THEORETICAL
AND MATHEMATICAL PHYSICS

Extreme Ultraviolet Nanolithography for ULSI:
A Review

R. P. Seisyan

Ioffe Physicotechnical Institute, Russian Academy of Sciences,
ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia
e-mail: rseis@ffm.ioffe.ru
Received July 26, 2004

Abstract—The basic concepts of high-resolution extreme ultraviolet nanolithography, which is aimed at producing ultra-large-scale integrated circuits (with an integration one or two orders of magnitude exceeding present-day integration levels), are reviewed and substantiated. The problems in and the current status of this field of technology are considered. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

The central issue behind the fast progress in microelectronics over the past 40 years has been the continuous shrinkage of the minimal feature size (a_{min}). Until recently, the advances have been provided by refining photolithography equipment operating in the visible, ultraviolet, and near-vacuum ultraviolet ranges of electromagnetic radiation. A resolution as high as $a_{min} \approx 0.1 \ \mu m$ has been achieved to date using modern photolithographic tools.

The key advantage of optical patterning is the possibility of parallel transfer of a pattern consisting of many millions of features owing to the wave nature of optical processes. It is this approach that has provided a high economic effectiveness of planar technology and made it possible to fabricate ultra-large-scale integrated circuits with an integration level of 10^7–10^8 transistors per chip.

Analysis of the current tendency toward a continuous decrease in the integrated circuit (IC) minimal feature size shows that this parameter will inevitably go far beyond the submicrometer range down to $a_{min} = 10$–60 nm.

Figure 1, based on a great deal of published data, demonstrates how a_{min} has varied with time. It is seen that the progress in microelectronics throughout its evolution has been related to an exponential decrease in a_{min}, with the exponent of the evolution curve being the same; namely, a_{min} has decreased by about 13% annually. However, the capabilities of optical systems are restricted by the diffraction limit: the pattern linewidth cannot be much smaller than the radiation wavelength. According to the Rayleigh criterion, the minimal feature size (linewidth) is proportional to radiation wavelength λ,

$$a_{min} = k\lambda/NA, \qquad (1)$$

where k is a proportionality coefficient and NA is the numerical aperture of the objective lens. A further decrease in the wavelength, which implies a deeper advance into the vacuum ultraviolet, poses challenging problems. These are the absence of natural transparent materials suitable for UV optics and pattern photomasks, as well as the difficulty in designing efficient radiation sources in this wavelength range. This necessitates a search for nontrivial technological solutions providing a further progress in microelectronics. Specifically, flying out of the sub-100-nm range becomes a bottleneck, since materials that do not too highly absorb electromagnetic radiation in this wavelength range (up to hard X rays) are lacking.

Up to now, focused electron-beam lithography has been viewed as a method that might fully meet the requirements of submicron electronics. The potential of electron-beam lithography in achieving a high resolution is well known: transmission electron microscopes currently available provide a resolution as high as several angstroms. However, fast electrons penetrating deep into the material cause a number of secondary effects smearing the pattern. Yet, nanometer feature sizes here are quite feasible; unfortunately, another problem is highlighted in this case. Focused electron-beam lithography implies sequential (point- or feature-wise) exposure, which greatly extends the patterning time. This circumstance is the irreconcilability of the tendency toward an exponential increase in the integration level (Fig. 1b).

The radical way out in this situation (as in other situations bearing on a very large number of events) or, in other words, provision of efficient further development consists in making a multitude of elementary processes proceed in parallel. In practice, the projection versions of electron- and ion-beam methods of lithography run into the mask problem, which can hardly be resolved. An electron-(ion-)beam mask contains small patterns
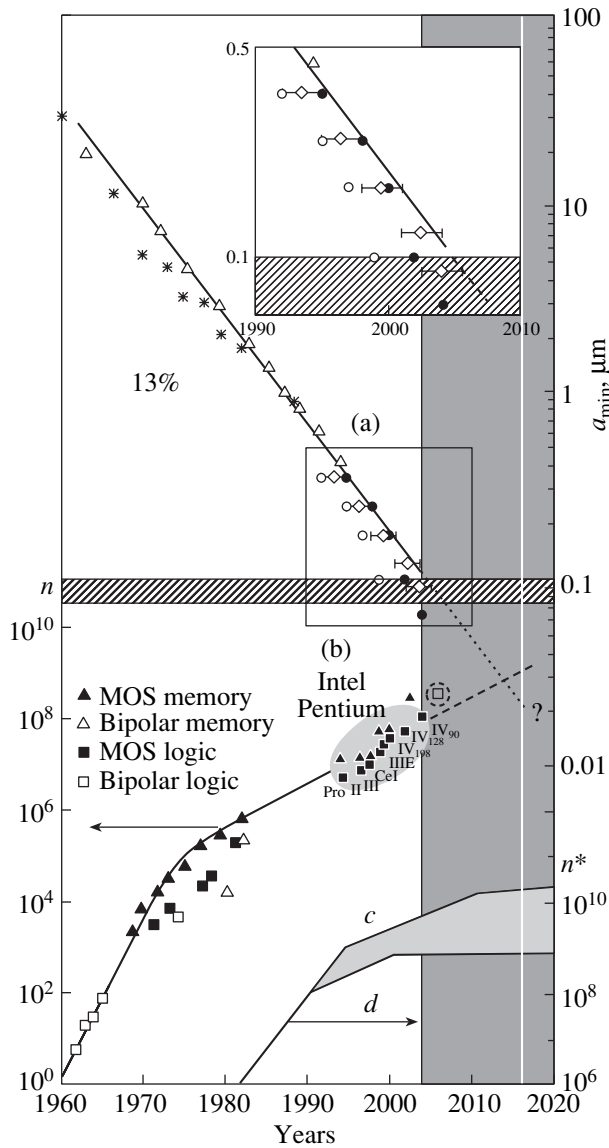
Fig. 1. Variation of minimal feature size (linewidth) a_{min} (upper part) and integration level (lower part). Year 2000: Pentium IV, $n = 4.42 \times 10^6$, 0.18 μm design rule; Pentium IV Northwood, $n = 5.5 \times 10^7$, 0.18 μm design rule. Year 2003: Pentium IV Prescott ($n = 1.2 \times 10^8$) and SRAM ($n = 3.3 \times 10^8$), 0.09 μm design rule. Year 2006: $n = 3.5 \times 10^8$ (Intel's forecast). Lines c and d: Meindl's forecast (Stanford University, the 1980s) according to which the minimal linewidth provided by electron-beam lithography will be $a_{min} = 0.25-0.50$ μm.

