tag{11}$$

Suppose that $\tau_0 = \tau_2$. Then, relationships (9) and (10) yield the inequality

$$\rho(\psi(\tau_1), \varphi(\tau)) < \delta\left(\frac{1}{2}\varepsilon\right). \tag{12}$$

Hence, we have

$$\rho(\psi(t + \tau_1), \varphi(t + \tau)) < \frac{1}{2}\varepsilon \quad \forall t \in R^+. \tag{13}$$

From Eqs. (11) and (13), we find

$$\rho(\psi(t + \tau_1), \pi(t + \tau_0)) < \varepsilon \quad \forall t \in R^+. \tag{14}$$

Since $\tau_0 = \tau_2$, the rigidity of path C_π follows from inequality (14). Theorem 2 is proven.

Theorem 3. *Let* C_φ *be a* L^+*-stable motion* $\varphi(t)$ *of Eq.* (1). *Then, there exists a recurrent path* C_π *embedded into set* Ω_φ.

Proof. Theorem 3 is based on the fact that the set of ultimate solutions is a dynamic flux in the Birkhoff sense [9, 10].

Theorem 4. *Let* C_φ *be a recurrent path rigid in the Joukowski sense and embedded into the spherical set* $B(0, r)$ *at all* $t \in R$. *Then,* C_φ *is a nearly periodic path.*

Proof. We reparametrize, if necessary, $\varphi(t)$ to obtain $\mathcal{T} = 0$. Since C_φ is rigid in the Joukowski sense, then, for a given $\varepsilon > 0$, there exists $\delta = \delta\left(\frac{\varepsilon}{2}\right)$. Since $\varphi(t)$ is recurrent, Lemmas 1 and 5 imply that there exists a relatively dense sequence of numbers $\{\tau\}$ such that

$$\rho(\varphi(0), \varphi(\tau)) < \frac{1}{2}\delta. \tag{15}$$

If τ is specified, then, from the continuous dependence of a solution on the initial conditions, it follows that there exists a number $\mu > 0$ such that if $u(t)$ is a solution to Eq. (1) and

$$\rho(\varphi(0), u(0)) < \mu, \tag{16}$$

then we have

$$\rho(\varphi(\tau), u(\tau)) < \frac{1}{2}\delta. \tag{17}$$

If *t* is a given real number and $T_1 > 0$, then, due to recurrence of $\varphi(t)$, there exists a number τ_1 such that we have

$$\tau_1 < -T_1, \quad \tau_1 < t, \quad |\varphi(0) - \varphi(\tau_1)| < \min\{\mu, \delta\}. \tag{18}$$

Therefore, according to the definition of number μ, we obtain

$$\rho(\varphi(\tau), \varphi(\tau + \tau_1)) < \frac{1}{2}\delta. \tag{19}$$

From Eqs. (15) and (19), we find

$$\rho(\varphi(0), \varphi(\tau + \tau_1)) < \delta. \tag{20}$$

Since path C_φ is rigid at $\mathcal{T} = 0$, it follows from (18) that, at $t \geq 0$,

$$\rho(\varphi(t), \varphi(t + \tau_1)) < \frac{1}{2}\varepsilon. \tag{21}$$

From Eqs. (20) and (21), we have that, if $t \geq 0$, then

$$\rho(\varphi(t+\tau_1), \varphi(t+\tau_1+\tau)) < \varepsilon. \quad (22)$$

Since τ_1 can be chosen in such a manner that $\tau_1 < -T_1$, where $T_1 > 0$ is a given number, Theorem 4 is proven.

Theorem 5. *Let* (1) C_φ *be an L-stable path for motion* $\varphi(t)$ *of Eq.* (1) *rigid in the Joukowski sense and* (2) *there are no equilibrium states of Eq.* (1) *in the closed spherical set* $B[0, r]$. *Then, there exists a nontrivial nearly periodic solution* $\pi(t)$ *such that*

$$C_\pi \subset \Omega_\varphi. \quad (23)$$

Indeed, the formulation of Theorem 5 implies that $\Omega_\varphi \subset B[0, r]$ and hence $C_\pi \subset \Omega_\varphi$.

Theorem 6. *1) Let* C_φ *be a L-stable path for motion* $\varphi(t)$ *of Eq.* (1) *asymptotically rigid in the Joukowski sense and let path* C_ν *for solution* $\nu = \nu(t)$ *to Eq.* (1) *is such that*

$$C_\nu \subset \Omega_\varphi. \quad (24)$$

Then, path C_ν *is asymptotically rigid in the Joukowski sense.*

Proof. Let the conditions of the theorem be met. Based on Theorem 2 it is sufficient to prove that condition (4) of Definition 1 is valid at $\varphi(t) = \nu(t)$. For a given number $\varepsilon > 0$, let number δ correspond to number $\delta(\varepsilon)$ related to motion $\varphi(t)$. We suppose that there exist numbers τ_1 and τ_2 such that

$$\rho(\psi(\tau_1), \nu(\tau_2)) < \frac{1}{2}\delta(\varepsilon). \quad (25)$$

From Eq. (24), it follows that there exists a number $\tau_3 \geq \mathcal{T}$ such that

$$\rho(\varphi(\tau_3), \nu(\tau_2)) < \frac{1}{2}\delta(\varepsilon), \quad (26)$$

and, from Eqs. (25) and (26), we find that $\rho(\psi(\tau_1), \varphi(\tau_3)) < \delta(\varepsilon)$. Since path C_φ is asymptotically rigid in the Joukowski sense, there exist numbers τ_4 and $\bar{\tau}_4$ such that

$$\lim_{t\to\infty} \rho(\psi(t), \varphi(\tau_4 + t)) = 0, \quad (27.1)$$

$$\lim_{t\to\infty} \rho(\nu(t), \nu(\bar{\tau}_4 + t)) = 0. \quad (27.2)$$

Let us choose $\tau = t - \tau_4 + \bar{\tau}_4$. Then, we have

$$\lim_{\tau\to\infty} \rho(\nu(\tau + (\tau_4 - \bar{\tau}_4)), \varphi(\tau + \bar{\tau}_4 + (\tau_4 - \bar{\tau}_4)) = 0, \quad (28)$$

or

$$\lim_{\tau\to\infty} \rho(\nu(\tau + (\tau_4 - \bar{\tau}_4)), \varphi(\tau + \tau_4)) = 0. \quad (29)$$

From Eqs. $(27)_2$ and (29), we arrive at

$$\lim_{t\to\infty} \rho(\psi(t), \nu(t + (\tau_4 - \bar{\tau}_4)) = 0. \quad (30)$$

Thus, Theorem 6 is proven.

KEY THEOREMS AND THEIR COROLLARIES

The following theorems, which are the key theorems of the present study, are valid.

Theorem 7. *Let* C_φ *be an L-stable path for motion* $\varphi(t) = \varphi(t, p)$ *of Eq.* (1). *If path* C_φ *is rigid in the Joukowski sense with respect to* $Q \subset R^n$, *then there exists a nearly periodic path* C_π *of motion* $\pi(t) = \varphi(t, q)$, $q \in Q$ *completely embedded into set* Ω_φ, *that is,* $C_\pi \subset \Omega_\varphi$.

Theorem 8. *Let* C_φ *be an L-stable path for motion* $\varphi(t) = \varphi(t, p)$ *of Eq.* (1). *If path* C_φ *is asymptotically rigid in the Joukowski sense with respect to* $Q \subseteq R^n$, *then there exists a periodic path* C_ν *of motion* $\nu(t) = \varphi(t, q)$, $q \in Q$ *coinciding with* set Ω_φ, *that is,* $C_\nu = \Omega_\varphi$.

Theorem 7 is a corollary to Theorems 2, 3, and 4. Theorem 8 is a corollary to Theorem 7 and Theorems 5 and 6. Indeed, it follows from the formulation of Theorem 7 that there exists a nearly periodic rigid path C_π such that $C_\pi \subset \Omega_\varphi$. According to Theorem 5, path C_π is asymptotically rigid in the Joukowski sense. Due to the fact that the path is nearly periodic, we have $C_\pi \subset \Omega_\varphi$. Suppose that $\psi(t)$ is a recurrent motion such that $C_\psi \cap C_\pi = \emptyset$ and, in addition, $C_\psi \subset \Omega_\pi$.

According to Theorem 4, motion $\psi(t)$ is nearly periodic, and according to Theorem 5, path C_ψ is asymptotically rigid. Owing to the condition $C_\psi \subset \Omega_\pi$, there exists a number τ_1, such that

$$\rho(\psi(t+\tau_1), \pi(t)) \to 0 \quad \text{at} \quad t \to +\infty. \quad (31)$$

Since $\psi(t)$ and $\pi(t)$ are nearly periodic, $\psi(t+\tau_1)$ and $\pi(t)$ also are nearly periodic. Then, the following relationship holds at all values of t:

$$\pi(t) = \psi(t+\tau_1). \quad (32)$$

Therefore, the only recurrent motion $\psi(t)$, for which the condition $C_\psi \subset \Omega_\pi$ is met, is either motion $\pi(t)$, or motion $\pi(t+c)$, where c is a constant. Set Ω_π is closed, invariant, compact and contains a minimal set M, which is closed and, hence, compact. According to Lemmas 3 and 4, each path of a minimal compact set is recurrent. Since we proved that $\pi(t)$ is the only recurrent motion in Ω_π, we have $M = C_\pi$ and hence $\Omega_\pi \subset C_\pi$. Thus, $\pi(t)$ is a special motion (equilibrium or periodic). Due to the asymptotic rigidity of path C_π, the equality $\Omega_\pi = C_\pi$ holds, and Theorem 8 is proven.

These key theorems lead to:

Corollary 1. *Let* C_φ *be a* L^+*-stable asymptotically rigid path for Eq.* (1). *If there are no stable equilibrium*

states in the spherical set $B[0, r]$, *then there exists a nontrivial periodic solution* $\pi(t)$ *to Eq.* (1) *such that* $C_\pi = \Omega_\varphi$.

Let us consider the following condition, referred to as Condition Π.

Condition Π. *Let* (1) $x(t, p)$ *be a solution to Eq.* (1) *such that* $x(t, p) = p$; (2) *there exists a bounded open set* $K \subset G$ *such that if* $p \in \partial K$, *then for all* $t \geq t_0$, *for which solution* $x(t, p)$ *is defined, we have* $x(t, p) \in \overline{K}$, *where* ∂K *is the boundary of set* K *and* $\overline{K}$ *is its closure.*

The key theorems yield the following corollary:

Corollary 2. *Let* (1) *condition* Π *be met*; (2) *path* C_φ *for Eq.* (1) *be a rigid* (*asymptotically rigid*) *path in the Joukowski sense. Then, Eq.* (1) *has a rigid nearly periodic path* (*asymptotically rigid periodic path*).

ACKNOWLEDGMENTS

The authors are grateful to Academician V.V. Rumyantsev for his attention to the presented study.

REFERENCES

1. A. M. Lyapunov, *General Problems on Stability of Motion* (Izd. Kharkov. Mat. Obshch., Kharkov, 1892).
2. V. I. Vorotnikov and V. V. Rumyantsev, *Stability and Control with Respect to the Phase Vector of Dynamical Systems. Theory, Methods, and Application* (Nauch. Mir, Moscow, 2001).
3. N. E. Zhukovskiĭ, Uchen. Zap. Mos. Univ., Otd. Fiz.–Mat., No. 4, 1 (1882).
4. Dzh. Sindzh, *Tensor Methods in Dynamics* (Inostrannaya Literatura, Moscow, 1947).
5. M. Sh. Aminov, Tr. Kazan. Aviats. Inst. **24**, 3 (1949).
6. V. I. Zubov, *Lyapunov Methods and Their Application* (Leningrad. Gos. Univ., Leningrad, 1957).
7. G. A. Leonov and D. V. Ponomarenko, Izv. Vyssh. Uchebn. Zaved. Ser. Mat., No. 4 (371), 88 (1993).
8. O. V. Dunaeva, Dokl. Akad. Nauk **355**, 51 (1997) [Phys. Dokl. **42**, 374 (1997)].
9. Dzh. Birkhoff, *Dynamical Systems* (Gostekhizdat, Moscow, 1941).
10. V. V. Nemytskiĭ and V. V. Stepanov, *Qualitative Theory of Differential Equations* (GITTL, Moscow, 1949).
11. P. Franklin, Math. Z. **30**, 325 (1929).
12. A. A. Markoff, Math. Zeitschr. **36**, 708 (1933).
13. K. S. Sibirskiĭ, Izv. AN Mold. SSR, No. 1, 38 (1963).
14. P. Hartman and C. Olech, Trans. Am. Math. Soc. **104**, 154 (1962).
15. B. P. Demidovich, *Lectures on Mathematical Stability Theory* (Nauka, Moscow, 1967).

Translated by K. Kugel

Doklady Physics, Vol. 49, No. 10, 2004, pp. 593–597. Translated from Doklady Akademii Nauk, Vol. 398, No. 5, 2004, pp. 620–624.

MECHANICS

Stability of Motion of a Solid in the Steklov's Case

A. P. Markeev

Presented by Academician V.V. Kozlov April 20, 2004

Received April 21, 2004

FORMULATION OF THE PROBLEM

We consider motion of a solid about a fixed point O in a uniform gravity field. The weight of the body is mg, and the distance from the center of gravity to the fixed point is l. Let $Oxyz$ be a reference system rigidly connected to the body and its axes be the principal inertia axes of the body with respect to point O. The symbols A, B, and C denote the corresponding principal moments of inertia. The reference system $OXYZ$ is fixed, and its OZ axis is directed upward. The notation p, q, r and γ_1, γ_2, γ_3 corresponds to the components of the angular velocity for the body and of the unit vector of the ascending vertical axis OZ in the mobile reference system $Oxyz$.

In the Steklov's case [1], it is assumed that the center of gravity lies on the principal inertia axis. Let it be Ox axis.

The Euler–Poisson equations have the form

$$A\dot{p} = (B-C)qr, \quad B\dot{q} = (C-A)rp + mgl\gamma_3, \quad C\dot{r} = (A-B)pq - mgl\gamma_2; \tag{1}$$

$$\dot{\gamma}_1 = r\gamma_2 - q\gamma_3, \quad \dot{\gamma}_2 = p\gamma_3 - r\gamma_1, \quad \dot{\gamma}_3 = q\gamma_1 - p\gamma_2. \tag{2}$$

V.A. Steklov found [1] that, if quantities A, B, and C, in addition to the conditions

$$A + B > C, \quad B + C > A, \quad C + A > B, \tag{3}$$

which are usual for solids, also obey the constraints

$$B > A > 2C, \tag{4}$$

then Eqs. (1) and (2) admit a particular periodic solution expressed in terms of the Jacobian elliptic functions. According to Eq. (4), the center of gravity for the body lies at the medium-length axis of the ellipsoid of inertia. The periodic motion of the solid corresponding to this Steklov's solution is orbitally unstable [2].

Institute of Problems in Mechanics,
Russian Academy of Sciences,
pr. Vernadskogo 101, Moscow, 119526 Russia

However, it was shown in [3] that the relationships found in [1] actually imply the existence of one more particular periodic solution to Eqs. (1) and (2). For this particular solution, the center of gravity for the body lies in the shortest axis of the ellipsoid of inertia. In this case, the values of moments of inertia should meet [in addition to conditions (3)] the following inequalities:

$$2B > A > B > C, \quad A > 2C. \tag{5}$$

The period of time τ is independent of initial conditions, and its value is calculated according to the formula

$$\tau = 4kK(k)\sqrt{\frac{A-C}{mgl}}, \quad k^2 = \frac{1-\theta_b}{1-\theta_c}, \quad 0 < k < 1.$$

Here and below, we use notation typical of the theory of elliptic functions and integrals. The symbols θ_b and θ_c denote dimensionless inertia parameters,

$$\theta_b = \frac{B}{A}, \quad \theta_c = \frac{C}{A}.$$

In the θ_b, θ_c plane, domain (3), (5) of admissible parameter values is the interior of the right triangle $Q_1Q_2Q_3$ with vertices Q_1 (1, 0), Q_2(0.5, 0.5), and Q_3(1, 0.5) (see Fig. 1).

In this paper, we present the results of solving the problem on the orbital stability of periodic motion of a solid, which corresponds to the Steklov's solution to Euler–Poisson equations (1) and (2).

SOLUTION EXPRESSED IN TERMS OF CANONICALLY CONJUGATE VARIABLES

As the generalized coordinates, we take here the Euler angles ψ, θ, and φ introduced in a usual manner. ψ is a cyclic coordinate, and the momentum p_ψ corresponding to it is the first integral of equations of motion. For the Steklov's solution, the constant entering into this integral is equal to zero [4]. It is assumed that, for perturbed motion, we are also dealing with $p_\psi = 0$.