(fragments) that are transferred simultaneously. To this end, however, the mask must be transparent to an electron (ion) beam. Sophisticated and expensive facilities and processes have been developed for this purpose, such as SCALPEL (the thickness of the transparent mask is no more than 100 nm). With such masks, today's ICs can still be fabricated; in the near term, however, they may become inappropriate.

Let us return to the lower part of Fig. 1. Lines c and d refer to the Meindl's forecast made in the 1980s at Stanford University, according to which electron lithography will provide $a_{min} = 0.25-0.50$ μm by 2010–2020. This prediction turned out to be understated: an even higher resolution has already been achieved with conventional lithographic tools using vacuum ultraviolet excimer lasers. For example, Intel Co. announced its 0.09-μm technology at the end of 2003. Such a resolution was achieved with standard 193-nm ArF excimer lasers, phase-contrast masks, and specially tailored resists. Moreover, it is expected that the resolution may be improved still further (down to 0.065 μm) even without resort to shorter wavelength (157-nm) F_2 lasers.

Owing to the wave nature of the processes and the possibility of parallel transfer of the IC entire pattern or its major part onto a semiconductor wafer, photolithography will be used until its potential is completely exhausted. Demand for nontrivial approaches will arise when a_{min} shrinks down to 0.07 μm or below. The lack of adequate transparent materials forces designers to try mirror optics. In this case, however, the numerical aperture of the objective lens reduces noticeably and, hence, the resolution drops.

CONCEPT AND DESIGN OF AN EXTREME ULTRAVIOLET EXPOSER FOR NANOLITHOGRAPHY

The aforesaid makes us turn to wave processes in the extreme ultraviolet (EUV) range ($10 \leq \lambda \leq 50$ nm). This range borders that of soft X rays, $\lambda < 10$ nm. It is in the EUV range that a great step forward in designing reflection optics and efficient radiation sources has been made in recent years. This progress has been achieved largely in the course of developing new-generation weaponry (specifically, X-ray lasers). The central point here is the creation of mirrors that efficiently reflect EUV radiation and soft X rays. To this end, a multilayer Bragg coating is applied on the atomically smooth surface of a massive substrate with a given curvature.

By way of illustration, Fig. 2 shows the typical wavelength dependence of the reflection coefficient for the Mo/Si multilayer coating, which proved to be one of the most efficient systems in this respect. On Mo/Si-coated Bragg mirrors, reflection coefficient R today approaches 70% at 13.4 nm (the theoretical value is $R_{max} = 74-78\%$). In the wavelength range 10–15 nm, a reflection coefficient as high as $R \geq 60\%$ has been achieved; however, the fundamental (absolute) maximum in this range lies near 13.4 nm [1, 2]. That is why main efforts are concentrated on using just this wavelength.

Figure 3 demonstrates the speculative dependences of the maximal resolution on the photon wavelength or energy. Obviously, the EUV range, as well as the longest wave part of soft X rays, is the most promising for
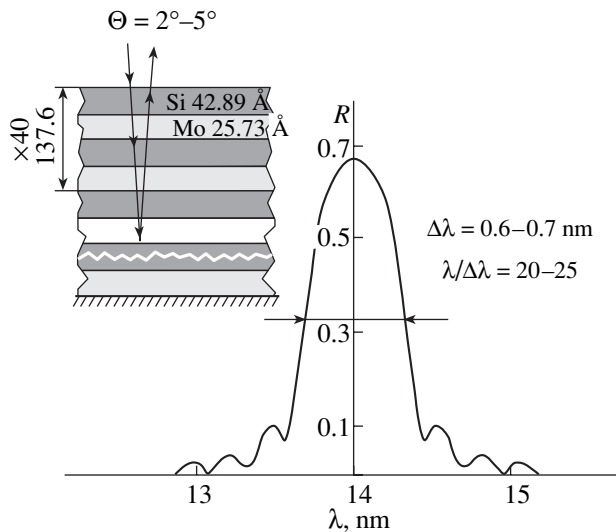
Fig. 2. Typical reflection spectrum of a Mo/Si-coated Bragg mirror.

Fig. 3. Estimated dependence of minimal linewidth a_{min} on the actinic radiation wavelength or photon (electron) energy for different lithography methods. Contact photolithography up to the X-ray range (proximity X-ray lithography); here, $a_{min} = (2\lambda\delta)^{1/2}$ (1, gap $\delta = 10$ μm; 1', "the zero gap"). Projection photolithography for NA varying from (2') 1.0 to (2) 0.6; here, $a_{min} = k\lambda/NA$. Curves 3 and 3' show the limitations due to image-smearing photoelectrons at different material densities (curve 3' corresponds to a higher density). Curve 4 shows the limitation due to backscattered electrons (for electron-beam lithography). (I) Infrared, (II) ultraviolet, (III) vacuum ultraviolet, (IV) far and extreme vacuum ultraviolet, (V) soft X rays, and (VI) hard X rays.

fine-line lithography. Straight lines 1 and 1' refer to contact lithography including proximity X-ray lithography. Here, $a_{min} = (2\lambda\delta)^{1/2}$. Continuous line 1 refers to $\delta = 10$ μm; line 1', to the "zero gap." The family of lines 2 refers to projection lithography for NA varying from 1.0 to 0.6. Here, $a_{min} = k\lambda/NA$. Curves 3 and 3' show the restrictions due to the generation of photoelectrons, which smear the pattern, at different material densities. Finally, straight line 4 is the restriction due to backscattered electrons (in electron-beam lithography). Thus, only EUV or soft X-ray lithography with reduction Bragg optics may radically overcome the nanolithography difficulties; however, losses due to incomplete reflection may be minimized (within 30–40%) only in the interval 13.4 ± 2.0 nm.

The leading tool developers and research organizations worldwide are now concentrating on design solutions for EUV exposers that are capable of patterning ULSI microprocessor slices of density 10^8–10^{10} per chip. The basic idea here is fragmentwise patterning by means of a projection reflection optical system using multilayer Bragg coatings. This is a fundamentally new IC pattern generator (having no analogues in conventional microtechnology), which exploits the step-and-repeat approach and is essentially an EUV projection wafer stepper. Here, EUV radiation is emitted by a plasma that is generated by a high-power pulsed laser focused on a target. The related optics and mask are fabricated in the same way as reflection X-ray optics with multilayer Bragg coatings. The coatings are applied on atomically smooth substrates, which are flat (EUV mask) or have a desired curvature (objective lens and condenser). The IC layout demagnified M times (M is the demagnification of the objective) is patterned in the absorbing layer by the standard methods of photolithography or focused electron-beam lithography (since
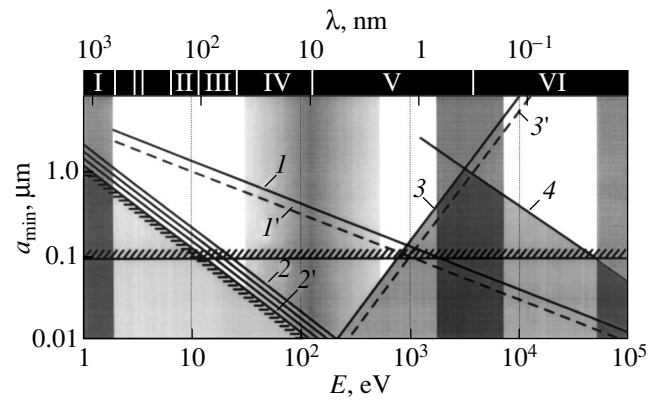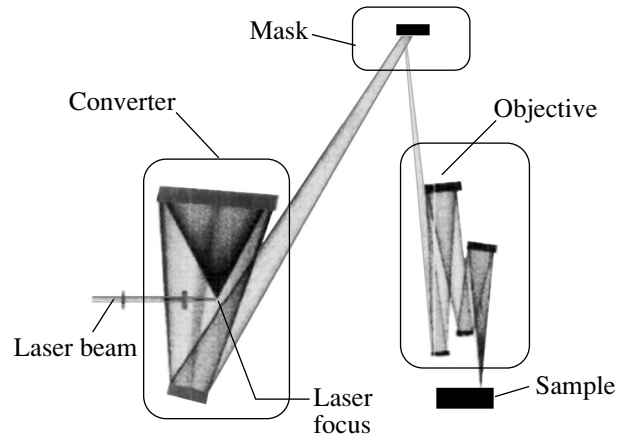

Fig. 4. Schematic of an EUV nanoexposer with a laser-plasma source of radiation.

the mask is used many times, the time of its formation is not a critical point here). The surface of the wafer to which the IC pattern is transferred in steps is covered by a special film (EUV resist). An EUV exposer is schematically shown in Fig. 4. Basically, it consists of four units: (i) an EUV source illuminating the mask, (ii) a mask assembly carrying the enlarged layout of an IC layer, (iii) a patterning optical system, and (iv) a sample assembly (the sample is coated by an EUV resist).

Significantly, the development of EUV facilities should be additionally supported by a number of advanced techniques, of which dose control in the vacuum ultraviolet, EUV, and soft X-ray ranges; spectral

control of radiation and materials; and mask technology are the most important. Relevant research is now under way.

EUV SOURCE FOR NANOLITHOGRAPHY

Synchrotron radiation, which was used in early studies by American and Japanese authors, is certainly a very efficient EUV source. It is very convenient for extensive research in this area. However, bulky and expensive storage rings may strongly retard the introduction of synchrotron radiation into the semiconductor industry. In addition, the specialists involved in the American project believe that application of synchrotrons may monopolize the semiconductor industry, which is undesirable and should be avoided [3]. In their opinion, the space occupied by an exposer for nanolithography (hereafter, nanoexposer) should not exceed $4 \times 4 \times 4$ m and its price must be moderate (no higher than \$10 M) so that any medium lithography user may buy it (this statement can hardly be open to argument). Therefore, most early works considered the laser plasma as an alternative EUV source. The plasma generated by a pulsed laser radiation striking solid, liquid, or gaseous objects is an effective source of EUV and soft X-ray radiation.

However, the industrial application of the nanoexposer poses the problem of high throughput, which is difficult to combat. For example, the throughput achieved on an ASML TWINSCAN AT : 1200B wafer scanner/stepper, which offers a resolution of 0.08 μm, is 103 circuit layers per hour on 300-mm silicon wafers at 109 exposures of 26×33-mm chips [4]. The power level providing such a throughput is unattainable with laser-plasma-based EUV sources currently available. This has given impetus to a search for new approaches including those based on the well-known methods of generating dense high-temperature discharge plasmas. Some of the new processes have already been tested as means for carrying out a controlled thermonuclear reaction but failed in producing ultrahigh temperatures. Among them are the Z (linear) pinch, Θ pinch, plasma focus, capillary discharge, vacuum arc, and their combinations (including combinations with a laser plasma). Also, much attention is now given to implementing an X-ray laser, which may lase directly at 13.4 nm (the wavelength of interest in the EUV).

Nevertheless, the laser plasma remains a convenient and efficient EUV radiation source. The source as such here is a rapidly (with a rate of $\sim 10^6$ cm/s) expanding (up to 20-fold) ionized cloud of the target material plasma heated to temperatures of $(0.12–1.20) \times 10^6$ K (or to 10–100 eV). A quasi-point source of radiation is very convenient for developing and optimizing the optics used. Initially, plasma of heavy metals was tried, since they offer a high conversion efficiency (about 1% per 2π sr in the nanoexposer "operating" range 13.4 nm $\pm$ 2% nm). Emphasis was on the continuous

part of the radiation spectrum. In this case, however, a plethora of debris contaminating the optical system and quickly making it inoperative has presented an impenetrable obstacle. Mask contamination, resulting in large-scale reproduction of faulty chips, is particularly dangerous. Many expedients to protect the optics have failed. Then, researchers turned to supersonic gas jets of a heavy inert gas, xenon. The laser plasma of xenon was used in radiation sources intended for the developmental EUV exposers produced by American and European companies. Basically, application of an inert gas as a target combats the problem of contamination. In the radiation source, the collecting mirror, being in the immediate vicinity of the target of the converter, runs into danger to the greatest extent. Only this component exhibits corrosion traces due to bombardment by high-energy ions and atoms of the inert gas (xenon jet), which cuts the service time of the mirror. More significantly, contamination is prevented at the expense of conversion efficiency η, which drops down to several tenths of a percent and is then difficult to raise to reasonable values. The value $\eta = 0.7\%$ is considered as ultimate for Xe gas (the maximal value achieved to date is $\eta = 0.55\%$ [5]). A reason for such a low value is the low density of the gaseous target. In an attempt to raise the conversion efficiency, the laser beam has to be directed to the near-nozzle region, where the density is the highest. In this case, one runs into the risk of nozzle damage and contamination by nozzle material particles. The way out might be to form Xe clusters or apply liquid or solid xenon. Recent experiments using the above approach have demonstrated that relatively high conversion efficiencies up to $\eta = 1.5\%$ are feasible.