1028-3358/04/4910-0593$26.00

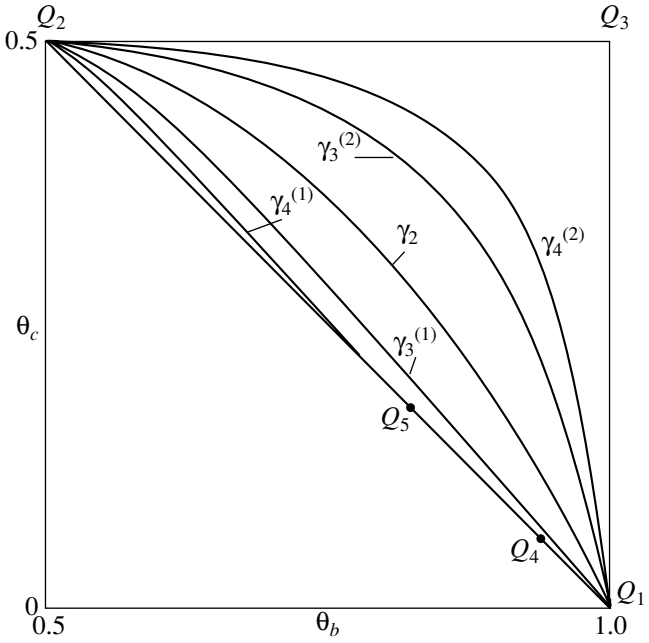


Fig. 1. The domain of admissible values of parameters and resonance curves.

We put $\varphi = q_1$, $\theta = q_2$, $p_\varphi = \sqrt{mglA}\,p_1$, $p_\theta = \sqrt{mglA}\,p_2$ and use a new independent variable,

$$w = \frac{2\pi}{\tau}(t + t_0), \quad t_0 = \text{const}.$$

The Hamiltonian function can be written as

$$F = \frac{kK(k)\sqrt{1-\theta_c}}{\pi}$$

$$\times\left\{\left[\frac{1}{\theta_c} + \left(\sin^2 q_1 + \frac{1}{\theta_b}\cos^2 q_1\right)\cot^2 q_2\right]p_1^2\right.$$

$$+\left(\frac{1}{\theta_b} - 1\right)\sin 2q_1 \cot q_2 p_1 p_2 \qquad (6)$$

$$\left.+\left(\cos^2 q_1 + \frac{1}{\theta_b}\sin^2 q_1\right)p_2^2 + 2\sin q_1 \sin q_2\right\}.$$

We can put the Steklov's solution of the Euler–Poisson equations under study in correspondence with a 2π-periodic solution with respect to w:

$$q_i = f_i(w), \quad p_i = g_i(w), \quad i = 1, 2, \qquad (7)$$

for the differential equations with Hamiltonian (6). In this case, we have

$$\sin f_1 \sin f_2 = -1 + \frac{1}{1-\theta_c}\mathrm{sn}^2 z,$$

$$\cos f_1 \sin f_2 = \sqrt{\frac{2\theta_b - 1}{(1-\theta_c)(\theta_b - \theta_c)}}\,\mathrm{snzcnz},$$

$$\cos f_2 = -\sqrt{\frac{1-2\theta_c}{(1-\theta_c)(\theta_b - \theta_c)}}\,\mathrm{snzdnz},$$

$$g_1 = \theta_c\sqrt{\frac{2\theta_b - 1}{(1-\theta_b)(1-\theta_c)(\theta_b - \theta_c)}}\,\mathrm{dnz}, \qquad (8)$$

$$g_2 = \sqrt{\frac{1-2\theta_c}{(1-\theta_c)(1-\theta_b)}}$$

$$\times\left[\sqrt{\frac{2\theta_b - 1}{1-\theta_c}}\cos f_1 \mathrm{snz} - \frac{\theta_b}{\sqrt{\theta_b - \theta_c}}\sin f_1 \mathrm{cnz}\right],$$

$$z = \frac{2K(k)}{\pi}w.$$

For solution (7), the Oz axis cannot pass through the vertical line; nutation angle q_2 periodically changes (with a period τ) between its minimum and maximum values lying within the segment $[0, \pi]$,

$$\arccos\mu < q_2 < \pi - \arccos\mu.$$

Here, $\mu = \frac{\sqrt{1-2\theta_c}}{1-\theta_c}$, if $0 < k^2 \le \frac{1}{2}$ and $\mu = \frac{1}{2}\sqrt{\frac{1-2\theta_c}{(1-\theta_b)(\theta_b - \theta_c)}}$, if $\frac{1}{2} \le k^2 < 1$.

The other property of solution (7) important for the further discussion is that the angle q_1 corresponding to the self-rotation is a monotonically growing function of time. Indeed, it follows from Eq. (8) that

$$\frac{dq_1}{dz} = \sqrt{\frac{2\theta_b - 1}{(1-\theta_c)(\theta_b - \theta_c)}}\frac{\mathrm{dnz}}{\sin^2 q_2}$$

$$\times\left(\mathrm{cn}^2 z + \frac{\theta_c}{1-\theta_c}\mathrm{sn}^2 z\right) > 0. \qquad (9)$$

THE HAMILTONIAN OF THE PERTURBED MOTION. ISOENERGETIC REDUCTION

To study the orbital stability of periodic motion (7), we introduce new canonically conjugate variables ξ_i, η_i

($i = 1, 2$) in such a manner that unperturbed motion (7) can be written in the form

$$\xi_1(w) = w + \xi_1(0), \quad \eta_1 = \xi_2 = \eta_2 = 0,$$

and the new Hamiltonian function $\Gamma(\xi_1, \xi_2, \eta_1, \eta_2)$ is a 2π-periodic solution with respect to ξ_1. Since (see the previous section) $\frac{df_1}{dw} \neq 0$, the variables ξ_i, η_i ($i = 1, 2$) could be introduced using the following canonical univalent transformation [5]:

$$\begin{aligned} q_1 &= f_1(\xi_1), \quad q_2 = f_2(\xi_1) + \xi_2, \\ p_1 &= g_1(\xi_1) + \left(\frac{df_1(\xi_1)}{d\xi_1}\right)^{-1} \\ &\times \left(\eta_1 + \frac{dg_2(\xi_1)}{d\xi_1}\xi_2 - \frac{df_2(\xi_1)}{d\xi_1}\eta_2\right), \\ p_2 &= g_2(\xi_1) + \eta_2. \end{aligned} \tag{10}$$

The problem concerning the orbital stability of the periodic motion (7) is equivalent to the stability problem for the system having the Hamiltonian Γ with respect to perturbations of parameters η_1, ξ_2, and η_2. Function Γ can be written in the form of a series expansion

$$\Gamma = \Gamma_2 + \Gamma_3 + \Gamma_4 + \ldots + \Gamma_k + \ldots, \tag{11}$$

where we omitted a constant equal to the value of the Hamiltonian for the unperturbed motion. Γ_k is a form of power k with respect to $|\eta_1|^{1/2}, \xi_2, \eta_2$. Here, we have

$$\begin{aligned} \Gamma_2 &= \eta_1 + \varphi_2(\xi_2, \eta_2, \xi_1), \\ \Gamma_3 &= \psi_1(\xi_2, \eta_2, \xi_1)\eta_1 + \varphi_3(\xi_2, \eta_2, \xi_1), \\ \Gamma_4 &= \chi(\xi_1)\eta_1^2 + \psi_2(\xi_2, \eta_2, \xi_1)\eta_1 + \varphi_4(\xi_2, \eta_2, \xi_1), \end{aligned}$$

where $\chi(\xi_1)$ is a 2π-periodic function with respect to ξ_1 and φ_m and ψ_m are forms of power m with respect to ξ_2 and η_2 having coefficients 2π-periodic with respect to ξ_1. The criteria for orbital stability and instability can be expressed [5–7] in terms of the coefficient in the normal form of the Hamiltonian for perturbed motion (11).

The necessary calculations can be significantly simplified, if we note that these conditions coincide with those of stability and instability for the states of equilibrium $\xi_2 = \eta_2 = 0$ for a reduced Hamiltonian system with a single degree of freedom, which describes the perturbed motion at the isoenergetic level $\Gamma = 0$. Solving the latter equality with respect to η_1, we find $\eta_1 = -H(\xi_2, \eta_2, \xi_1)$, where H is the 2π-periodic function with respect to ξ_1, which can be presented as the series

$$H = H_2 + H_3 + H_4 + \ldots + H_k + \ldots. \tag{12}$$

Here, H_k is the form of power k with respect to ξ_2 and η_2, and we have

$$\begin{aligned} H_2 &= \varphi_2, \quad H_3 = \varphi_3 - \varphi_2\psi_1, \\ H_4 &= \varphi_4 - \varphi_3\psi_1 + \varphi_2\psi_1^2 - \varphi_2\psi_2 + \chi\varphi_2^2. \end{aligned}$$

At the isoenergetic level $\Gamma = 0$, the equations for the perturbed motion have the Hamiltonian form (the Whittaker equation) with Hamiltonian (11). The coordinate ξ_1 plays the role of an independent variable. Note that the actual procedure (10) of changing the variables and obtaining expansions (11) and (12) leads to rather cumbersome calculations. They were performed using computer packages for analytical calculations.

STABILITY IN THE LINEAR APPROXIMATION. RESONANCE CURVES.

Let $\mathbf{X}(\xi_1)$ be a fundamental matrix of solutions to the linear system with the Hamiltonian H_2. Its elements $x_{ij}(\xi_1)$ obey the equations

$$\begin{aligned} \frac{dx_{1j}}{d\xi_1} &= \frac{\partial H_2(x_{1j}, x_{2j}, \xi_1)}{\partial x_{2j}}, \\ \frac{dx_{2j}}{d\xi_1} &= -\frac{\partial H_2(x_{1j}, x_{2j}, \xi_1)}{\partial x_{1j}}, \quad j = 1, 2, \end{aligned} \tag{13}$$

and initial conditions

$$x_{11}(0) = x_{22}(0) = 1, \quad x_{12}(0) = x_{21}(0) = 0. \tag{14}$$

The characteristic equation for matrix $\mathbf{X}(2\pi)$ has the form

$$\rho^2 - 2a\rho + 1 = 0, \quad a = \frac{1}{2}(x_{11}(2\pi) + x_{22}(2\pi)). \tag{15}$$

The calculations have demonstrated that, everywhere within the domain $Q_1Q_2Q_3$ of admissible values of the parameters θ_b and θ_c (Fig. 1), the quantity a meets inequality $|a| \leq 1$. Therefore, to obtain a rigorous solution [8] to the stability problem, it is not enough to consider only the linear approximation (13). Nonlinear analysis is needed.

Let $\rho = \exp(i\alpha)$ and $\bar{\rho} = \exp(-i\alpha)$ be the roots of Eq. (15) (multiplicators). Since $|a| \leq 1$, α is a real number and $\cos\alpha = a$. If $\rho^k = 1$, we have a resonance of the kth order. As a rule, for the stability problem, resonances up to the fourth order, inclusive, are the most important. The calculations demonstrated that there are

five curves γ_2, $\gamma_3^{(n)}$, and $\gamma_4^{(n)}$ ($n = 1, 2$) corresponding to these resonances. These curves are shown in Fig. 1. Each resonance curve begins at the vertex Q_1 of the triangle $Q_1Q_2Q_3$ and terminates at its vertex Q_2. The resonance of the second order takes place in the curve γ_2, where $a = -1$. Curves $\gamma_3^{(1)}$ and $\gamma_3^{(2)}$ correspond to the resonance of the third order; for them, $a = -\frac{1}{2}$. In curves $\gamma_4^{(1)}$ and $\gamma_4^{(2)}$, we have the resonance of the fourth order, and here, $a = 0$. The curve $\gamma_4^{(1)}$ consists of two portions Q_1Q_4 and Q_5Q_2 ($Q_4(0.936868, 0.063132)$, $Q_5(0.821425, 0.178575)$). The segment Q_1Q_4 of this curve differs only slightly from the segment Q_1Q_4 of the straight line Q_1Q_2. In Fig. 1, these segments are indistinguishable.

NONLINEAR ANALYSIS OF STABILITY

Algorithm of the study. We solve the nonlinear problem for stability of the equilibrium position $\xi_2 = \eta_2 = 0$ of the reduced system by reducing it to the stability problem for an immobile point of the mapping conserving the area. This mapping is generated by a set of equations corresponding to Hamiltonian (12).

Let q_0 and p_0 be the initial values of variables ξ_2 and η_2, whereas q_1 and p_1 are their values at $\xi_1 = 2\pi$. If q_0 and p_0 are sufficiently small, q_1 and p_1 are analytical functions of q_0 and p_0, and they specify the mapping T of the neighborhood of the equilibrium position onto itself. The stability problem for the equilibrium position $\xi_2 = \eta_2 = 0$ is equivalent to this problem for the immobile point $q_0 = p_0 = 0$ of mapping T.

To find the mapping, we perform the canonical univalent change of variables,

$$\begin{aligned} \xi_2 &= x_{11}(\xi_1)u + x_{12}(\xi_1)v, \\ \eta_2 &= x_{21}(\xi_1)u + x_{22}(\xi_1)v. \end{aligned} \tag{16}$$

The new Hamiltonian $G(u, v, \xi_1)$ does not contain the second-power terms G_2 and can be represented as a series

$$G = G_3 + G_4 + \ldots + G_k + \ldots, \tag{17}$$

where G_k is the form of H_k from expansion (12), where ξ_2 and η_2 are expressed in terms of u and v according to relationships (16). This change of variables reduces the problem of finding the mapping T to finding the mapping $q_0, p_0 \longrightarrow u_1, v_1$. Then, we have $q_0 = u_0$, $p_0 = v_0$, and

$$\begin{aligned} q_1 &= x_{11}(2\pi)u_1 + x_{12}(2\pi)v_1, \\ p_1 &= x_{21}(2\pi)u_1 + x_{22}(2\pi)v_1. \end{aligned} \tag{18}$$

We implicitly specify the mapping $q_0, p_0 \longrightarrow u_1, v_1$ by equalities

$$\begin{aligned} q_0 &= \frac{\partial S}{\partial p_0}, \quad v_1 = \frac{\partial S}{\partial u_1}, \\ S &= u_1p_0 + S_3(u_1, p_0) + S_4(u_1, p_0) \ \ldots \\ &\ldots + S_k(u_1, p_0) + \ldots, \end{aligned} \tag{19}$$

where S is the function

$$\begin{aligned} \Phi &= u_1p_0 + \Phi_3(u_1, p_0\xi_1) + \Phi_4(u_1, p_0, \xi_1) \ \ldots \\ &\ldots + \Phi_k(u_1, p_0, \xi_1) + \ldots, \end{aligned} \tag{20}$$

calculated at $\xi_1 = 2\pi$ and obeying the Hamilton–Jacobi equation

$$\frac{\partial \Phi}{\partial \xi_1} + G\left(u_1, \frac{\partial \Phi}{\partial u_1}, \xi_1\right) = 0 \tag{21}$$

and initial conditions $\Phi_k(u_1, p_0, 0) \equiv 0$ ($k = 3, 4, \ldots$).

Substituting expansions (17) and (20) into the left-hand side of Eq. (21) and equating to zero the coefficients at all powers of u_1 and p_0, we arrive at a system of ordinary differential equations for coefficients in the forms Φ_k, $k = 3, 4, \ldots$. At $\xi_1 = 0$, these coefficients should vanish. Since the forms G_3, G_4, … depend on $x_{ij}(\xi_1)$, this system should be considered simultaneously with system (13), (14). If mapping T is found with an accuracy up to terms on the order of m with respect to q_0, p_0, the number of equations in the resulting system is equal to $\frac{m(m+5)}{2} + 1$. Integrating this system from $\xi_1 = 0$ to $\xi_1 = 2\pi$, we find functions S_3, S_4, … and, from Eqs. (19), we determine the explicit form of q_0, $p_0 \longrightarrow u_1, v_1$:

$$\begin{aligned} u_1 &= q_0 - \frac{\partial S_3}{\partial p_0} + \frac{\partial^2 S_3}{\partial p_0 \partial q_0}\frac{\partial S_3}{\partial p_0} - \frac{\partial S_4}{\partial p_0} + O_4, \\ v_1 &= p_0 + \frac{\partial S_3}{\partial q_0} - \frac{\partial^2 S_3}{\partial q_0^2}\frac{\partial S_3}{\partial p_0} + \frac{\partial S_4}{\partial q_0} + O_4. \end{aligned} \tag{22}$$

Here, $S_k = S_k(q_0, p_0)$. The symbol O_4 denotes all terms of a power higher than three with respect to q_0, p_0. Substituting the found expressions (22) for u_1 and v_1 into the right-hand sides of Eqs. (18), we obtain the explicit form of mapping T.