The theoretical studies and spectroscopic analyses performed up to the present time have made it possible to establish an extremely important fact. Namely, EUV photons in the range of interest come largely from a 10-fold ionized xenon ion Xe^{10+}, and the emission here is due to the discrete absorption lines that are attributed to a series of transitions involving the fourth and fifth atomic shells: $4p^6 4d^8 \longrightarrow 4p^5 4d^9 + 4d^7(4f + 5p)$. However, the radiation at the necessary wavelength 13.4 nm is accounted for by the weak transitions $4p^6 4d^8 \longrightarrow 5p 4d^7$ and is much weaker than the radiation at 11 nm, which is accounted for by other transitions within the fourth shell ($n = 4 \longrightarrow n^* = 4$ [6]) (Fig. 5). An increase in the excitation intensity shifts the maximum toward shorter wavelengths rather than enhances the emission. As follows from calculations, the laser power density exciting the line at 13.4 nm must not exceed 10^{10}–10^{12} W/cm^2 and the maximal electron temperature must be no higher than 55 eV [6].

Moreover, the conversion efficiency may be raised only if the process of plasma generation is separated from heating of the plasma to a desired temperature. This may be done using a sequence of two pulses several nanoseconds distant from each other [7].

Thus, Xe cannot be viewed as an appropriate material compatible with mirror optics, which offers the highest reflection coefficient (i.e., the reflection at 13.4 nm). Analysis shows that the material of choice here is that with an atomic number of 50 (Sn), which provides the highest conversion efficiency at 13.4 nm. As the atomic number increases (Sb, I), the wavelength shifts to the shorter wave range and vice versa (In, Cd) (Fig. 6). Furthermore, Li, F, Sc, and O have also been shown to be of interest as far as emission at 13.4 nm is concerned. When highly ionized, these elements exhibit appropriate lines of the discrete spectrum. Tin as a medium generating a plasma that emits in the EUV range has recently attracted considerable attention. Even preliminary experiments turned out to be encouraging. A conversion efficiency of about 2% has been reached, which is predicted to rise to 3% or even higher. Using the Z pinch technique and its modifications, one can provide an optical power on the order of 10 W at the intermediate focus on the mask [5, 8]. One more advantage of the discharge-type plasma source is the absence of a high-power plasma-generating laser beam. However, in this case, one again faces the problem of contamination of the optical path, which must be necessarily solved.

In most recent works, a diode-pumped solid-state laser is used as a primary source. This laser offers a number of advantages, among which are operating stability, a high quality of the beam, and a high overall efficiency. An array of semiconductor lasing diodes used for pumping makes it possible to attain a high optical output (from 5 to 20 kW) in the pulsed mode at a pulse repetition rate of 0.5–15 kHz. A high wavelength (1.06 μm) is more likely to be an advantage, since it minimizes the optimal density of the plasma and reduces the contamination level [5, 6]. An alternative might be excimer lasers, e.g., 248-nm KrF or 308-nm KrCl devices, in which high peak outputs are easier to attain. However, reliable operation at high average outputs in the quasi-stationary mode seems to be problematic here. In both cases, the primary source is of sophisticated design and is made using leading-edge technologies, since, eventually, it governs the throughput of the nanoexposer. Figure 7 presents the generalized scheme of energy losses in the nanoexposer. Along with a low conversion coefficient, there are many other factors causing considerable energy losses (see also Tables 1 and 2). Losses due to multiple reflection stand out here. In the case of an advanced eight-mirror objective lens, they may increase 300 times. For a standard resist sensitivity of 5 mJ/cm^2, the energy of spectral-pure EUV radiation that is necessary for today's throughput requirements to be met varies between 10 and 130 W depending on the number of mirrors in the imaging objective lens. A spectral filter used in the system has a transmission coefficient of no higher than 50%. Even at its exit part, an eight-mirror system will consume electric power of about 2.0 MW, which poses the problem of heat removal from the source units.
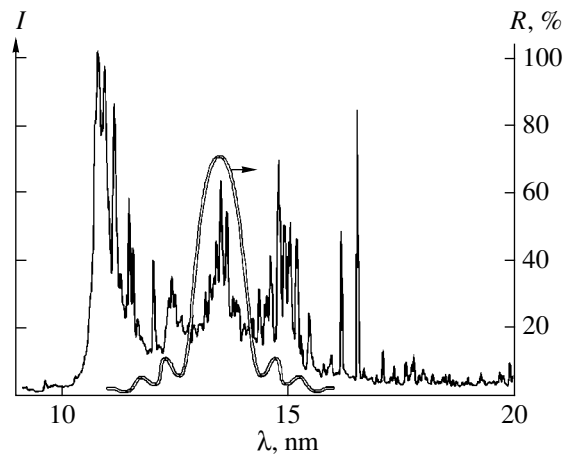
Fig. 5. Radiation spectrum of 10-fold ionized Xe and the reflection spectrum of a typical Mo/Si-coated Bragg mirror [6].
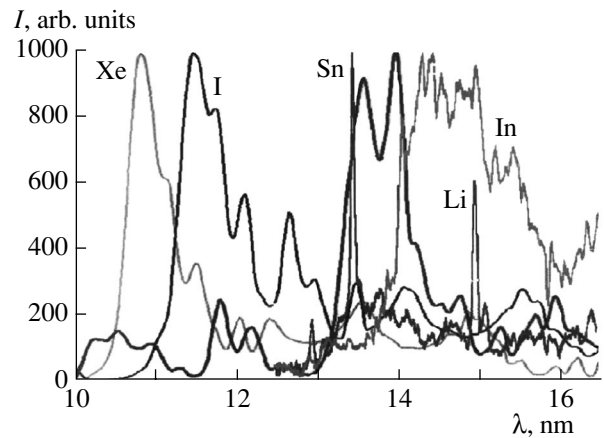

Fig. 6. Radiation spectra for multiply ionized Xe, In, Sn, I, and Li [6].

OPTICAL SYSTEM

The nanoexposer has a number of special features associated with the mirror optics in general and with Bragg reflection coatings in particular.

(1) The mirror optics is free of dispersing elements; therefore, the ray path is wavelength-independent and the system may be tested with electromagnetic radiation of any range provided that the radiation is intensely reflected from the surface of the mirrors.

(2) The mirror optics cannot have a numerical aperture close to unity. To this end, the last mirror but one must be placed exactly at the focus of the last (patterning) mirror, i.e., in the image field. The numerical aperture is maximal if the image field is minimized.