After the mapping T has been found, it is possible [9] to write out the criteria of stability and instability of its fixed point $q_0 = p_0 = 0$. Thus, we arrive at conclusions concerning the orbital stability and instability of the periodic motion (7) in the Steklov's case. Putting aside

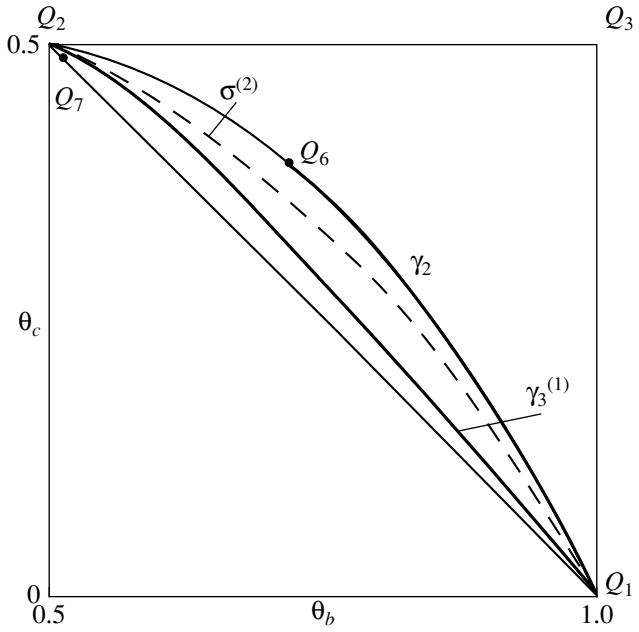


Fig. 2. Domains of orbital stability and instability.

the details of the calculations, we present below only the results of the examination of these criteria.

Results. Curve γ_2 for the second-order resonance is divided by point $Q_6(0.7320, 0.3776)$ into two segments Q_1Q_6 and Q_6Q_2. In the first segment, the fixed point $q_0 = p_0 = 0$ of mapping T is unstable, while at the second segment, it is stable.

In curve $\gamma_3^{(1)}$ for the third-order resonance, the fixed point of the mapping is unstable, while in curve $\gamma_3^{(2)}$ it is stable.

For both curves, $\gamma_4^{(1)}$ and $\gamma_4^{(2)}$ of the fourth-order resonance, the stability takes place.

For the values of parameters θ_b and θ_c lying within the triangle $Q_1Q_2Q_3$ but not belonging to the resonance curves γ_2, $\gamma_3^{(n)}$, and $\gamma_4^{(n)}$ ($n = 1, 2$), we verified the nondegeneracy condition for the normal form of mapping T in the terms of third power with respect to q_0, p_0 (the Arnold–Moser "torsion" condition [6]). It turned out that there exists a curve $\sigma(\theta_b, \theta_c) = 0$, at which this condition is violated. Curve $\sigma = 0$ consists of two portions, $\sigma^{(1)}$ and $\sigma^{(2)}$. Portion $\sigma^{(1)}$ is located near vertex Q_2. It starts at the point $Q_7(0.512643, 0.487357)$ of the segment Q_1Q_2 and at $\theta_c \longrightarrow 0.5$, it approaches arbitrarily close to the point Q_2 remaining between the segment Q_1Q_2 and the resonance curve $\gamma_4^{(1)}$. Portion $\sigma^{(2)}$ is located between resonance curves $\gamma_3^{(1)}$ and γ_2 and connects vertices Q_1 and Q_2 of triangle $Q_1Q_2Q_3$. In Fig. 2, the portion $\sigma^{(2)}$ is drawn by the dashed line, and the portion $\sigma^{(1)}$ is not shown in this figure.

In conclusion, we can say that Fig. 2 illustrates the aforementioned results of the nonlinear analysis of stability for Steklov's periodic motions (7). In the curve $\gamma_3^{(1)}$ for the third-order resonance and in the segment Q_1Q_6 of the γ_2 resonance curve marked out by the solid line, motion is orbitally unstable. In segment $\sigma^{(2)}$ (as well as in segment $\sigma^{(1)}$ not shown in Fig. 2) of curve $\sigma = 0$ and at the point Q_6 of curve γ_2, the problem of stability remains still unsolved. For the rest values of parameters θ_b and θ_c lying within the triangle $Q_1Q_2Q_3$, the periodic motion of the solid is orbitally stable in the Steklov's case.

ACKNOWLEDGMENTS

The work was supported by the Russian Foundation for Basic Research (project no. 02-01-00831) and by a Grant of the President of Russian Federation for the Support of Leading Scientific Schools (project no. NSh-1477.2003.1).

REFERENCES

1. V. A. Steklov, Tr. Otd. Fiz. Nauk Obshch. Lyubiteleĭ Estestvoznaniya **10** (1), 1 (1899).
2. E. Yu. Kucher, Mekh. Tverd. Tela, No. 33, 33 (2003).
3. P. A. Kuz'min, Prikl. Mat. Mekh. **16**, 243 (1952).
4. A. I. Dokshevich, *Solving Euler–Poincaré Equations in the Finite Form* (Naukova Dumka, Kiev, 1992).
5. A. P. Markeev, Prikl. Mat. Mekh. **66**, 929 (2002).
6. V. I. Arnol'd, V. V. Kozlov, and A. I. Neĭshtadt, *Mathematical Aspects of Classical and Celestial Mechanics* (URSS Editorial, Moscow, 2002).
7. A. P. Markeev, *Libration Points in Celestial Mechanics and Space Dynamics* (Nauka, Moscow, 1978).
8. I. G. Malkin, *Theory of Stability of Motion* (Nauka, Moscow, 1966).
9. A. P. Markeev, Izv. Ross. Akad. Nauk: Mekh. Tverd. Tela, No. 2, 37 (1996).

Translated by K. Kugel

Doklady Physics, Vol. 49, No. 10, 2004, pp. 598–602. Translated from Doklady Akademii Nauk, Vol. 398, No. 5, 2004, pp. 625–629.

MECHANICS

A Model of an Asymmetric Weakened Zone in Problems of Interaction Between Cracklike Defects in an Elastic Medium

I. V. Simonov

Presented by Academician N.F. Morozov May 17, 2004

Received May 15, 2004

A weakened zone of a deformed solid medium can be defined as a cut in its volume. In this case, the entire surface of cut edges is subjected to the action of distributed adhesive forces, whereas displacements exhibit discontinuities [1]. This object can be treated as a forerunner of a crack. With increasing load on a solid when the bonds begin to break and traction-free zones appear on the edges of a cut, the weakened zone transforms into a crack. The fundamental difference between a weakened zone and a crack is that the opening of the weakened zone does not occur immediately, but on attaining a certain load level sufficient to overcome the initial adhesive forces. These forces can be of different physical natures, e.g., of an atomic-molecular or dislocation character. In addition, these forces can arise due to the presence of porosity, fibers, plastic hinges, etc. The scale of linear sizes of a weakened zone ranges from nanometers and micrometers (new materials) to kilometers (welded cracks in glaciers and faults in the earth's crust).

In this study, an attempt is made to construct a model of an asymmetric weakened zone that would be suitable for studying the interaction of cracklike defects, with the physical nature of the adhesive forces being not concretely specified. The idea is to explicitly set the opening of this zone in the form of a two-parametric family of functions as a special-type dislocation. In addition, certain physical constraints on the behavior of a solution at the end zones are *a priori* taken into account. One of the goals of this investigation is to test the validity of the model in solving the problem on interaction between a crack and a weakened zone of a small length at the interface of two elastic media. This is one of stages of a possible scenario of discontinuous crack motion, when the crack propagation is retarded by a firm microstructure element (a barrier) and, beyond the barrier, a weakened zone is formed on the extension of a crack. From the solution to this problem, the ultimate load responsible for transformation of this weakened zone into a crack and the critical load braking up the barrier are found. A typical—and the most interesting from the physical standpoint—situation is when a different-scale interaction occurs: the crack size considerably exceeds the length of the weakened zone, and the barrier width is much less than the length of the weakened zone. A method for defining the adhesive forces in measurements of the opening at the sites of an incomplete fracture is also described.

When constructing the model, we base ourselves on the investigation of the behavior of the weakened zone in the case of symmetric loading [1] when the adhesive forces are set in the form of polynomials of a different order in coordinates, and the coefficients are found from additional conditions. In contrast to [1], we consider here an asymmetric opening of a cut $w(X)$, $|X| \le 1$, which is located on the X axis, in the form

$$w(X) = w_0 T^{3/2}, \quad T = (1 - X^2)G(X, \eta), \quad -1 < X < 1, \tag{1}$$

where w_0 and η are, respectively, the amplitude and displacement of the maximum to the left from the center $X = 0$ of the zone provided that $\eta > 0$. For the unknown function $G(X, \eta)$, it is assumed that the conditions of evenness and normalization are fulfilled, the function $T(X, \eta)$ attains its maximum at $X = -\eta$, this function is bounded at the peak value $X = 1$ for $\eta \longrightarrow 1 + 0$, and an inequality denoting that, at $\eta > 0$, the larger opening intensity coefficient corresponds to the edge $X = -1$:

$$G(X, \eta) = G(-X, -\eta), \quad T(-\eta, \eta) = 1, \quad G(X, 0) = 1,$$
$$\frac{\partial T}{\partial X}(-\eta, \eta) = 0, \quad 0 < G(1, \eta) < G(-1, \eta), \quad \eta > 0. \tag{2}$$

The function $G(X, \eta)$ is sought in the form of a rational function in variables X, η, which obeys conditions (2). This leads to a finite recurrence system of equations

Institute for Problems in Mechanics, Russian Academy of Sciences, pr. Vernadskogo 101, Moscow, 119526 Russia

1028-3358/04/4910-0598$26.00 © 2004 MAIK "Nauka/Interperiodica"

with respect to the coefficients. Unexpectedly, the solution has a simple form,

$$T(X, \eta) = \frac{1 - X^2}{(1 + \eta^2)(1 + \alpha X)}, \quad \alpha = \frac{2\eta}{1 + \eta^2}. \quad (3)$$

For $\eta = 0$, opening (1), (3) corresponds to the solution for the symmetric case [1] if the adhesive forces are specified as a quadratic function of a coordinate. In addition, the asymptotic behavior of the opening in the neighborhood of the edges of the weakened zone (the so-called beak) qualitatively coincides with the approximation of the experimental values of this quantity at the crack end zone [2]. The parameter η is found from the condition that the adhesive forces at the point of the maximum opening $w(x)$ are minimal. The amplitude w_0 remains a free parameter. This ambiguity is a result of the general character of the approach, i.e., its arbitrariness in the description of the relation between a force and an opening. However, the ultimate value w_* is assumed to be the known parameter [1]. For $w_0 > w_*$, traction-free regions appear at the edges. Thus, the criterion $w_0 = w_*$ characterizes the onset of transforming the weakened zone into an imperfect crack, i.e., into a qualitatively different state. This condition is similar to the Leonov–Panasyuk criterion known in theory of cracks: for $w_0 = w_*$, the adhesive forces disappear at the point $X = -\eta$ and this equality is used to find the ultimate load.

It seems likely that the possibility of a mechanism responsible for the formation of a zone of microdefects in front of a mobile crack and their subsequent merging, which causes nonuniform crack motion, was first proposed in [3]. Later, this hypothesis was confirmed by experimental observations (see [4]). In [5], a number of discrete mathematical fracture models were proposed, and it was shown that such a mechanism can manifest itself only in the presence of a microstructure. Thus, one possible explanation of the phenomenon of discontinuous propagation of cracks is the following. Periodically, the crack front meets a structural element—a barrier of enhanced firmness—and stops if there is not enough energy to overcome the barrier. Beyond the barrier, in a zone of a weak firmness, microdefects (dislocations, microcracks, pores, etc.) are formed, which requires time. They locally change the stress field in such a manner that the effect produced on the barrier is enhanced. As a result, the barrier is destroyed, and the front moves for the distance of the lattice period up to the following barrier. In the literature, there are many variants describing the interaction of a crack and microdefects. However, it is interesting to analyze the situation when the weakened zone arises in front of the crack.

We consider the plane problem of the contact between two linearly elastic isotropic half-planes $y > 0$ (medium 1) and $y < 0$ (medium 2). On the straight line $y = 0$, which connects the media, a half-infinite imperfect crack ($x < 0$) and a weakened zone ($0 < a < x < b$) are located. They are opened by normal forces Σ applied to the crack edges at a unit distance from the crack tip (Fig. 1). The adhesive forces impede this opening. In a small zone $-l < x < 0$ of the fracture process, they are distributed according to a quadratic law,

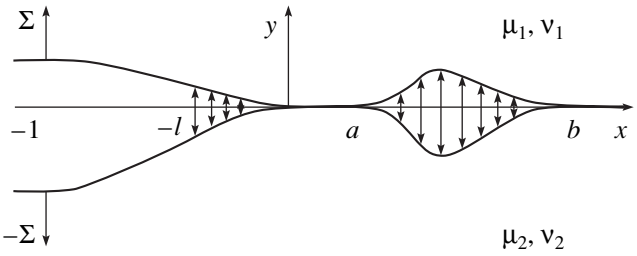


Fig. 1.

$$\sigma = \sigma_0(1 - R^2), \quad R = \frac{x}{l}, \quad (4)$$

and, in the region $a < x < b$, the displacement jump of form (1), (3) is set, where the coordinate X should be replaced by

$$X = \frac{x}{L} - D, \quad D = \frac{b + a}{b - a} > 1, \quad L = \frac{b - a}{2}. \quad (5)$$

The weakened zone, when it is located near the crack, is immersed in an essentially asymmetric stress field. This was taken into account above in constructing the weakened-zone model. In the segments $0 < x < a$ and $x > b$, the stresses and displacements are continuous. Thus, a problem appears for an elastic compound plane with the boundary conditions for the stresses at the semiaxis $x < 0$ and for the jumps of the normal displacements at $x > 0$.

Expressing the stress fields and derivatives with respect to x of the displacement vector in terms of the complex potential satisfying the continuation conditions across the interface [1], we arrive at the coupled generalized vector Riemann–Hilbert problem [6]. This problem can be essentially simplified and reduced to the Keldysh–Sedov problem for a single complex function [6], provided that terms inessential for the analysis are ignored. The problems are equivalent if the Dunders parameter β ($|\beta| \le 0.5$ [1]), which defines, in essence, the difference in the elastic properties of materials (in particular, identical materials), is equal to zero. For small values of β, the simplified problem provides a major part of the exact solution with an error of $O(\beta^2)$. For finite values of β, the solution to this problem qualitatively correctly and quantitatively satisfactorily describes the normal contact forces that are of interest for us in the case of a "poor" determination of shear stresses. The available experience in solving various contact problems confirms this statement.

The formulation of the problem becomes similar to that of the problem for a uniform plane, and the diver-

sity of elastic properties is determined only by the parameter q (for its definition, see below). The solution to this simplified problem is given by the integrals of the Cauchy type. We now write out the final expressions for the normal adhesive forces in front of the crack after normalizing them to the value of Σ and with allowance for (4) and (5):

$$\sigma(x) = \sigma_1(x) + \sigma_2(x), \quad x > 0,$$

$$\sigma(0) = \frac{5}{8\sqrt{l}}(1 + W_0 I_0),$$

$$\sigma_1(x) = \frac{3W_0}{\pi L}\int_{-1}^{1}\sqrt{\frac{\xi + D}{X + D}}\frac{g(\xi)d\xi}{\xi - X},$$

$$W_0 = \frac{w_0}{q\Sigma}, \quad g(\xi) = \frac{(1+\eta\xi)(\eta+\xi)\sqrt{1-\xi^2}}{(1+\eta^2+2\eta\xi)^{5/2}},$$

$$\sigma_2(x) = \frac{\Theta(x,l)}{\pi\sqrt{x}} = \frac{1}{\pi\sqrt{x}}\left\{\frac{1}{1+x} + O\left(\frac{1}{R}\right)\right\}, \quad R \gg 1,$$

$$\Theta(x,l) = S + W_0 I_0 (S-1) - \frac{x}{1+x}, \qquad (6)$$

$$I_0 = \int_{-1}^{1}\left(\frac{T(\xi)}{\xi + D}\right)^{3/2}\frac{d\xi}{2\sqrt{L}},$$

$$S = \frac{5}{4}\left\{\sqrt{R}(1-R^2)\arctan\frac{1}{\sqrt{R}} - \frac{R}{3} + R^2\right\}$$

$$\approx 1 - \frac{4}{21R} + O\left(\frac{1}{R^2}\right), \quad R \gg 1,$$

$$q = q_1 + q_2, \quad 4\mu_j q_j = 1 + \kappa_j,$$

$$\kappa_j = 3 - 4\nu_j \text{ (plane deformation)},$$

$$\kappa_j = (3-\nu_j)(1+\nu_j)^{-1}$$

(two-dimensional stress state),

where $\sigma_1(x)$ is the weakened-zone contribution, $\sigma_2(x)$ is the nominal stresses in the absence of the weakened zone, and the quantity $\sigma(0)$ is calculated from the boundedness condition after eliminating the root singularity; μ_j and ν_j are the shear moduli and Poisson's ratios ($j = 1, 2$), respectively. The integrals in (6) exist as $x \longrightarrow a + 0$ and $b - 0$, and the function $\sigma(x)$ is bounded.