(3) Bragg coatings are, in essence, interference filters for wavelength λ. For coatings used in the system, the wavelengths of reflection maxima must fall into diffraction zone $\Delta\lambda$. Losses are considered reasonable if spread $\delta\lambda_{max}$ ($<\Delta\lambda$) is such that $\delta\lambda_{max}/\lambda \leq 2\%$.
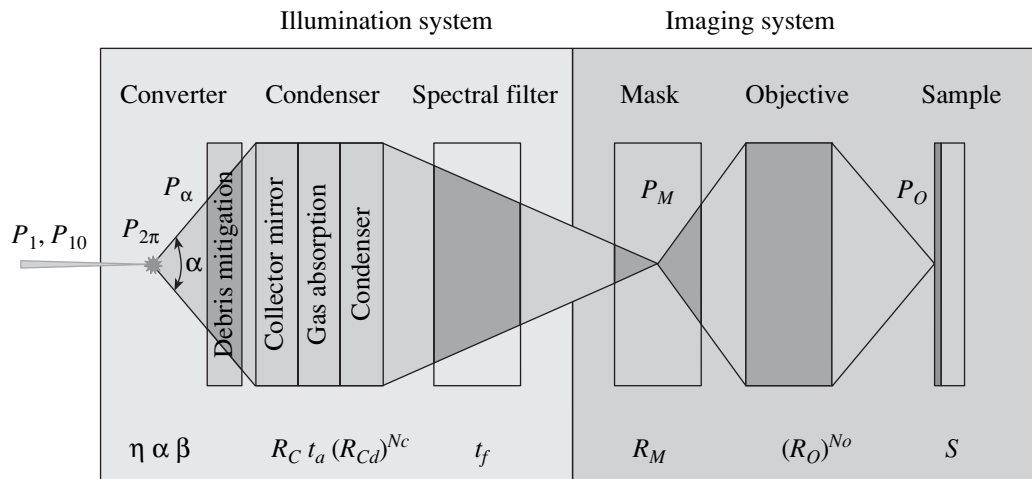
Fig. 7. Energy losses in the nanoexposer.

(4) The parameters of Bragg coatings are calculated for normal incidence. If the multilayer coating has the same period throughout the mirror, there exists allowable deviation $\Delta\Theta_{max}$ from $\pi/2$ for the angles of incidence and reflection. This additionally decreases the attainable numerical aperture of the patterning EUV objective.

The condenser is a separate optical unit of the nanoexposer. It is in direct contact with the converter and so must be isolated from the other optical elements, as well as equipped with high-capacity pumping facilities and means for removing conversion products. Thus, the design of the condenser strongly depends on the conversion method on the one hand. On the other

Table 1. Factors responsible for energy losses in an EUV nanoexposer

Conversion efficiency η, %	Collector efficiency α (angle of collection divided by 2π)	Transmission β of debris mitigation system	Collector mirror reflectivity R_c	Reflectivity R of other mirrors	Residual gas transmission t_r	Spectral filter transmission t_f	Mask reflectivity R_m	Mirror (M) objective transmissivity T_0	
0.55	0.2–0.3					0.5	0.65	2 M	0.4225
0.70		1.0	0.5	0.65–0.86	0.9			4 M	0.1785
1.5									
0.7			Condenser losses					6 M	0.0754
2.2		0.9			1.0			8 M	0.03186
3.0			0.7						

Table 2. Electrical and optical power levels in different units of an EUV nanoexposer with a throughput corresponding to an advanced ASML TWINSCAN AT : 1200B scanner/stepper

Resist sensitivity S, mJ/cm^2	Power P needed for exposing 100 300-mm wafers per hour, W	Total radiation power P_m on mask surface, W		Collectable EUV power P_α, W	Total EUV radiation power $P_{2\pi}$, W	Total radiation power P_{10} of primary source** , kW	Total electrical power P_1^{***}, kW
5	2.68*	2 M	9.8	77	310	15	150
		4 M	23	183	734	37	370
		6 M	55	434	1736	87	870
		8 M	130	1027	4110	205	2050

* Die-to-die exposure time losses are 50%, the power values are increased 10 times because of time losses due to scanning.
** Conversion efficiency $\eta = 2\%$.
*** Laser efficiency $\eta_l = 10\%$.

hand, the optical path of the condenser must be built on the same Bragg mirrors as the remaining part of the optical system if a laser-plasma converter is employed (the central wavelengths of the reflection bands must be very close to each other) and is an integral part of the entire optical scheme, including the imaging objective and the mask. Therefore, the condenser, mask, and objective should be optimized as a whole. As was noted above, Bragg mirrors narrow interval $\delta\lambda$ into which the reflection maxima of all the mirrors must fall. Since $\delta\lambda_{max}/\lambda \approx 2\%$, this interval is roughly 0.27 nm for $\lambda = 13.4$ nm. Note that this value is not the FWHM of a spectral line: only that part of the line is taken into account where the reflection coefficient is fairly high. Particular emphasis should be placed on the first (collecting) mirror, which is in the immediate vicinity of the conversion point and, thus, degrades most severely during operation. The challenge is to extend the service time of this mirror to 10 000 h of continuous operation with a decrease in reflection coefficient R by no more than 10% [5] (in light of this, R of the condenser mirror was taken to be relatively small, $R = 0.5$, in calculation of Table 2). At the same time, the closer the collecting mirror to the converter and the larger its radius, the higher the solid angle collecting the EUV radiation and the higher the efficiency of the source. Typically, the collecting solid angle equals 1.8 sr, corresponding to 28.6% of the omnidirectional radiation (i.e., covering 2π starting from the conversion point) [5]. Other designs of the condenser, i.e., those based on grazing incidence mirrors, are applied in discharge sources of the plasma. A radically new design of the condenser may be related to an X-ray laser with an appropriate wavelength.

Since energy losses due to incomplete reflection of the EUV in Bragg mirrors are considerable, their number should be minimized in both the condenser and the imaging objective. A two-mirror objective consisting of a large concave mirror and small convex mirror (the so-called Schwarzschild objective) has found application in X-ray astronomy. For lithography purposes, this objective was first used in the American nanolithography project and has numerical aperture $NA = 0.06–0.09$ [9]. Such a low aperture is the inevitable cost for a relatively large image field. The real image has the form of a narrow (less than 1 mm in width) circular arc. To fully expose (transfer) an IC topological layer, it is necessary to scan within a die over the semiconductor wafer surface; so, such an exposer represents a "scanning stepper." The same idea was employed in subsequent (refined) versions.