Equating the derivative $\sigma'(x)$ to zero at the point $x = x_m = L(D - \eta)$, where the maximum of $w(x)$ is attained, we arrive at the equation for the determination of the displacement parameter η:

$$2L^{3/2}\left[2\Theta'(x_m, l) - \frac{\Theta(x_m, l)}{x_m}\right] + 3W_0 J(\eta) = 0, \qquad (7)$$

$$J(\eta) = \int_{-1}^{1}\frac{\sqrt{D+\xi}(\alpha^2\xi^4 + 4\alpha\xi^3 + (8-2\alpha^2)\xi^2 + 4\alpha\xi + 5\alpha^2 - 4)}{(1+\eta^2)^{3/2}(1+\alpha\xi)^{7/2}(1-\xi^2)^{1/2}(\xi+\eta)}d\xi.$$

The uniqueness of the solution is provided by the continuous and strictly monotonic dependence of the special integral J on η and by the fact that it takes all real values within the interval $|\eta| < 1$. In the ultimate case, we find the relation of $W_0 = W_*$ to other parameters and, provided that $R_m \gg 1$, Eq. (7) is simplified, becoming independent of l:

$$W_* = \frac{S(R_m) - \dfrac{x_m}{1+x_m}}{(1 - S(R_m))I_0 + 3I_1 L^{-1/2}},$$

$$\frac{J}{I_1} \approx \frac{2}{D-\eta}\left\{\frac{1+3x_m}{1+x_m}\right\}. \qquad (8)$$

Length of the crack mouth l can be found from the Leonov–Panasyuk condition $w(l) = w_c$, where parameter w_c defines the value of the mouth opening to which the adhesive forces vanish:

$$\sqrt{\frac{l}{L}} = \frac{\dfrac{\pi w_c}{q\Sigma}}{\dfrac{25}{18}\sqrt{L} - W_0 I_2} \approx \frac{\dfrac{\pi w_c}{w_*}}{\dfrac{25}{6}(1+x_m)I_1 - I_2},$$

$$w_0 = w_*, \quad R_m \gg 1,$$

$$I_1 = \int_{-1}^{1}\frac{\sqrt{T(\xi)(\xi+D)}}{(1+\eta^2+2\eta\xi)^2}(1+\eta\xi)d\xi, \qquad (9)$$

$$I_2 = \int_{-1}^{1}\frac{T^{3/2}(\xi)d\xi}{\sqrt{\xi+D}\left(\xi + D + \dfrac{l}{L}\right)}.$$

The solution to the integroalgebraic equation (9) [along with (7) or (8)], can be efficiently sought by the method of successive approximations.

To determine the onset of the barrier fracture, it is appropriate to use the Novozhilov–Neuber criterion, according to which fracture takes place when the averaged tension stress at the barrier attains the critical value σ_c [5]. The critical force can be found after explicitly eliminating the singularity for stress (6) and averaging over the segment $0 < x < a$:

$$\Sigma_c = \frac{\pi\sqrt{l}\sigma_c - \frac{w_0}{q}[I_3 + I_0 I_4]}{I_4 - 2a^{-1}\sqrt{l}(\sqrt{a} - \arctan\sqrt{a})},$$

$$I_4 = \frac{1}{h}\int_0^h \frac{S(R)dR}{\sqrt{R}}$$

$$\approx \frac{5}{4}\arctan\frac{1}{\sqrt{h}}, \quad h \ll 1,$$

$$I_3 = \frac{3\sqrt{l}}{2a}\int_0^\epsilon \sqrt{r}J_0(r)dr \approx \sqrt{\frac{al}{L^3}}J_0(0), \quad \epsilon \ll 1, \tag{10}$$

$$J_0 = \int_{-1}^{1}\left[\frac{T(\xi)}{D+\xi}\right]^{3/2}\frac{D+\xi-r/3}{(D+\xi-r)^2}d\xi, \quad h = \frac{a}{l},$$

$$r = \frac{x}{L}, \quad \epsilon = \frac{a}{L}.$$

The quantity Σ_c drops with rising amplitude of opening w_0 and attains the minimum value Σ_c^m for $w_0 = w_*$.

The ultimate load under which the weakened zone transforms into a crack is found under the condition of vanishing the adhesive forces at the point of their minimum $x = x_m$:

$$\Sigma_* = \frac{w_*}{qW_*} \approx \frac{3w_*(1+x_m)I_1}{q\sqrt{L}}, \quad R_m \gg 1. \tag{11}$$

The evolution of the system of two defects for continuously increasing the load can be different depending on the system parameters. For example, if $\Sigma_c^m < \Sigma_*$, then criterion (10) is realized first: the crack passes by a barrier and, at the same time, captures the weakened zone. If $\Sigma_c^m > \Sigma_*$, first, the weakened zone transforms into a crack. In this case, a more complex problem appears with additional unknowns, namely, with the lengths of the adhesive-force actions near the tips of the formed crack. As is noted in [1], the above transformation can be unstable and be accompanied by rapid processes of vanishing of the adhesion forces and development of one of the cracks. Thus, the description of this transformation needs additional hypotheses.

The calculated values of the parameter η, the ratio $\frac{l}{L}$, the peak values of stresses $\sigma(0)$, and the force coefficient $f = \frac{P}{\Sigma} = \langle\sigma\rangle a$ (where P is the load breaking the barrier) for two values of $a = 0.01$ and 0.1 are presented in the table as a function of half-length L of the weakened zone.

$a = 0.01$							
L	0.005	0.010	0.020	0.050	0.100	0.200	0.500
η	0.95	0.118	0.153	0.177	0.190	0.217	0.288
$\frac{l}{L}$	0.079	0.130	0.197	0.286	0.323	0.323	0.275
$\sigma(0)$	31.74	17.93	10.63	5.914	3.965	2.840	2.071
f	0.060	0.056	0.051	0.041	0.033	0.026	0.019
$a = 0.10$							
L	0.005	0.010	0.020	0.050	0.100	0.200	0.500
η	0.0131	0.0251	0.0466	0.0966	0.152	0.217	0.313
$\frac{l}{L}$	0.0090	0.0171	0.0312	0.0625	0.0959	0.132	0.158
$\sigma(0)$	93.198	47.854	25.115	11.322	6.578	4.078	2.468
f	0.228	0.206	0.195	0.185	0.176	0.163	0.139

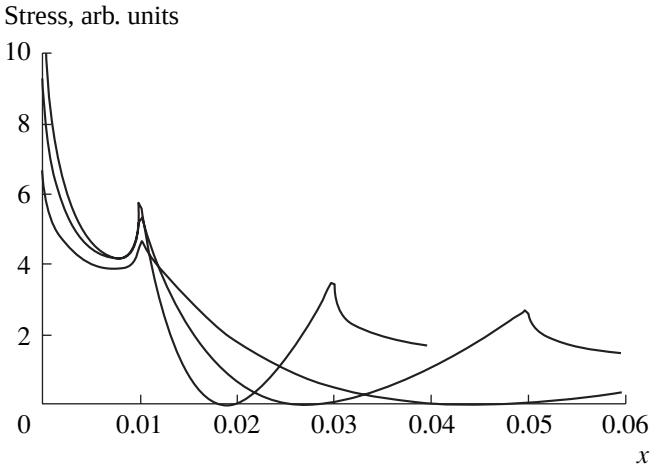


Fig. 2.

It was assumed in this case that $w_0 = w_* = w_c$, i.e., the ultimate situation was considered. For small L, the parameter $\eta < 0.31$ sharply increases with L, while it rises moderately for $L > 0.1$. With increasing a, the parameter η noticeably drops. The ratio $\frac{l}{L}$ is important as a measure of the accuracy, since approximate equation (8) was used. According to the estimates, the error must not exceed 10%. (It is inappropriate to require a higher accuracy, since the approach used, in itself, is approximate.) The length is $l < 0.14$ and strongly changes: the effect of the weakened zone on the adhesive region $-l < x < 0$ is significant. The value of normalized stress at the crack tip $\sigma(0)$ decreases with increasing the length of the weakened zone: $\sigma(0) \sim \frac{1}{\sqrt{L}}$ and $\frac{a}{L} < 0.5$. Upon receding of the constant-length weakened zone from the crack tip, the coefficient of force increases, and averaged stress $\langle\sigma\rangle$ decreases. When a = const and the length of the weakened zone increases, the values of f and $\langle\sigma\rangle$ decrease logarithmically.

From the solution, it is possible to evaluate the deformation law for bonds $\sigma = \sigma(w)$. This dependence turned out to be close to linear one, with a small deviation in the neighborhood of the point $w = 0$. We recall that all these conclusions are related to the ultimate situation $w_0 = w_*$. Therefore, in order to retain the situation with changing a or L, both the applied force Σ ($\sim L^{-1/2}$), and the quantities $\sigma(0)$, l, and f, also must vary nonlinearly. Figure 2 shows the distribution of the stresses in front of the crack for $a = 0.01$ and $L = 0.01$, 0.02, and 0.05. In the left tip of the weakened zone, the maximum of these stresses depends slightly on the length of the weakened zone and is determined by the barrier width. In the case of a close weakened zone, the stress pattern near the barrier $0 < x < a$ is strongly distorted, and the distortion becomes local only at a certain distance ($a \gg L$).

By virtue of the σ–τ dualism well-known in elasticity theory, a weakened-zone model of shear type and the solution to the corresponding problem associated with loading a crack–weakened-zone system by a point shear force, as well as by a oblique force, are automatically derived from the above results.

The theoretical approach being developed provides the possibility of experimentally determining the constitutive parameters and functions. For example, using laser techniques, it is possible to determine with a high degree of accuracy the opening of a weakened zone in plates [2]. Comparison of the experimental $w_e(x)$ and predicted (1) forms of the opening will indicate the degree of deviation between the ideal and actual patterns. In addition, in solving a problem similar to that described above and using the measured opening $w_e(x)$, we can find the true bond law, which is incapable of direct measurements. Therein lies the indirect method of determining the deformation-bond law for incomplete fracture.

ACKNOWLEDGMENTS

This work was supported by the OÉMMPU Program of the Russian Academy of Sciences, no. 13.

REFERENCES

1. V. Simonov and B. L. Karihaloo, Int. J. Solids Struct. **37** (46/47), 7055 (2000).
2. A. S. Vavakin, Yu. I. Kozyrev, and R. L. Salganik, Izv. Akad. Nauk SSSR: Mekh. Tverd. Tela, No. 2, 111 (1969).
3. K. B. Broberg, in *Proc. IUTAM Symp. on High Velocity Deformation of Solids* (Springer, Berlin, 1979), p. 182.
4. V. N. Nikolaevskiĭ, in *Mechanics. Impact, Explosion, and Fracture* (Mir, Moscow, 1981), No. 26, p. 166.
5. N. F. Morozov, *Mathematical Problems in Mechanics of Cracks* (Nauka, Moscow, 1984).
6. N. I. Muskhelishvili, *Certain Basic Problems in Mathematical Theory of Elasticity* (Nauka, Moscow, 1966).

Translated by Yu. Vishnyakov

Doklady Physics, Vol. 49, No. 10, 2004, pp. 603–607. Translated from Doklady Akademii Nauk, Vol. 398, No. 5, 2004, pp. 630–634.

MECHANICS

Invariant Integrals in Problems of a Crack at the Locus of Inhomogeneity and in Contact Problems

A. M. Khludnev

Presented by Academician V.N. Monakhov April 27, 2004

Received May 7, 2004

The problem of equilibrium of an anisotropic elastic body containing a crack at the locus of inhomogeneity is considered in the present study. It is found that, in this problem, there exist invariant integrals (that is, integrals that are independent of the integration surface). The existence of invariant integrals is also established in the problem of a contact of an elastic body that interacts, over some part of its surface, with a rigid stamp, in which case nonlinear boundary conditions of mutual nonpenetration are imposed at the contact boundaries.

The existence of invariant integrals in the linear theory of cracks—these are usually referred to as Cherepanov–Rice integrals—has been discussed in a number of studies (see, for example, [1–3]). The analysis there was confined to linear problems, where linear boundary conditions were imposed at the edges of a crack. Here, we consider nonlinear problems in crack theory (for a survey of such problems, the interested reader may address the monograph quoted in [4]). Boundary conditions at the edges of a crack in the form of a set of equalities and inequalities are peculiar to nonlinear problems. Previously, invariant integrals for smooth tensors of elasticity moduli were constructed in nonlinear problems of crack theory [4–6]. In the present study, invariant integrals are constructed for an elastic body containing a crack at the locus of inhomogeneity. In this case, the tensor of elasticity moduli is not smooth over the domain being considered. The invariant integrals are constructed both for the two- and the three-dimensional case. The fictitious-domain method, which was recently developed for problems involving the Signorini boundary conditions [7, 8], is used to derive invariant integrals in contact problems. Here, the problem of equilibrium of a body containing a crack is included in the family of problems dependent on a parameter, a contact problem being associated with the limiting value of this parameter. As a matter of fact, invariant integrals are obtained simultaneously in the problems being considered. The fundamentals of the fictitious-domain method as applied to linear boundary value problems can be found in [9–11]. Concurrently, use is being made here of a formula for the derivative of the energy functional with respect to a perturbation parameter in elasticity-theory problems where nonlinear boundary conditions are imposed at the edges of cracks in solids. The technique of differentiation of energy functionals in the nonlinear problems of crack theory is given in [4–6, 12]. Applications of problems of crack theory in mechanics of a deformed solid can be found in [1, 2].

FORMULATION OF THE PROBLEMS.

Let $\Omega_1 \subset R^3$ be a bounded simply connected domain having Lipschitzian boundary Γ_1. Let $\Gamma_c \subset \Gamma_1$ be a contact boundary—that is, part of the boundary where the Signorini boundary conditions $\Gamma_0 = \Gamma_1 \backslash \Gamma_c$ and $\operatorname{meas}\Gamma_0 > 0$ are satisfied. For the sake of simplicity, we assume that Γ_c, which is a two-dimensional surface in R^3, can be represented as a graph of the function $x_3 = \phi(x_1, x_2)$, $(x_1, x_2) \in \overline{D}$, where ϕ is a rather smooth function. Here, $D \subset R^2$ is a bounded, simply connected domain with the boundary γ_0 of class $C^{0,1}$; being a curve in R^3, γ_0 can be represented in the form $\gamma_0 = \{(r, \varphi, 0) | r = g(\varphi), \varphi \in [0, 2\pi], g(0) = g(2\pi), g > 0, g \in C^{0,1}\}$. Moreover, there exists $\delta_0 > 0$ such that we have

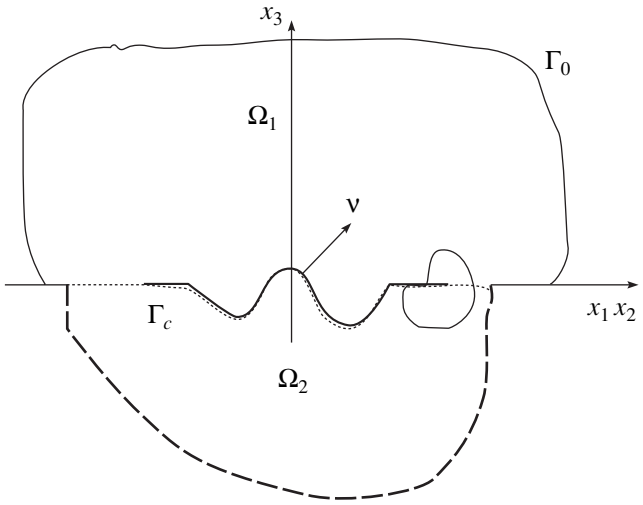
$$\gamma_1 \subset \Gamma_1,$$
$$\gamma_1 = \{(r, \varphi, 0) | g(\varphi) - \delta_0 < r < g(\varphi) + \delta_0\},$$

where (r, φ, ξ) are cylindrical coordinates in R^3. The above condition implies that, near the edge γ_0 of the contact boundary Γ_c, there exists a flat area belonging to the boundary Γ_1.

The formulation of a contact problem is as follows [13]. In the domain Ω_1, it is required to find the

Lavrent'ev Institute of Hydrodynamics,
Siberian Division, Russian Academy of Sciences,
pr. Lavrent'eva 15, Novosibirsk, 630090 Russia
e-mail: khlud@hydro.nsc.ru

1028-3358/04/4910-0603$26.00 © 2004 MAIK "Nauka/Interperiodica"



Vertical cross section of the domain Ω_c.

functions $u^0 = (u_1^0, u_2^0, u_3^0)$ and $\sigma = \{\sigma_{ij}\}$ $(i, j = 1, 2, 3)$ such that

$$-\mathrm{div}\sigma = f \quad \text{in} \quad \Omega_1, \tag{1}$$

$$\sigma = C^1\varepsilon(u^0) \quad \text{in} \quad \Omega_1, \tag{2}$$

$$u^0 = 0 \quad \text{on} \quad \Gamma_0, \tag{3}$$

$$\begin{gathered} u^0\nu \geq 0, \quad \sigma_\nu \leq 0, \quad \sigma_\tau = 0, \\ u^0\nu \cdot \sigma_\nu = 0 \quad \text{on} \quad \Gamma_c. \end{gathered} \tag{4}$$

Here and below, $\varepsilon_{ij}(v) = \frac{1}{2}(v_{i,j} + v_{j,i})$ are the components of the strain tensor, $v_{i,j} = \dfrac{\partial v_i}{\partial x_j}$ with $x = (x_1, x_2, x_3) \in \Omega_1$, $f = (f_1, f_2, f_3) \in C^1_{\mathrm{loc}}(R^2)$ is a known function; and $C^1 = \{c^1_{ijkl}\}$ is the tensor of elasticity moduli $(i, j, k, l = 1, 2, 3)$,

$$c^1_{ijkl} = c^1_{klij} = c^1_{jikl}, \quad c^1_{ijkl} = \mathrm{const},$$

$$c^1_{ijkl}\xi_{kl}\xi_{ij} \geq c|\xi|^2, \quad c > 0, \quad \forall \xi = \{\xi_{ij}\},$$

$$\sigma_\nu = \sigma_{ij}\nu_j\nu_i, \quad \sigma_\tau = \sigma\nu - \sigma_\nu \cdot \nu,$$

$$\sigma\nu = \{\sigma_{ij}\nu_j\}_{i=1}^3.$$

We note that Eq. (1) is the equation of equilibrium; Eq. (2) represents Hooke's law; boundary condition (3) corresponds to the fastening of an elastic body on Γ_0; and boundary conditions (4), referred to as the Signorini boundary conditions, describe the contact of an elastic body with a nondeformable surface at zero friction.

All of the quantities carrying two subscripts are assumed to be symmetric with respect to the interchange of these subscripts—for example, $\sigma_{ij} = \sigma_{ji}$; in addition, summation over dummy indices is implied.

It is well known that the problem specified by (1)–(4) admits a variational formulation. To be more specific, we consider the space of Sobolev functions, $H^1_{\Gamma_0}(\Omega_1) = \{v = (v_1, v_2, v_3) \in H^1(\Omega_1) \mid v = 0 \text{ on } \Gamma_0\}$, and the set of admissible translations $K = \{v \in H^1_{\Gamma_0}(\Omega_1) \mid v\nu \geq 0$ almost everywhere on $\Gamma_c\}$. The problem specified by (1)–(4) can then be represented in the form of the variational inequality

$$u^0 \in K, \quad \int_{\Omega_1} \sigma(u^0)\varepsilon(v - u^0) \geq \int_{\Omega_1} f(v - u^0) \quad \forall v \in K. \tag{5}$$

Here and below, we use the notation $\sigma(v) = C^1\varepsilon(v)$.

Along with the problem formulated in terms of (1)–(4), we will consider the problem of equilibrium of an elastic body containing a crack at the locus of inhomogeneity. For this, a bounded domain Ω_2 that has a Lipschitzian boundary Γ_2 is added to the domain Ω_1 in a way that is indicated in the figure, whereupon the boundary-value problem involving nonlinear boundary conditions on Γ_c is solved in the domain $\Omega_c = \Omega_1 \cup \Omega_2 \cup (\Sigma \backslash \Gamma_c)$ with $\Sigma = \Sigma_0 \backslash \partial\Sigma_0$, $\Sigma_0 = \Gamma_1 \cap \Gamma_2$. As a matter of fact, it is assumed here that, in R^3, there exists a domain such that it is divided by a regular surface Σ_0 into two subdomains Ω_1 and Ω_2 and that $\Gamma_c \subset \Sigma_0$. The resulting problem describes the equilibrium of an elastic body occupying the domain Ω_c and containing the crack Γ_c, the boundary conditions at its edges $\Gamma_c^\pm$ corresponding to the absence of penetration. Moreover, we will consider a family of boundary-value problems dependent on a parameter λ such that the problem of equilibrium of a body containing a crack and the problem specified by (1)–(4) correspond to positive values of λ and $\lambda = 0$, respectively. The existence of invariant integrals will be established simultaneously for the entire family of problems associated with positive values of λ. The existence of invariant integrals for the contact problem specified by (1)–(4) will be proven by going over to the limit $\lambda \longrightarrow 0$. From the standpoint of the contact problem specified by (1)–(4), the auxiliary domain Ω_2 introduced above can be called a fictitious domain. The geometry of the domains Ω_1 and Ω_2 is assumed to be such that the cut Γ_c does not reach the external boundary Γ; that is, $\Gamma_c \cap \Gamma = \emptyset$, where $\Gamma = \partial\Omega_c \backslash (\Gamma_c^+ \cup \Gamma_c^-)$.

Thus, we introduce the tensor $B^\lambda = \{ b^\lambda_{ijkl} \}$ ($\lambda > 0$; $i, j, k, l = 1, 2, 3$) as

$$b^\lambda_{ijkl} = \begin{cases} c^1_{ijkl} & \text{in} \quad \Omega_1, \\ \lambda^{-1} c^2_{ijkl} & \text{in} \quad \Omega_2. \end{cases}$$

Here, the tensor $C^2 = \{ c^2_{ijkl} \}$ possesses the same properties as the tensor C^1. In the domain Ω_c, which contains the crack (cut) Γ_c, we will solve the following problem. It is required to find functions $u^\lambda = (u^\lambda_1, u^\lambda_2, u^\lambda_3)$ and $\sigma^\lambda = \{ \sigma^\lambda_{ij} \}$ ($i, j = 1, 2, 3$) such that

$$-\mathrm{div}\sigma^\lambda = f \quad \text{in} \quad \Omega_c, \tag{6}$$

$$\sigma^\lambda = B^\lambda \varepsilon(u^\lambda) \quad \text{in} \quad \Omega_c, \tag{7}$$

$$u^\lambda = 0 \quad \text{on} \quad \Gamma, \tag{8}$$

$$\begin{gathered} [u^\lambda]\nu \geq 0, \quad [\sigma^\lambda_\nu] = 0, \quad \sigma^\lambda_\nu \leq 0, \quad \sigma^\lambda_\tau = 0, \\ [u^\lambda]\nu \cdot \sigma^\lambda_\nu = 0 \quad \text{on} \quad \Gamma_c. \end{gathered} \tag{9}$$

Here, $[v] = v^+ - v^-$ is the discontinuity of the function v on Γ_c, where the plus (minus) sign corresponds to the positive (negative) direction of the normal ν; $\sigma^\lambda_\nu = \sigma^\lambda_{ij}\nu_j\nu_i$; and $\sigma^\lambda_\tau = \sigma^\lambda\nu - \sigma^\lambda_\nu \cdot \nu$. The equality $\sigma^\lambda_\tau = 0$ on Γ_c implies that $\sigma^\lambda_\tau = 0$ on $\Gamma^\pm_c$.

The problem specified by (6)–(9) has only one solution at each positive value of λ. Namely, we consider the space of functions $H^1_\Gamma(\Omega_c) = \{ v = (v_1, v_2, v_3) \in H^1(\Omega_c) \mid v = 0 \text{ on } \Gamma \}$ and the set of admissible translations $K_c = \{ v \in H^1_\Gamma(\Omega_c) \mid [v]\nu \geq 0$ almost everywhere on $\Gamma_c \}$. The problem specified by (6)–(9) is then equivalent to that of minimizing the functional

$$\Pi_\lambda(\Omega_c; v) = \frac{1}{2}\int_{\Omega_c} \sigma^\lambda(v)\varepsilon(v) - \int_{\Omega_c} fv$$

on the set K_c and can be formulated in terms of a variational inequality; that is,

$$u^\lambda \in K_c,$$

$$\int_{\Omega_c} \sigma^\lambda(u^\lambda)\varepsilon(v - u^\lambda) \geq \int_{\Omega_c} f(v - u^\lambda) \quad \forall v \in K_c.$$

Here, $\sigma^\lambda(v)$ is determined from an equation of the form in (7); that is, $\sigma^\lambda(v) = B^\lambda\varepsilon(v)$.

We will consider a family of perturbed problems dependent on the parameter δ that are defined in the perturbed domain Ω^δ_c. At each fixed value of λ and a small value of δ, we will find a solution $u^{\lambda\delta}$ to the respective perturbed problem and the derivative of the energy functional $\Pi_\lambda(\Omega^\delta_c; u^{\lambda\delta})$ with respect to the parameter δ at $\delta = 0$. At an appropriate choice of perturbations, the resulting formula for the above derivative will provide invariant integrals in the problem specified by (6)–(9). After that, we will go over to the limit $\lambda \longrightarrow 0$ in the formula for this derivative.

Let us consider a perturbation of the domain Ω_c in the form $y = \Psi_\delta(x)$, $x \in \Omega_c$ and $y \in \Omega^\delta_c$, where $\Psi_\delta(x) = x + \delta V(x)$ and $V(x) = (V_1(x), V_2(x), 0)$, $V(x) \in W^{1,\infty}_{\mathrm{loc}}(R^3)$. At small values of δ, this transformation establishes a one-to-one correspondence between Ω_c and Ω^δ_c. We assume that the vector field $V(x)$ is such that $\Psi_\delta(\Gamma) = \Gamma$ and $\nu^\delta(y) = \nu(x)$, where $\nu^\delta(y)$ is the normal to the perturbed cut $\Gamma^\delta_c = \Psi_\delta(\Gamma_c)$. At each value of δ, we obtain the perturbed domain Ω^δ_c and the boundary value problem that is perturbed with respect to that specified by (6)–(9) and formulated as follows. It is required to find the functions $u^{\lambda\delta} = (u^{\lambda\delta}_1, u^{\lambda\delta}_2, u^{\lambda\delta}_3)$ and $\sigma^{\lambda\delta} = \{ \sigma^{\lambda\delta}_{ij} \}$ ($i, j = 1, 2, 3$) such that

$$-\mathrm{div}\sigma^{\lambda\delta} = f \quad \text{in} \quad \Omega^2_c, \tag{10}$$

$$\sigma^{\lambda\delta} = B^{\lambda\delta}\varepsilon(u^{\lambda\delta}) \quad \text{in} \quad \Omega^\delta_c, \tag{11}$$

$$u^{\lambda\delta} = 0 \quad \text{on} \quad \Gamma, \tag{12}$$

$$\begin{gathered} [u^{\lambda\delta}]\nu \geq 0, \quad [\sigma^{\lambda\delta}_\nu] = 0, \quad \sigma^{\lambda\delta}_\nu \leq 0, \\ \sigma^{\lambda\delta}_\tau = 0, \quad [u^{\lambda\delta}]\nu \cdot \sigma^{\lambda\delta}_\nu = 0 \quad \text{on} \quad \Gamma^\delta_c. \end{gathered} \tag{13}$$

We assume that the coefficients $b^{\lambda\delta}_{ijkl}$ in (11) are defined in Ω^δ_c in such a way that their properties of smoothness are preserved under the map $y = \Psi_\delta(x)$; that is, they remain piecewise smooth,

$$b^\delta_{ijkl} = \begin{cases} c^1_{ijkl} & \text{on} \quad \Psi_\delta(\Omega_1), \\ \lambda^{-1} c^2_{ijkl} & \text{on} \quad \Psi_\delta(\Omega_2). \end{cases}$$

Let $u^{\lambda\delta}$ be a solution to the problem specified by (10)–(13) from the space $H^1(\Omega^\delta_c)$. This solution can be found according to the following scheme. We consider the set of admissible translations in the problem specified by (10)–(13):

$$K^\delta_c = \{ v \in H^1_\Gamma(\Omega^\delta_c) \big| [v]\nu \geq 0 \text{ almost everywhere on } \Gamma^\delta_c \}.$$

Furthermore, we introduce the notation

$$\Pi_\lambda(\Omega_c^\delta;\ v) = \frac{1}{2}\int\limits_{\Omega_c^\delta}\sigma^{\lambda\delta}(v)\varepsilon(v) - \int\limits_{\Omega_c^\delta} fv$$

and consider the problem of minimizing the functional $\Pi_\lambda(\Omega_c^\delta;\ v)$ on the set K_c^δ. A solution to this problem exists and is determined from the variational inequality

$$u^{\lambda\delta} \in K_c^\delta,$$

$$\int\limits_{\Omega_c^\delta}\sigma^{\lambda\delta}(u^{\lambda\delta})\varepsilon(v - u^{\lambda\delta}) \geq \int\limits_{\Omega_c^\delta} f(v - u^{\lambda\delta}) \quad \forall v \in K_c^\delta.$$

This makes it possible to determine the energy functional as

$$\Pi_\lambda(\Omega_c^\delta;\ u^{\lambda\delta}) = \frac{1}{2}\int\limits_{\Omega_c^\delta}\sigma^{\lambda\delta}(u^{\lambda\delta})\varepsilon(u^{\lambda\delta}) - \int\limits_{\Omega_c^\delta} fu^{\lambda\delta}.$$

For the derivative of the energy functional with respect to the parameter δ, we further introduce the notation $I^\lambda = \dfrac{d}{d\delta}\Pi_\lambda(\Omega_c^\delta;\ u^{\lambda\delta})|_{\delta=0}$. According to [6, 12], we then have

$$I^\lambda = \int\limits_{\Omega_c}\left\{\frac{1}{2}\operatorname{div}(Vb_{ijkl}^\lambda)\varepsilon_{kl}(u^\lambda)\varepsilon_{ij}(u^\lambda) - \sigma_{ij}^\lambda(u^\lambda)E_{ij}\left(\frac{\partial V}{\partial x};\ u^\lambda\right)\right\} - \int\limits_{\Omega_c}\operatorname{div}(Vf_i)u_i^\lambda, \tag{14}$$

where $E_{ij}(\Phi;\ v) = \dfrac{1}{2}(v_{i,k}\Phi_{kj} + v_{j,k}\Phi_{ki})$ with $\Phi = \{\Phi_{ij}\}$ $(i, j = 1, 2, 3)$. In view of the above assumption concerning the vector field V, it is not necessary to differentiate the coefficients b_{ijkl}^λ with respect to x_3, which, in general, have a discontinuity along the set Σ. It is well known (see [7, 8]) that, for $\lambda \longrightarrow 0$, there are the strong convergence $\dfrac{1}{\sqrt{\lambda}}u^\lambda \longrightarrow 0$ in $H^1(\Omega_2)$ and the strong convergence $u^\lambda \longrightarrow u^0$ in $H^1(\Omega_1)$, where u^0 is a solution to the problem specified by (1)–(4). From (14), we then derive $I_0 = \lim\limits_{\lambda\to 0} I^\lambda$. The result is

$$I^0 = \int\limits_{\Omega_1}\left\{\frac{1}{2}\operatorname{div}V\cdot\sigma_{ij}(u^0)\varepsilon_{ij}(u^0) - \sigma_{ij}(u^0)E_{ij}\left(\frac{\partial V}{\partial x};\ u^0\right)\right\} - \int\limits_{\Omega_1}\operatorname{div}(Vf_i)u_i^0. \tag{15}$$

INVARIANT INTEGRALS

We now consider specific realizations of the vector field V, which lead to invariant integrals via a transformation of formulas (14) and (15). Since the components of the stress tensor are not defined in general in the domain Ω_2 at $\lambda = 0$, the respective invariant integrals for the problem specified by (1)–(4) and the problem specified by (6)–(9) will be written separately.