Clearly, the two-mirror Schwarzschild objective (Fig. 8a) is inadequate in terms of resolution (see the Rayleigh criterion) and/or image quality. In this case, the Rayleigh optical resolution could not overcome the barrier $a_{min} = 90–100$ nm. It should be noted, however, that the Rayleigh condition alone, which characterizes the properties of the image optical field, cannot provide
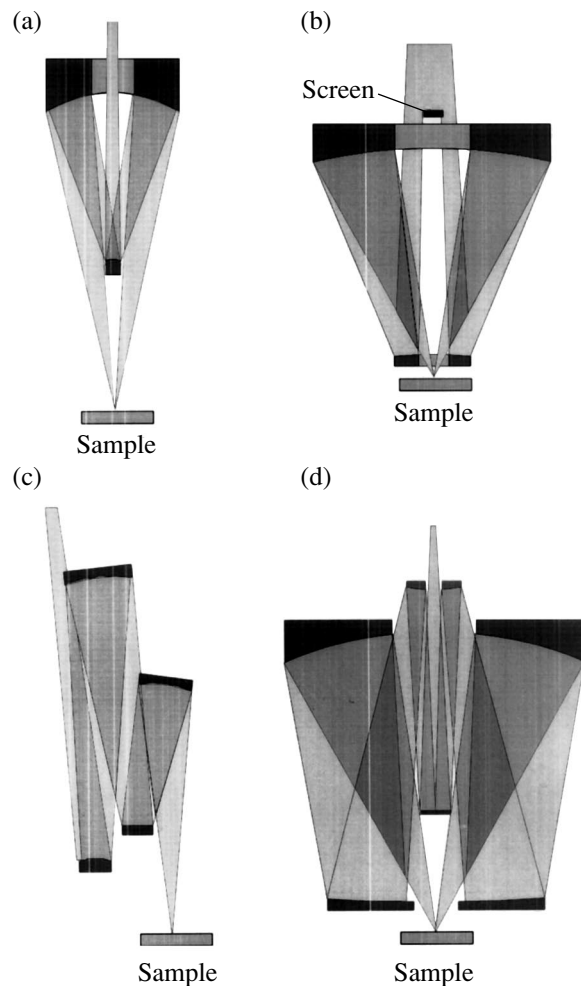

Fig. 8. Optical schemes of Bragg mirror objectives: (a) low-aperture two-mirror Schwarzschild objective, (b) two-mirror high-aperture objective [10, 11], (c) commercial four-mirror low-aperture objective, and (d) four-mirror high-aperture objective with a small image field [14].

a high-quality pattern unless special measures are taken. For example, in the absence of phase-shift masks and/or high-contrast resists, only a feature size equal to two or three Rayleigh "lengths" a_{min} can be reliably reproduced. If the image field size is of minor importance, the numerical aperture can be greatly increased by using the modified Schwarzschild objective. Two-mirror Schwarzschild objectives with $NA = 0.3$ (ASET, Zeiss) are currently available [10, 11], and a Schwarzschild objective with $NA = 0.36$ (Fig. 8b) has been developed and launched in production. However, here the image field does not exceed 1 mm^2. Four-mirror systems (Fig. 8c) might basically remedy the situation: they offer an improved NA and still retain a considerable image field in the scanning stepper mode. In the commercial versions of four-mirror objectives, however, NA is no higher than 0.10–0.14 [11, 12]. A unique four-mirror objective (Fig. 8d) has recently been developed that offers NA as high as 0.485 at the expense of
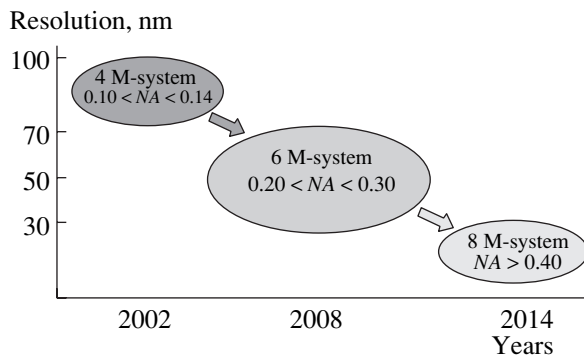
Resolution, nm

Fig. 9. Forecast of minimal resolution a_{min} attainable with mirror objectives vs. the number of mirrors (M) and numerical apertures [13].

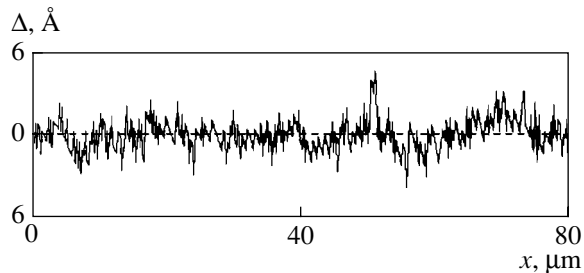**Fig. 10.** Profilogram of the atomically smooth surface with an rms roughness of 0.1 nm.

the image field size (0.5×0.5 mm). With such an objective, linewidth $a_{min} = 10$ nm can be reached owing solely to the image field quality without using tailored masks and resists. A considerably higher *NA* is attained in commercial six-mirror systems, and eight-mirror systems may provide $NA \approx 0.4$ (Fig. 9) even in commercial versions of the scanning stepper (Zeiss specialists' prediction). Thus, the resolution of a commercial nanoexposer based on an eight-mirror objective, which supposedly will come onto the market by 2014, may reach 20–30 nm [13]. However, energy losses due to six- or eightfold reflection become critical here at $R < 1$.

In view of the aforesaid, one cannot ignore image correction techniques. For example, simple optics may be equipped with a system of holographic correction. The quality of mirrors, specifically, the maximal reflection coefficient achieved with a multilayer Bragg coating, directly depends on the surface roughness of a substrate to be covered by the layers. Figure 10 shows the surface roughness measurements on simple test mirrors made of Sitall (devitrified glass such as Pyroceram). Similar results were obtained on quartz glass and single-crystal silicon. It follows from Fig. 10 that there exist techniques providing an rms roughness of 0.10–0.33 nm. Thus, an atomically smooth surface becomes a reality! Moreover, the distortion of the aspherical shape of patterning mirrors may be minimized down to 0.05λ or less (at $\lambda = 13.4$ nm).

EUV MASK

Essentially, EUV masks are similar to any other Bragg mirror of the optical system. The only difference is that the mask is flat and incorporates a layer (W, Ta, TaN, Cr, or other materials) efficiently absorbing in the EUV range. On this extra layer, the *M* times magnified pattern of a ULSI topological layer is formed by one of the conventional methods (e.g., submicron photolithography or focused electron-beam lithography). As nanolithography matures as a field of technology, production of EUV masks becomes an independent chore. One may expect the emergence of EUV mask suppliers in the immediate future. The most feasible substrate for the masks is a single-crystal silicon wafer, which is classically used in the IC production. It is not improbable, however, that quartz or Sitall will turn out to be more appropriate for preparing atomically smooth surfaces.

EUV RESIST

The basic problem associated with the EUV resist is that all natural materials highly absorb in the EUV range. The EUV absorption depth of standard organic photoresists does not exceed 100 nm. It seems most likely that EUV resists will form a desired pattern in a thin surface layer. Today, single-layer silylated, as well as two-layer and three-layer, compositions are being studied most extensively. A resist intended for large-scale production must offer both a high contrast (in image transfer) and a high sensitivity, so as to provide an acceptable production rate. A sensitivity of 5–10 mJ/cm² is considered as meeting the production requirements most adequately, since this value is a tradeoff between a production rate and a level of fluctuations due to photonic shot noise [14].

The EUV resist must have not only good optical properties but also a high chemical performance. For fine-line resists, line-edge roughness becomes a critical issue irrespective of the lithography method used. The EUV resist must be able to combat this problem as well. When the feature profiles in real resists are simulated, the illumination nonuniformity and resist–developer interaction are classically included in the following differential equation for the linewidth:

$$dy/dx = (\partial y/\partial E)(\partial E/\partial x),$$

where E is the absorbed energy.

For a weakly absorbing resist and weakly reflecting substrate, the first factor depends on the properties of the resist–developer pair and the second only on the performance of the optical system. Then, the first factor is directly proportional to contrast γ and inversely proportional to energy exposure E_0 sufficing to transfer the image (to develop the resist); i.e., $\partial y/\partial E \propto \gamma/E_0$. Next, taking into account that the contrast depends on the inverse optical density of the resist material $\gamma \propto (\beta + \alpha d)^{-1}$, where β is a constant, α is the absorption coeffi-
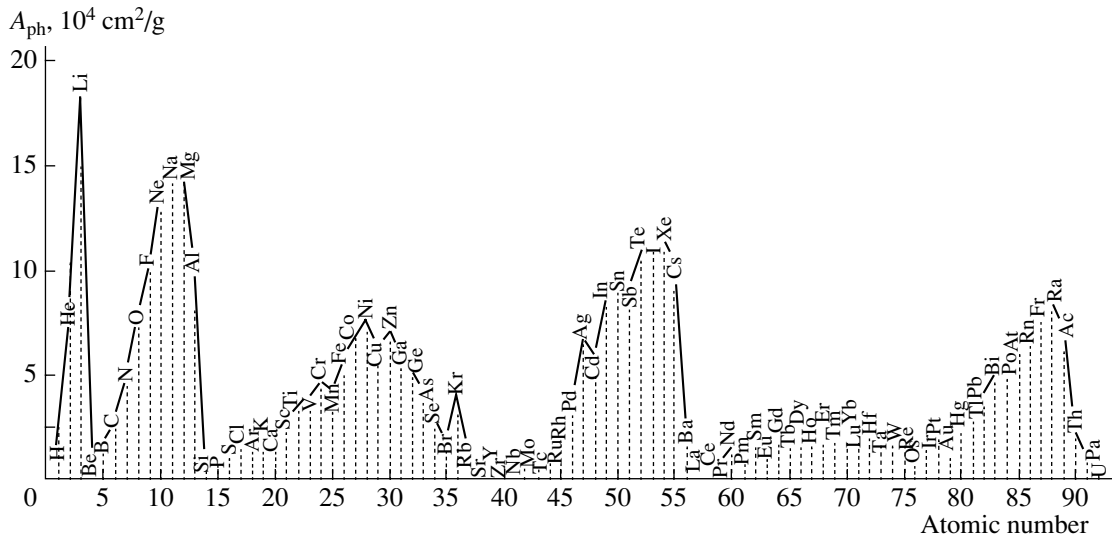
Fig. 11. Photoabsorption cross section A_{ph} vs. the atomic number for different elements.

cient, and d is the resist thickness), we find, in a first approximation, the functional dependence of the feature profile on the parameters of the optical system and resist,

$$dy/dx \propto [NA/(\lambda(\beta + \alpha d)E_0)][1 - k(\Delta z(NA)^2/\lambda)]^2.$$

Here, Δz is the offset from the center in the image plane and k is a coefficient close to unity. From this expression, it follows that high-contrast resists with low optical density αd provide a shaper feature image upon developing. A decrease in the optical density through thinning the resist is possible down to a certain limiting thickness at which the resist film is still free of defects and composition and thickness fluctuations are negligible. Also, the limiting (least) thickness depends on the resistance of unexposed areas to the developer. Therefore, searching for resist materials with an acceptable sensitivity and low absorptivity remains to be a topical issue [15].

Under our conditions (13.4 nm, 92.5 eV), the photoabsorption cross section is determined mostly by inner atomic shells; in other words, the absorption of the resist material depends only on its stoichiometric composition and is almost independent of the chemical bond structure. Figure 11 plots the calculated photoabsorption cross section against the atomic number for the wavelength 13.4 nm.

Conventional organic resists consist largely of carbon, hydrogen, and oxygen but may also contain nitrogen, fluorine, and other elements. Calculations show that fluorinated polymers cannot be used as photoresists because of their high absorptivity. It has been found that aromatic hydrocarbon radicals incorporated into polymers considerably decrease the absorption coefficient at $\lambda = 13.4$ nm. In addition, these radicals may greatly improve the chemical resistance of a material to a developer and, thereby, favor the use of single-layer EUV resists.

Silane polymers have the lowest linear absorption coefficient and are, thus, the most promising for EUV lithography. Siloxane polymers also have a low absorption coefficient but mostly because of their low density. The advantages due to the low absorption of silicon are inconspicuous when siloxane polymers are used [15].

When introduced into polymers, radiation-sensitive side groups increasing the radiation yield of reactions of polymer destruction under exposure may raise the sensitivity to several mJ/cm^2. However, the contrast of the resists thus prepared drops to unity. Note that the sensitivity and the contrast are in inverse proportion to each other [16]. Important features of organic EUV resists are their linearity and also the fact that the energy exposure is independent of the exposing radiation intensity. The intensity and the exposure time are "interchangeable" here: only the integral dose taken up by the resist is of significance. Therefore, a low radiation intensity can be compensated for by a large number of radiation pulses; hence, operation at a high repetition rate is preferable.