In the examples given below, we will have to choose vicinities S_1 and S_2 having Lipschitzian boundaries ∂S_1 and ∂S_2. We assume that the boundaries of the domains also satisfy the Lipschitzian condition.

Example 1. Let $\theta(x)$ be a smooth function equal to zero off a small vicinity S_1 of a curve γ_0 and $\theta = 1$ in a vicinity S_2 of the curve γ_0, $S_2 \subset S_1$. For instance, S_1 and S_2 may be tori that contain γ_0 and which are so small that $(\partial S_1) \cap \Gamma_1$ is part of a flat segment of γ_1. We then choose a perturbation of the domain Ω_c in the form

$$y_1 = x_1 + \delta\theta(x)p_1, \quad y_2 = x_2 + \delta\theta(x)p_2, \quad y_3 = x_3, \tag{16}$$

where $x \in \Omega_c$, $y \in \Omega_c^\delta$, and $p_1^2 + p_2^2 = 1$. In this case, $V(x) = (\theta(x)p_1, \theta(x)p_2, 0)$, while formula (14) takes the form

$$I^\lambda = \int\limits_{\Omega_c}\left\{\frac{1}{2}(\theta_{,l}p_l)\sigma_{ij}^\lambda(u^\lambda)\varepsilon_{ij}(u^\lambda) - \sigma_{ij}^\lambda(u^\lambda)(u_{i,l}^\lambda p_l)\theta_{,j}\right\} - \int\limits_{\Omega_c}(\theta f_i)_{,l}p_l u_i^\lambda. \tag{17}$$

Suppose that $f \equiv 0$ in $S_2 \cap \Omega_c$. It turns out that the smoothness of the respective solution in H^2 up to the points of $\gamma_1\backslash\gamma_0$ and the relevant boundary conditions make it possible to perform integration by parts in (17) and to obtain thereby, for the problem specified by (6)–(9), an invariant integral of the form

$$I^\lambda = \int\limits_{(\partial S_2)\cap\overline{\Omega}_c}\left\{\frac{1}{2}(n_l p_l)\sigma_{ij}^\lambda(u^\lambda)\varepsilon_{ij}(u^\lambda) - \sigma_{ij}^\lambda(u^\lambda)(u_{i,l}^\lambda p_l)n_j\right\}, \tag{18}$$

where $n = (n_1, n_2, n_3)$ is an inward normal to ∂S_2. It should be noted that, if part of the surface $(\partial S_2) \cap \overline{\Omega}_c$ lies on γ_1, integration in (18) can be performed along any edge of the cut Γ_c (see figure).

In the case of the contact problem specified by (1)–(4), similar argument as applied to formula (15) leads to an invariant integral of the form

$$I^0 = \int\limits_{(\partial S_2)\cap\Omega_1} \left\{ \frac{1}{2}(n_l p_l)\sigma_{ij}(u^0)\varepsilon_{ij}(u^0) - \sigma_{ij}(u^0)(u^0_{i,l}p_l)n_j \right\}. \quad (19)$$

Here, $(\partial S_2) \cap \Omega_1$ is a surface belonging to the "cap" type, lying in Ω_1, and covering the curve γ_0.

Example 2. We choose a smooth function θ with a support lying in a small vicinity S_1 of the surface Γ_c and assume that $\theta = 1$ in a vicinity S_2 of the surface Γ_c, $S_2 \subset S_1$. The smallness of the vicinity S_1 means that the edge of the surface $(\partial S_1) \cap \Omega_c$ is part of the flat segment γ_1 of the boundary Γ_1. As before, we consider a perturbation of the domain Ω_c in form (16). For the problem specified by (6)–(9) at $f \equiv 0$ in $S_2 \cap \Omega_c$, the form of the invariant integral coincides with (18) in this case.

From (15), we obtain an invariant integral in the problem specified by (1)–(4), the form of this integral being coincident with (19). In this case, $(\partial S_2) \cap \Omega_1$ is a surface belonging to the "cap" type, lying in Ω_1, and covering Γ_c.

Invariant integrals can also be constructed in the two-dimensional case and for different geometries of the domains Ω_1 and Ω_2. For the two-dimensional analog of Example 1, in particular, the invariant integral for the problem specified by (6)–(9) has the form

$$I^\lambda = \int\limits_{(\partial S_2)\cap\overline{\Omega}_c} \left\{ \frac{1}{2}n_1\sigma^\lambda_{ij}(u^\lambda)\varepsilon_{ij}(u^\lambda) - \sigma^\lambda_{ij}(u^\lambda)u^\lambda_{i,1}n_j \right\},$$

where $(\partial S_2) \cap \overline{\Omega}_c$ is a closed curve having the inward normal $n = (n_1, n_2)$ and surrounding the crack vertex and where summation is performed over $i, j = 1, 2$.

In conclusion, it should be noted that the existence of invariant integrals can be established in some other cases as well. In the above situations, the value of an invariant integral coincides numerically with the value of the derivative of the corresponding energy functional with respect to the perturbation parameter δ at $\delta = 0$. In particular, invariant integrals can be used in seeking approximate expressions for energy functionals in perturbed problems. For this purpose, one can make use of the formula

$$\Pi_\lambda(\Omega^\delta_c; u^{\lambda\delta}) = \Pi_\lambda(\Omega_c; u^\lambda) + \delta I^\lambda + o(\delta),$$

which is valid for all positive values of λ. A similar expansion holds at $\lambda = 0$ as well, but one must replace Ω_c by Ω_1 in this case.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 03-01-00124.

REFERENCES

1. G. P. Cherepanov, *Mechanics of Brittle Fracture* (Nauka, Moscow, 1974) [in Russian].
2. N. F. Morozov, *Mathematical Problems in Crack Theory* (Nauka, Moscow, 1984) [in Russian].
3. S. A. Nazarov, Prikl. Mat. Mekh. **62** (3), 489 (1998).
4. A. M. Khludnev and V. A. Kovtunenko, *Analysis of Cracks in Solids* (WIT, Southampton, 2000).
5. J. Sokolowski and A. M. Khludnev, Prikl. Mat. Mekh. **64** (3), 464 (2000).
6. V. A. Kovtunenko, Prikl. Mat. Mekh. **67** (1), 109 (2003).
7. V. D. Stepanov and A. M. Khludnev, Sib. Mat. Zh. **44** (6), 1350 (2003).
8. K.-H. Hoffmann and A. M. Khludnev, SFB Preprint No. 136 (Bonn, 2004).
9. V. D. Kopchenov, Differ. Uravn. Ikh Primen. **4** (1), 151 (1968).
10. M. B. Brusnikin, Dokl. Akad. Nauk **387**, 151 (2002).
11. N. S. Bakhvalov, Preprint No. 191, OVM AN SSSR (Department of Computational Mathematics, USSR Acad. Sci., Moscow, 1988) [in Russian].
12. A. M. Khludnev, K. Ohtsuka, and J. Sokolowski, Q. Appl. Math. **60** (1), 99 (2002).
13. G. Fichera, Boundary Value Problems of Elasticity with Unilateral Constraints, in *Handbuch der Physik, Bd. 6a/2* (Springer Verlag, Berlin, 1972; Mir, Moscow, 1974).

Translated by A. Isaakyan

Doklady Physics, Vol. 49, No. 10, 2004, pp. 608–613. Translated from Doklady Akademii Nauk, Vol. 398, No. 6, 2004, pp. 764–770.

MECHANICS

Blocking of a Rayleigh Wave by a Subsurface Crack

E. V. Glushkov, N. V. Glushkova, and E. M. Shapar

Presented by Academician V.A. Babeshko April 8, 2004

Received April 20, 2004

1. INTRODUCTION

By now, diffraction of Rayleigh waves (RWs) by surface and subsurface obstacles has been studied relatively well. Flaw detection using reflected and transmitted signals yields information on zones of exfoliation of film structures and nanocoatings, on inceptive defects (fractures) in structural elements, and on surface corrosion of pipelines. Sharp screening of surface waves at certain frequencies is used in acoustoelectronics for producing frequency filters. In geophysics, surface waves provide information on the presence and properties of seismic brakes.

When an RW runs on a crack that does not even reach the surface, partial reflection of the wave with a transformation of a fraction of its energy into bulk waves is also observed. One of the most thorough theoretical studies of dependences of the transmission κ^+, reflection κ^-, and surface-to-bulk conversion κ_ν coefficients for the energy of a plane RW on the size, tilt, and depth of a strip crack was performed by van der Hijden and Neerhoff [1]. In particular, they have shown that, at a certain ratio of depth d of a horizontal crack to the crack half-width a, it is possible to almost completely screen a surface wave. The minimum value of the transmittance $\kappa^+ = 0.0036$ was obtained at $\omega = 3$, $d/a = 0.46$, and $\nu = \frac{1}{3}$ (hereafter, $\omega = \frac{2\pi f a}{v_S}$ is the dimensionless circular frequency, f is the dimensional frequency, v_S is the S-wave propagation velocity, and ν is the Poisson's ratio).

On the other hand, it is known that cracks, as well as bodies of finite dimensions, have natural frequencies (scattering resonance frequencies) that coincide with spectral points of the corresponding boundary value problem. For a crack in an infinite homogeneous space, all resonance poles ω_k are located strictly in the lower half-plane $\mathrm{Im}\omega < 0$ of the complex frequency plane. Particular values of ω_k for circular and elliptic cracks are presented in [2, 3], and their location in the complex plane as a function of the crack shape was studied in [4]. For a strip crack in an infinite medium ($d = \infty$), the obtained sets of values of ω_k, which are closest to the real axis and are localized within a rectangle $0 < \mathrm{Re}\omega < 12$, $-4 < \mathrm{Im}\omega < 0$, are given in Table 1. (The poles ω_k^I and ω_k^{II} correspond to the first mode of the crack opening, i.e., to normal displacement and to shear displacement of the crack sides, respectively.)

Since the poles are complex, their contribution into a transient scattered field determined through the residues in ω_k results in signals $e^{-i\omega_k t}$ exponentially decaying with time and having a damping decrement $|\mathrm{Im}\,\omega_k|$.

For defects in a finite-thickness waveguide (a layer), in which waves are repeatedly reflected between the surfaces, the energy of incoming waves can be trapped, and then trapped modes can be formed with a localization of the wave process (see, e.g., [5]). Here, it is possible that spectral points may lie on the real axis. As applied to cracks, this phenomenon was called "vibration-strength viruses," which manifest themselves only under certain conditions [6]. It is evident that, if the energy is completely trapped at the frequency of the real-valued ω_k, the transmittance κ^+ must be zero.

The following problems remained unsolved for a crack in a half-plane: whether it is possible to completely block an RW, whether real resonance frequen-

Table 1

ω_k^{I}	ω_k^{II}
$0.940 - i \times 0.526$	$1.175 - i \times 1.091$
$2.362 - i \times 0.633$	$2.322 - i \times 1.429$
$3.794 - i \times 0.692$	$3.483 - i \times 1.245$
$5.209 - i \times 0.698$	$5.190 - i \times 0.976$
$6.669 - i \times 0.654$	$6.870 - i \times 1.161$
$8.183 - i \times 0.636$	$8.214 - i \times 1.323$
$9.680 - i \times 0.660$	$9.537 - i \times 1.216$
$11.150 - i \times 0.661$	$11.122 - i \times 1.055$

Kuban State University,
P. O. Box 4102, Krasnodar, 350080 Russia
e-mail: evg@math.kubsu.ru; nvg@math.kubsu.ru;
americal@math.kubsu.ru

1028-3358/04/4910-0608$26.00

cies exist in this case, and how the screening of surface waves is related to the spectral properties of the crack–surface system.

A thorough analysis based on an exact integral representation of the wave field has shown that the transmission of an RW can be blocked by a subsurface crack as a result of two different wave effects: (*1*) the resonant energy trapping at frequencies close to ω_k and (*2*) the antiphase overlapping of the initial incident RW and secondary RWs as a result of the interaction with the surface waves reflected by the crack.

In this case, although some of the resonance poles ω_k are very close to the real axis, no pure real poles have been found. Therefore, in the model under consideration, the first (resonance) mechanism does not ensure complete blocking. Nevertheless, complete blocking can be attained through the second mechanism of antiphase suppression for a limited set of fixed input parameters $\left(\omega, \frac{d}{a}\right)$.

2. GENERAL SCHEME OF SOLVING THE PROBLEM

We consider the two-dimensional problem of steady-state harmonic vibrations $\mathbf{u}(\mathbf{x})e^{-i\omega t}$ of an elastic half-plane containing a strip crack. In a Cartesian coordinate system $\mathbf{x} = (x, z)$, the medium occupies the lower half-plane $z \le 0$ and a crack (an infinitesimally thin straight cut of length $2a$ with stress-free sides) is assumed to be arbitrarily oriented. Its position is determined by depth d of the location of its center $\mathbf{x}_c = (0, -d)$ and by its tilt angle θ.

Along with the global coordinate system $\mathbf{x}$, we also introduce a local coordinate system $\mathbf{x}_1 = (x_1, z_1)$, in which the crack occupies an interval $|x_1| \le a$, $z_1 = 0$. The coordinates of points in these systems are interrelated by formulas $\mathbf{x}_1 = C(\mathbf{x} - \mathbf{x}_1)$ and $\mathbf{x} = \mathbf{x}_c + C_1\mathbf{x}_1$; $C_1 = C^{-1}$, $C(\theta)$ is the matrix of rotation by the angle θ.

Let, for the sake of definitiveness, a given load $\boldsymbol{\tau}|_{z=0} = \mathbf{q}_0(x)$ applied to the medium surface in a vicinity of the point x_0 be a source of vibrations; $\boldsymbol{\tau} = (\tau_{xz}, \sigma_{zz})$ (hereinafter, the harmonic factor $e^{-i\omega t}$ is omitted).

The source wave field $\mathbf{u}_0(\mathbf{x})$ can be represented as a convolution over x of the Green's matrix of the elastic half-plane $k(x, z)$ and the load $\mathbf{q}_0(x)$ or, in an alternative form, through their Fourier transforms $K(\alpha, z) = \mathcal{F}[k]$ and $\mathbf{Q}_0(\alpha) = \mathcal{F}[\mathbf{q}_0]$:

$$\mathbf{u}_0(x, z) = \frac{1}{2\pi}\int_\Gamma K(\alpha, z)\mathbf{Q}_0(\alpha)e^{-i\alpha x}d\alpha. \qquad (1)$$

Here, $\mathcal{F}$ and $\mathcal{F}_1$ are the operators of the Fourier transform with respect to x and x_1, respectively. The form of both $K(\alpha, z)$ components for an isotropic elastic half-plane and of the interrelation path Γ is well known. The details of deriving this integral representation of wave fields can be found, e.g., in [7, 8].

The total wave field $\mathbf{u}$ is composed of the source field $\mathbf{u}_0$, the field $\mathbf{u}_1$ of waves reflected by the crack, and the field $\mathbf{u}_2$ appearing as a result of a repeated reflection of $\mathbf{u}_1$ from the half-plane surface $z = 0$: $\mathbf{u} = \mathbf{u}_0 + \mathbf{u}_1 + \mathbf{u}_2$.

For the field $\mathbf{u}_1$, the representation in the local coordinate system through the jump of displacements across the crack edges: $\mathbf{v}(x_1) = \mathbf{u}_1\big|_{z_1 = 0^-} - \mathbf{u}_1\big|_{z_1 = 0^+}$ is known. This representation was also obtained using the symbols K^+ and K^- of the Green's matrix for the upper ($z_1 \ge 0$) and lower ($z_1 \le 0$) half-planes [9]:

$$\mathbf{u}_1^\pm(\mathbf{x}_1) = \frac{1}{2\pi}\int_\Gamma K^\pm(\alpha_1, z_1)L_1(\alpha_1)\mathbf{V}(\alpha_1)e^{-i\alpha_1 x_1}d\alpha_1. \qquad (2)$$

Here,

$$\mathbf{V}(\alpha_1) = \mathcal{F}_1[\mathbf{v}(x_1)],\ L_1(\alpha_1) = [K^+(\alpha_1, 0) - K^-(\alpha_1, 0]^{-1}.$$

The field $\mathbf{u}_2$ is introduced for compensating the traction $\boldsymbol{\tau}_1$ induced by the field $\mathbf{u}_1$ at the stress-free surface $z = 0$. Therefore, denoting $\mathbf{q}_2(x) = -\boldsymbol{\tau}_1|_{z=0}$, we obtain for $\mathbf{u}_2$ the same representation (1) as for $\mathbf{u}_0$ with $\mathbf{Q}_0$ replaced by $\mathbf{Q}_2 = \mathcal{F}[\mathbf{q}_2]$. By virtue of interrelation between $\mathbf{q}_2$ and $\mathbf{u}_1$, $\mathbf{Q}_2$ can be expressed analytically in terms of $\mathbf{V}$:

$$\mathbf{Q}_2(\alpha) = M(\alpha(\alpha_1))\mathbf{V}(\alpha_1(\alpha)). \qquad (3)$$

Matrix M is expressed through a chain of products of the matrices of the Fourier transforms K and L, of the stress operator T and rotations C and C_1.

An unknown jump $\mathbf{v}$ is determined from the condition of a zero total stress at the crack sides:

$$(\mathbf{q}_1 + \boldsymbol{\tau}_2 + \boldsymbol{\tau}_0)\big|_{z_1 = 0} = 0, \quad |x_1| \le a. \qquad (4)$$

Here, $\mathbf{q}_1 = \mathcal{F}_1^{-1}[L\mathbf{V}]$ is a traction in the line $z_1 = 0$, which correspond to the field $\mathbf{u}_1$ (in the local system $\mathbf{x}_1$), $\boldsymbol{\tau}_n = CT_1\mathbf{u}_n$, $n = 0, 2$ are stresses corresponding to the fields $\mathbf{u}_0$ and $\mathbf{u}_2$, and T_1 is the stress operator in the crack plane.

In view of (1)–(3), condition (4) leads to an integral equation with respect to $\mathbf{v}$:

$$\mathcal{L}_1\mathbf{v} + \mathcal{L}_2\mathbf{v} \equiv \int_{-a}^{a}[l_1(x_1 - \xi_1) + l_2(x_1, \xi_1)]\mathbf{v}(\xi_1)d\xi_1 = \mathbf{f}(x_1), \quad |x_1| < a, \quad \mathbf{f} = -\boldsymbol{\tau}_0\big|_{z_1 = 0}. \qquad (5)$$

Its matrix kernel is composed of the principal hypersingular part $l_1(x_1) = \mathcal{F}_1^{-1}[L_1(\alpha_1)]$, which arises in the process of solving the problem of scattering of elastic

waves by a crack in an infinite space, and a smooth (at $d > 0$) addition $l_2 = \mathcal{F}_1^{-1}[L_2(\alpha(\alpha_1))]$, which takes into account an effect of all repeated reflections between the crack and the half-plane surface on the value of the Fourier symbol L_2 is expressed as the product of matrices C, T_1, and K and comultipliers entering into $\mathbf{Q}_2$ [except for $\mathbf{v}(\mathbf{x}_1)$]. The smoothness of l_2 is ensured by the exponentially decaying symbol $L_2(\alpha) \sim O(e^{-d|\alpha|})$ as $|\alpha| \longrightarrow \infty$. For a crack touching the surface ($d = 0$), the kernel l_2 also becomes singular.

The solution to integral equation (5) is sought using the Galerkin scheme in the form of the expansion

$$\mathbf{v}_N(x_1) = \sum_{k=1}^{N} \mathbf{c}_k \varphi_k(x_1),$$

with unknown coefficients $\mathbf{c}_k$ to be determined from the set of linear algebraic equations

$$\sum_{k=1}^{N} a_{jk}\mathbf{c}_k = \mathbf{f}_j, \quad j = 1, 2, \ldots, N,$$
$$a_{ij} = ((\mathcal{L}_1 + \mathcal{L}_2)\varphi_k, \varphi_j), \quad \mathbf{f}_j = (\mathbf{f}, \varphi_j). \tag{6}$$

Here, round brackets $(f, g) = \int f \cdot g^* dx$ denote a scalar product in $L_2(-a, a)$.

As basis functions φ_k, we used both orthogonal polynomials with the weight factor $\sqrt{a^2 - x_1^2}$ conforming with the behavior of $\mathbf{v}$ as $x_1 \longrightarrow \pm a$ and conventional splines providing a piecewise-linear approximation.

The zeros of the matrix determinant of the system ω_k, $\det A(\omega_k) = 0$, approximate spectral points of the integral operator. For the introduced dimensionless variable ω, the roots ω_k depend only on the ratio $\frac{d}{a}$ and on crack tilt angle θ.

The RW $\mathbf{u}_{0,R}$ excited in a half-plane by a load $\mathbf{q}_0$ is described by the residue of the integrand in (1) at the Rayleigh pole $-\zeta$ or ζ of matrix K:

$$\mathbf{u}^{\pm}_{0,R}(\mathbf{x}) = \mathbf{a}^{\pm}_0(z)e^{\pm i\zeta x}, \quad \mathbf{a}^{\pm}_0 = R^{\pm}\mathbf{Q}_0(\mp\zeta),$$
$$R^{\pm} = \mp i \operatorname{res} K|_{\alpha = \mp\zeta} \tag{7}$$

($\mathbf{u}^{+}_{0,R}$ and $\mathbf{u}^{-}_{0,R}$ are waves propagating to the right and to the left from the source, respectively). Similarly, $\mathbf{u}^{\pm}_{2,R}$ is the RW of the field $\mathbf{u}_2$ of form (7), where $\mathbf{a}^{\pm}_2$ is expressed in terms of $\mathbf{Q}_2(\mp\zeta)$ of form (3). Field $\mathbf{u}_1$ contains only bulk waves.

Thus, the time-averaged (over the vibration period) energy $E_{0,R}$ of an RW incoming from the left to the crack is determined by the amplitude $\mathbf{a}^{+}_0$ of the field $\mathbf{u}_0$, whereas the energies of the reflected and transmitted waves E^{-}_R and E^{+}_R are determined by the amplitude $\mathbf{a}^{-}_2$ and the sum of the amplitudes $\mathbf{a}^{+}_0 + \mathbf{a}^{+}_2$, respectively. The RW transmittance and reflectance are determined by the relationships $\kappa^{\pm} = \frac{E^{\pm}_R}{E_{0,R}}$, and the surface-to-bulk conversion coefficient is $\kappa_v = \frac{E_v}{E_0}$. Here, E_v is the energy of bulk waves $\mathbf{u}_1$ and a part of $\mathbf{u}_2$ without $\mathbf{u}^{\pm}_{2,R}$. It is determined by integrating the power-density over the lower half-circle $x = R\cos\psi$, $y = R\sin\psi$, $-\pi \le \psi \le 0$ as $R \longrightarrow \infty$ [7–9].

3. ANALYSIS OF TRANSMITTANCE AND SPECTRAL PROPERTIES

Although the numerical analysis was performed at various tilt angles θ of a crack, in order to simplify our description, the basic result (i.e., the presence of two screening mechanisms) is demonstrated for a horizontal crack ($\theta = 0$) as an example. The coincidence of the dependences of $\kappa^{\pm}$ on ω and the ratio d/a for all crack angles θ presented in [1] along with the verification of the energy balance $E_{0R} = E^{+}_R + E^{-}_R + E_v$ served as a test of the reliability of the numerical results.

Figure 1 shows a typical behavior of $\kappa^{\pm}$ as a function of θ at $\frac{d}{a} = 0.2$ and 0.3311 and $\nu = 1/3$. At $\frac{d}{a} = 0.2$, narrow (resonant) dips in the κ^{+} plot are observed. They indicate strong screening of the RWs at the corresponding frequencies. At $\frac{d}{a} = 0.3311$, wide and deep depressions with minima at $\omega = 2.3$, 3.5, and 4.8 are added to the dips. To elucidate the relation of the screening effect to the spectral properties of the model under consideration, Fig. 2 shows the dependences on the ratio $\frac{d}{a}$ of the real and imaginary parts of five resonance scattering frequencies ω_k nearest to the real axis. For the values of $\frac{d}{a}$ demonstrated in Fig. 1, these poles are listed in Table 2.

As is seen, with decreasing the ratio $\frac{d}{a}$, these poles virtually lie on the real axis, but $\operatorname{Im}\omega_k$ still differs from zero as $d > 0$. Nevertheless, the closeness of ω_k to the real axis leads, first, to an abrupt increase in $|\mathbf{v}|$ at $\omega \approx \operatorname{Re}\omega_k$ and, as a consequence, to a resonant increase in the stress-intensity factors near crack edges, which are expressed in terms of $v^{\pm}_0 = \lim_{x \to \pm a} \mathbf{v}(x)/\sqrt{a^2 - x^2}$. Sec-

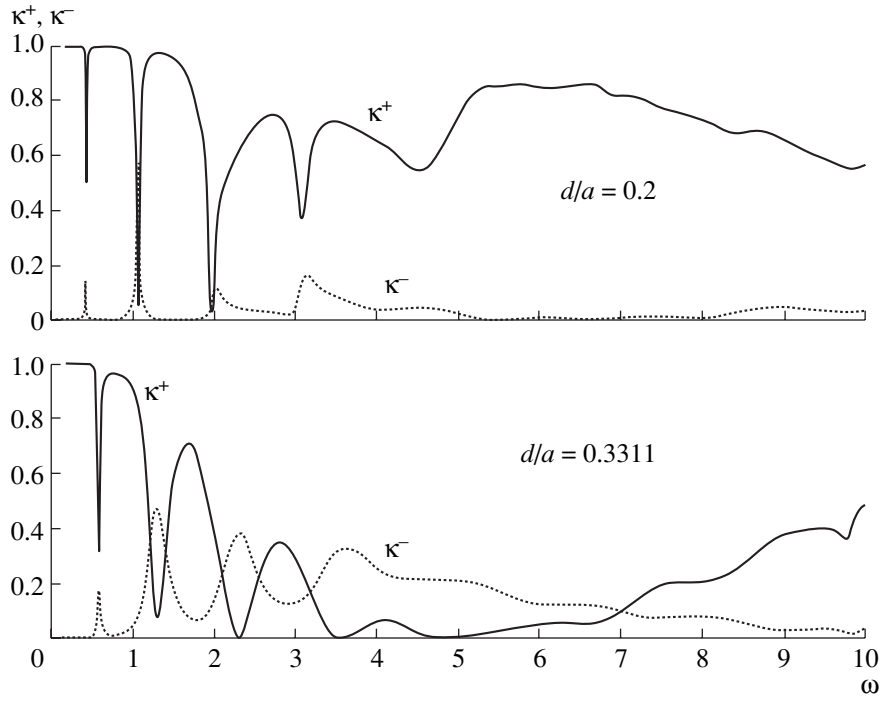


Fig. 1. Examples of partial resonant $\left(\frac{d}{a} = 0.2\right)$ and complete nonresonant $\left(\frac{d}{a} = 0.3311\right)$ screening of a Rayleigh wave by a subsurface crack.

ond, these are the frequencies at which sharp dips of the function $\kappa^+(\omega)$ are observed.

Of course, resonant energy trapping and localization of the wave process occur in this case. This is also implied by a sharp (several orders of magnitude) increase in the surface displacements $(\mathbf{u}_1 + \mathbf{u}_2)|_{z=0}$ above the crack at $\omega = \mathrm{Re}\omega_k$, $|\mathrm{Im}\omega_k| \ll 1$. However, wide minima of κ^+ at $\frac{d}{a} = 0.3311$ cannot be compared to a value of ω_k lying close to the real axis. The absence of both a resonant increase in the intensity coefficients and surface displacements above the crack also indicates that no energy trapping and no localization of the oscillatory process occur in these cases.

On the other hand, using a numerical minimization of function $\kappa^+\left(\omega, \frac{d}{a}\right)$ just in the vicinity of these frequencies, we have discovered discrete points in the plane $\left(\omega, \frac{d}{a}\right)$ in which the transmission of the RW is completely blocked ($\kappa^+ \equiv 0$!). The coordinates of these points for different Poisson's ratios ν are listed in Table 3.

For ω values not too close to zero, $\omega \geq \varepsilon_1$ and $\frac{d}{a} \geq \varepsilon_2$ (ε_1 and ε_2 are small), the set of complete-blocking parameters presented in Table 3 is exhaustive. This is confirmed by the form of the constructed surfaces $y = \kappa^+\left(\omega, \frac{d}{a}\right)$, which does not allow the appearance of minima $\kappa^+ = 0$ in other zones. For example, in the case

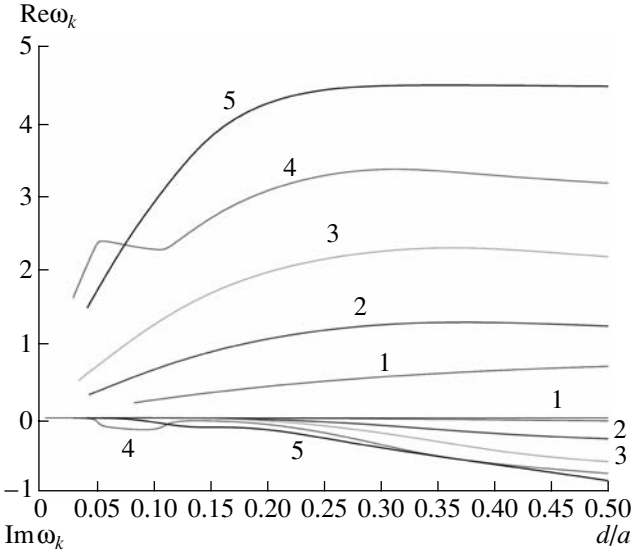


Fig. 2. Dependences of the poles ω_k closest to the real axis on the crack depth $\frac{d}{a}$.

Table 2

$\frac{d}{a}$	ω_1	ω_2	ω_3	ω_4	ω_5
0.2	$0.43–i\times0.004$	$1.07–i\times0.02$	$1.97–i\times0.04$	$3.07–i\times0.09$	$4.24–i\times0.16$
0.3311	$0.58–i\times0.02$	$1.27–i\times0.13$	$2.26–i\times0.27$	$3.32–i\times0.46$	$4.45–i\times0.47$

Table 3

ν	$\omega, \frac{d}{a}$			
	1	2	3	4
0.1	2.036, 0.431	–	–	–
0.2	2.157, 0.398	3.025, 0.439	3.220, 0.359	–
0.3	2.296, 0.371	2.911, 0.463	3.446, 0.336	–
0.3333	2.348, 0.364	2.865, 0.459	3.528, 0.331	–
0.4	2.461, 0.351	2.753, 0.429	3.709, 0.323	5.069, 0.334

of $\nu = \frac{1}{3}$, the surface relief $0 \le \kappa^+\left(\omega, \frac{d}{a}\right) \le 1$ is shown in Fig. 3 by level lines. A characteristic feature of the relief is the presence of a deep valley extending from northwest to southeast, approximately, from the point $\left(\omega, \frac{d}{a}\right) = (2.4, 0.5)$ to (7, 0.3). Gray and black colors indicate zones in which $\kappa^+ \le 0.01$ and $\kappa^+ \le 0.001$, respectively.

The global minima of the function κ^+ were sought in these zones, since motion from the valley in an arbitrary direction leads only to an increase in κ^+. This increase, with κ^+ tending to unity as $\frac{d}{a} \longrightarrow \infty$, is obviously associated with the fact that, with increasing depth, the RW interaction with the crack ceases, as also occurs at $\omega \longrightarrow \infty$ when the depth-to-wavelength ratio becomes large. A monotonic increase in κ^+ is violated only near the axis $\frac{d}{a} = 0$, where the relief becomes very irregular due to almost real ω_k. Therefore, the presence of global-minimum points in the band $\frac{d}{a} < \varepsilon_2$ also cannot be totally excluded.

The complete-blocking mechanism becomes clear if we analyze the reasons for vanishing of the amplitude $\mathbf{u}_R^+$ of the transmitted RW. Let, for simplicity, a vertical concentrated load $\mathbf{q}_0 = (0, \delta(x - x_0))$ be the source. In this case, $\mathbf{Q}_0 = (0, e^{i\alpha x_0})$, $\mathbf{a}_0^+ = \mathbf{r}e^{i\zeta x_0}$, where $\mathbf{r}$ is the second column of matrix $R^+ = (s\mathbf{r}, \mathbf{r})$ [see (7)]. The first column of R^+ is proportional to the second one with a certain factor s [a linear dependence of the columns of matrix $\Delta \cdot K(\alpha, z)$, where $\Delta(\alpha)$ is the Rayleigh denominator, exists only for $\alpha = \zeta$]. Correspondingly, $\mathbf{a}_2^+ = p\mathbf{r}$ with an amplitude factor $p = sQ_2^{(1)}(-\zeta) + Q_2^{(2)}(-\zeta)$. Obviously, $\kappa^+ \equiv 0$ for $f(\omega) = 1 + p(\omega)e^{-i\zeta x_0} \equiv 0$.

Analyzing the plots of function $f(\omega)$ has shown that, at points of complete blocking, Ref and Imf simultaneously change their signs, smoothly passing through zero. Thus, at such $\left(\omega, \frac{d}{a}\right)$, the RW $\mathbf{u}_{2,R}^+$ appearing upon the incidence of $\mathbf{u}_1$ on the surface $z = 0$ becomes equal in amplitude to $\mathbf{u}_{0,R}^+$ but oscillates in antiphase to it.

On the contrary, in the vicinity of the resonance frequencies ω_k, the plot of $f(\omega)$ has typical oppositely directed spikes indicating the presence of polar factors $(\omega - \omega_k)^{-1}$. Within a narrow interval between the spikes, Ref and Imf also change their signs, but no complete annihilation is observed here, since Ref and Imf have different zeros with complex ω_k.

Hence, there exist two mechanisms of screening a Rayleigh wave by a subsurface crack. The first one is related to a resonance in the crack–surface system, but it does not ensure complete screening for the model under study. The second mechanism results from compensation of the initial wave by the secondary RW excited in antiphase. This mechanism is not related to the resonant wave localization and, at certain combina-

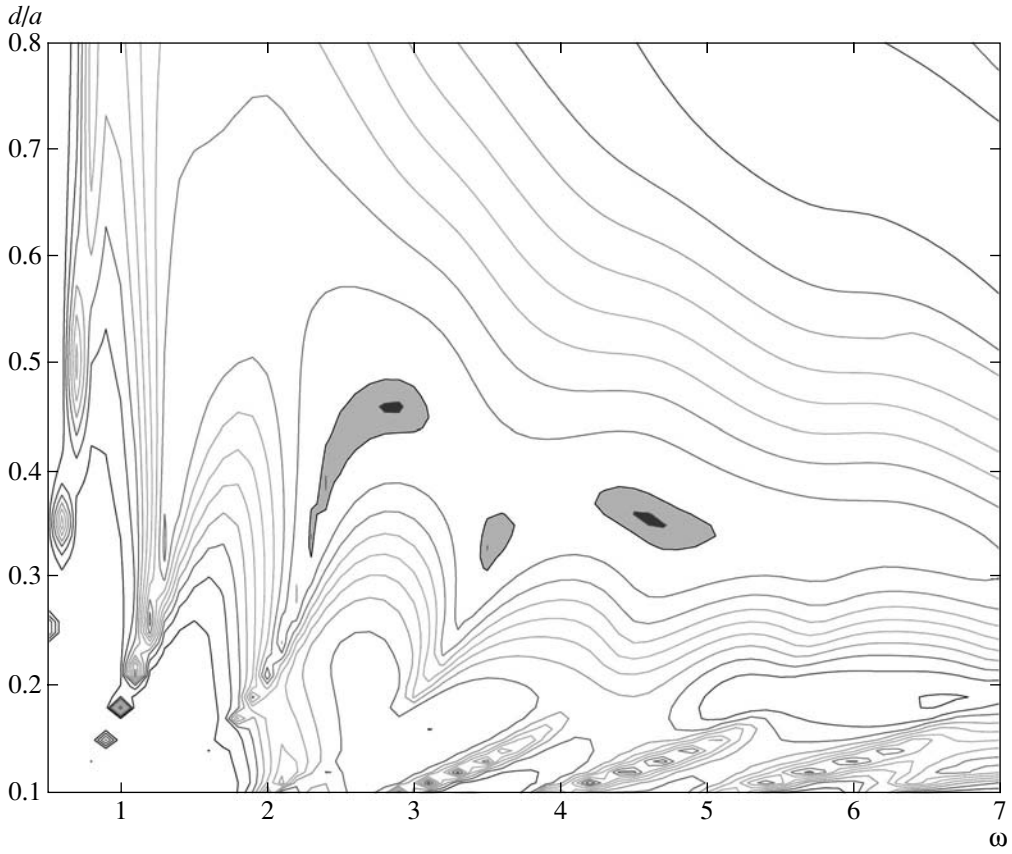


Fig. 3. Level lines for the transmittance κ^+ on the $\left(\omega, \frac{d}{a}\right)$ plane.

tions of the frequency, with the crack depth and size, ensures complete screening of a surface wave.

ACKNOWLEDGMENTS

The authors are grateful to Academician V.A. Babeshko and Prof. D.A. Indeĭtsev for fruitful discussions of various aspects of the problem.

This work was supported by the Russian Foundation for Basic Research, project nos. 03-01-00526 and 03-01-96626.

REFERENCES

1. J. H. M. T. van der Hijden and F. Neerhoff, J. Acoust. Soc. Am. **756**, 1694 (1984).
2. A. S. Eriksson, Trans. ASME, J. Appl. Mech. **62** (1), 59 (1995).
3. C. J. S. Alves and T. Ha Duong, Int. J. Numer. Methods Eng. **38**, 2347 (1995).
4. E. V. Glushkov and N. V. Glushkova, Prikl. Mat. Mekh. **62**, 866 (1998).
5. C. M. Linton and D. V. Evans, Q. J. Mech. Appl. Math. **44** (3), 487 (1991).
6. V. A. Babeshko, Izv. Vyssh. Uchebn. Zaved. Sev.-Kavk. Region, Spec. Issue **90** (1994).
7. V. A. Babeshko, E. V. Glushkov, and Zh. F. Zinchenko, *Dynamics of Inhomogeneous Linearly Elastic Media* (Nauka, Moscow, 1989).
8. N. V. Glushkova, Doctoral Dissertation in Physics and Mathematics (Rostov. Gos. Univ., Rostov-on-Don, 2000).
9. E. V. Glushkov and N. V. Glushkova, Prikl. Mat. Mekh. **60**, 282 (1996).

Translated by A. Seferov

Doklady Physics, Vol. 49, No. 10, 2004, pp. 614–616. Translated from Doklady Akademii Nauk, Vol. 398, No. 6, 2004, pp. 771–773.

MECHANICS

On Conclusions of the Action-Inversion Equation in the Case of Heat Removal in the Nozzle of a Liquid-Fuel Jet Engine

L. E. Sternin

Presented by Academician B.I. Katorgin April 9, 2004

Received April 20, 2004

The action-inversion law and the corresponding differential equation have been published by L.A. Vulis [1, pp. 93–95] and also have been presented by G.N. Abramovich in popular textbooks on applied gas dynamics (see, e.g. [2, pp. 189, 190]). At present, the law and the equation are well known and used for solving numerous quasi-one-dimensional scientific and applied problems. Without making allowance for expendable mechanical action and the effect of friction, the equation is written in the following simple form:

$$(\mathrm{M}^2 - 1)\frac{dw}{w} = \frac{dF}{F} - \frac{\gamma - 1}{a^2}dQ, \qquad (1)$$

where M is the Mach number, w is the flow velocity, F is the area, Q is the delivered heat (in the case of the heat removal, $dQ < 0$), a is the local velocity of sound, and γ is the adiabatic exponent.

Analyzing this equation in the supersonic region, we can draw the conclusion that, as far as the case of the heat removal in the nozzle is concerned, when the right-hand side of Eq. (1) increases, this must result in an additional velocity increase with respect to adiabatic flow. This appears to have been taken in [1, 2] as a decisive argumentation. However, here, a contradiction appears in the results of direct calculations performed on the basis of the complete set of equations for quasi-one-dimensional flows in cooled supersonic nozzles of liquid-fuel jet engines in which heat removal attains 1–2% of the deceleration enthalpy. These calculations show that, in fact, due to heat removal from gas moving in a nozzle, not an increase—but a decrease—of flow velocity with respect to the adiabatic flow (hereinafter, this velocity is called the relative velocity) occurs. A similar pattern is also manifested for high levels of the heat removal that rises along the nozzle. This effect on the flow velocity is associated with the intense increase in the Mach number entering into Eq. (1) that is stipulated by the heat removal. By virtue of a small flow deceleration caused by the heat removal with respect to the adiabatic flow, the flow velocity along the nozzle in any case increases, owing to the more intense geometric action on the flow.

In order to demonstrate the relative decrease in the flow velocity due to the heat removal, we perform the following simple analysis. It is easy to show that, as a result of heat removal in a supersonic nozzle, the Mach number becomes larger, whereas the static flow temperature drops.

Equation (1) can be rewritten for a supersonic nozzle in the form

$$\frac{dw}{d\ln F} = \left(1 - \frac{1}{c_p T}\frac{dQ}{d\ln F}\right)\frac{w}{\mathrm{M}^2 - 1}.$$

The effect of the heat removal on the variation of the flow velocity w in the nozzle is determined by comparing the derivatives $\frac{dw}{d\ln F}$ for flows with the heat removal (subscript Q) and without it (no subscript) in a certain current cross section 2 of the nozzle. This cross section is located at a finite distance downstream of the initial cross section 1.

First, we consider the case in which heat removal begins from cross section 1 ($Q_1 = 0$, $\mathrm{M}_1 > 1$). The difference of the derivatives for the velocity with respect to the logarithm of the cross-section area for flows with and without heat removal in cross section 2 (the derivative is zero in the cross section 1) features the effect of the heat removal on the velocity variation along the nozzle,

$$\Delta = \frac{dw_{Q,2}}{d\ln F} - \frac{dw_2}{d\ln F} = \left(\frac{w_{Q,2}}{\mathrm{M}_{Q,2}^2 - 1} - \frac{w_2}{\mathrm{M}_2^2 - 1}\right) - \left(\frac{w_{Q,2}}{c_p T_{Q,2}(\mathrm{M}_{Q,2}^2 - 1)}\frac{dQ}{d\ln F}\right). \qquad (2)$$

Glushko NPO Énergomash,
ul. Burdenko 1, Khimki,
Moscow oblast, 141400 Russia
e-mail: energo@online.ru

1028-3358/04/4910-0614$26.00 © 2004 MAIK “Nauka/Interperiodica”

Since $\frac{dQ}{d\ln F} < 0$ always, the second bracket (with the minus sign in front of it) yields a positive contribution to the right-hand side of expression (2). We now can determine the sign standing ahead of the first bracket in this formula,

$$A \equiv \left(\frac{w_{Q,2}}{M_{Q,2}^2 - 1} - \frac{w_2}{M_2^2 - 1}\right) = \frac{M_{Q,2}a_{Q,2}}{M_{Q,2}^2 - 1} - \frac{M_2 a_2}{M_2^2 - 1}. \quad (3)$$

It is evident that $A < 0$, because, due to the heat removal, M increases, whereas velocity a of sound decreases.

At the initial point 1 of the calculation, $A = 0$; originally, the right-hand side of expression (2) is positive, and the relative velocity increases in this case. However, with increasing F and, hence, increasing modulus of Q, the quantity A, remaining negative, rises in absolute value. As a result, for a certain F, the right-hand side of (2) vanishes and, furthermore, in the case of heat removal, the relative flow velocity in the supersonic nozzle begins to decrease. We should emphasize once more that the relative-velocity drop caused by the heat removal in the nozzle is stipulated by the combined effect on the Mach number of the terms in the right-hand side of Eq. (1), although, if taken separately, each of them increases the relative velocity. This fact, apparently, was missed by the authors of [1, 2].

In the general case, when the heat removal begins prior to cross section 1, $M_{Q.2} > M_2$, and the parameter A becomes negative as it follows from (3) (but not zero, as it took place in the preceding case). The right-hand side of (2) decreases and can become negative.

Figure 1 presents a variation of the relative difference in the flow velocities in the supersonic nozzle for flows with heat removal and adiabatic flow as a function of the relative radius of the nozzle wall (i.e., the radius of the nozzle contour is related to the nozzle–neck radius). The local rise of the relative difference in the velocities at the onset of the heat removal (bump) from point 1 is stipulated by the calculation model according to which the heat removal is absent upstream to point 1. The variation of relative velocities due to heat removal in subsonic and supersonic parts of the nozzle is shown in Fig. 2. The calculations of flow in the presence of the heat removal were performed by the inverse method, i.e., initially, the flow velocity was specified and, then, the flow area of the nozzle cross section was calculated. This approach ensured an easy pass by the singular saddle point in the vicinity of the velocity of sound. As follows from Fig. 2, in the actual case of heat removal in the nozzle of a liquid-fuel jet engine (i.e., in the case of existence of heat removal in the input subsonic nozzle cross section), a decrease in the relative velocity caused by the heat removal occurs partly over the narrowing part of the nozzle, as well as over its entire expanding part, and is monotonous.

The decrease in the relative velocity stipulated by the existence of heat removal in the supersonic nozzle

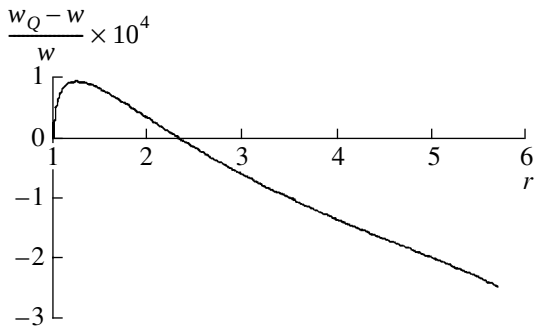


Fig. 1. Variation of the relative difference in the flow velocities in a nozzle.

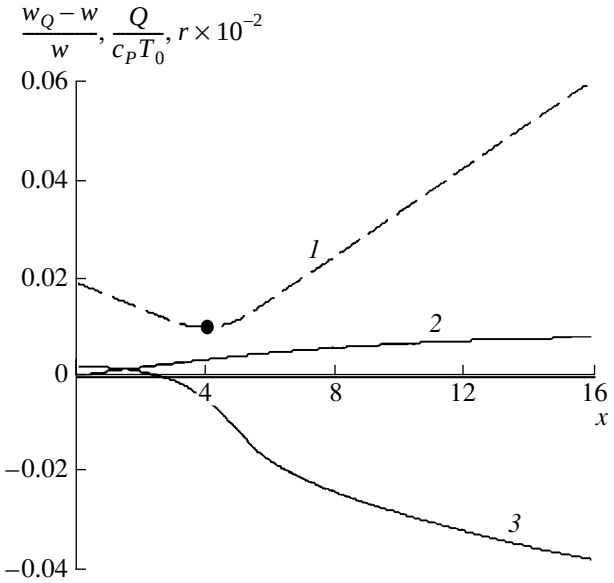


Fig. 2. Flow in a nozzle is accompanied by passing the value of its velocity through that of velocity of sound: (*1*) nozzle contour, (*2*) heat removal, and (*3*) relative velocity difference. x is the ratio of the nozzle length to the radius of the critical cross section, the dot corresponds to the passing through the velocity of sound. $M_Q = 1$ and T_0 is the deceleration temperature prior to the nozzle.

was noted in [3, pp. 82–85]. In the conditions under consideration, a maximum of the relative velocity also exists. In a cylindrical channel, $\frac{dF}{F} = 0$, and, for $M_Q > 1$, the second bracket in formula (2) disappears. Therefore, the right-hand side of this formula is always positive and, in the case of the heat removal, the flow velocity in the cylindrical channel always increases. Flows subjected to separate actions including friction are analyzed in [4]. In the case of an intense rise of the heat removal along the nozzle, which is not characteristic of liquid-fuel jet engines, a certain increase in the gas velocity with respect to adiabatic flow in the nozzle is also possible [3]. The mechanical action on the flow [1, 2] is equivalent to the delivery (or removal) of the energy to

(from) gas, so we can consider this action as a thermal one.

Thus, by virtue of heat removal, the velocity along a supersonic cooled nozzle of a liquid-fuel jet engine slightly decreases with respect to the adiabatic flow. The conclusions sometimes made on the basis of the action-inversion formula predicting the acceleration of supersonic flow in a nozzle cannot be considered to be true. Indeed, the analysis of the differential equation describing the action inversion, which was carried out in [1, 2], determines only the local character of the solution. Therefore, it is impossible to find flow parameters in the entire flow region without using other equations of the system under consideration.

ACKNOWLEDGMENTS

The author is grateful to B.I. Katorgin, G.G. Chernyĭ, and A.N. Kraĭko for fruitful discussions.

REFERENCES

1. L. A. Vulis, *Thermodynamics of Gas Flows* (Gosénergoizdat, Moscow, 1950).
2. G. N. Abramovich, *Applied Gas Dynamics* (Nauka, Moscow, 1969).
3. L. E. Sternin, *Principles of Gas Dynamics* (Izd. MAI, Moscow, 1995).
4. G. G. Chernyĭ, *Gas Dynamics* (Nauka, Moscow, 1988).

Translated by G. Merzon