The left side of Fig. 12 shows 2004–2005 predictions for NA and coefficient k_1 in the Rayleigh criterion, together with the achievements in this area [17] during the evolution of the UV scanner/stepper technology. The natural trend toward making NA as close to unity as possible (in 2003, the best objective had $NA \approx 0.85$) and reducing k_1 as much as possible is obvious. By using phase-shifting masks, high-contrast photoresists, etc., it was made possible to considerably reduce the value of k_1 (down to ≈ 0.425 or lower). With these means, a design rule of 0.09 μm was introduced into practice using an ArF laser wavelength of 193 nm (i.e., $a_{min} = 0.5\lambda$). Unfortunately, neither phase-shifting
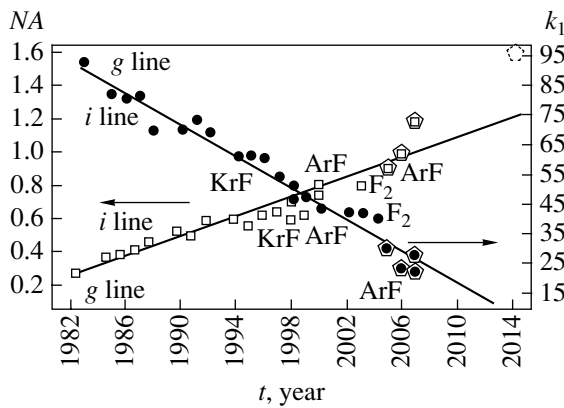
Fig. 12. Tendency toward an increase in numerical aperture *NA* of transmission objectives and for a decrease in coefficient k_1 in the formula for minimal linewidth a_{min}. On the left (up to year 2004), conventional photolithography [17]; on the right, predictions (pentagons) relying on the application of immersion liquids [20]. *g* line and *i* line are the radiation wavelengths 436 and 365 nm, respectively, of a mercury arc lamp; KrF, ArF, and F_2 stand for the wavelengths 248, 193, and 157 nm of the related excimer lasers.
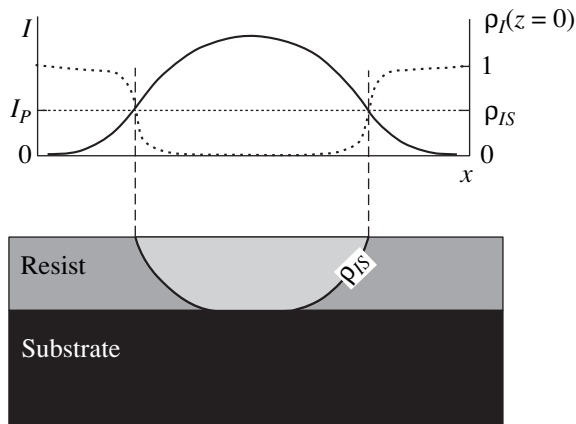

Fig. 13. Contrast enhancement in a photoresist imaging features of size finer than that following from the Rayleigh criterion.

masks nor high-contrast resists appropriate for the EUV range are available.

Thus, development of efficient high-contrast EUV resists becomes a key issue. Tackling this problem is really of great importance, since objectives with Bragg coatings inherently cannot possess a high numerical aperture, as was noted above. It was shown experimentally that a number of thin inorganic films exhibit a giant contrast in image transfer, which shows up under specific irradiation conditions [18]. Based on this effect, one could compensate for resolution losses due to a low numerical aperture of the optics. This effect can be phenomenologically described with three equations (akin to the well-known Dill equations) that take into account the nonlinear character of photochemical interaction and a change in the optical constants at the actinic radiation front [19]. Central here is the consid-

eration of a phenomenological function accounting for nonlinear processes in photochemical interaction. The best image quality is obtained if a certain relationship between the resist material parameters and exposure parameters is fulfilled, as follows from analysis. With this condition met, the edge profile of a feature developed (transferred) becomes much sharper relative to the light field (the contrast enhancement effect, Fig. 13). Experiments also show that, along with a high contrast, a high optical sensitivity reaching 3 cm^3/kJ at a threshold radiation flux of less than 17 kJ/(cm^2 s) (AsSe) may be attained. The related energy exposure needed for full development does not exceed 2–3 mJ/cm^2 for a film less than 100 nm thick. However, this photoresist is nonlinear and, therefore, the radiation intensity and the exposure time are not interchangeable: the energy of each of the pulses applied must exceed some threshold energy. As follows from simulation, the threshold energy density for AsSe equals 0.34 mJ/cm^2 (at a pulse duration of 20 ns). Basically, other EUV resists may be found. Clearly, application of nonlinear inorganic resists requires that the basic units of the nanoexposer be modified. The number of mirrors should be minimized, and, possibly, special correction schemes will be necessary. Furthermore, the peak (rather than the average output) power density exceeding a threshold value of 10–20 kW/cm^2 will be the critical parameter of the source. The exposure dose necessary for development can be accumulated by repeatedly exposing the resist to higher-than-threshold EUV pulses. Eventually, an extremely high resolution (higher than the resolution of a similar system based on linear resists) may be reached. Note, however, that this goal is easier to accomplish on relatively small image areas and may require *x–y* (two-dimensional) scanning in IC production. It should be emphasized that a resolution of 10–20 nm makes it possible to reproducibly fabricate regular nanostructures with quantum dots and wires, as well as mesoscopic devices.

CONCLUSIONS

The fabrication of a commercial nanoexposer requires that the physicotechnical problems mentioned above and also a number of problems of fine mechanics be solved. Among the latter are the design of the sample (and, perhaps, mask) assembly, which must incorporate a precision table reliably providing a nanometer positioning accuracy at an overall travel of 300–450 mm. A nanometer accuracy of focusing and alignment of all topological layers on the sample has also to be provided. A more routine yet labor-consuming problem is designing the nanoexposer as a whole (particularly, its body) and its mechanical part. Optical adjustment and focusing of the exposer in the EUV range is also a challenge. Significantly, the nanoexposer should be protected against dust and vibration and configured with differential pumping facilities so as to provide the ultrafine adjustment of the optical system. Discussion

of these problems is beyond the scope of this review. Note only would solving them manifest creating a very precise contemporary instrument, which would take advantage of many technology breakthroughs. A considerable barrier to the development of a commercial nanoexposer is the energy problems of the radiation source, which were mentioned above.

However, a much more serious circumstance that may significantly postpone the entrance of the nanoexposer to the market is the recent idea of immersion as applied to a standard scanner/stepper based on a 193-nm ArF excimer laser [20]. It has turned out that water (with refracting index $n = 1.43$) may be injected into the space between the imaging objective and semiconductor wafer virtually without adversely affecting the image quality. Because of this, the operating wavelength can be decreased n-fold, $193/n = 135$ nm. The numerical aperture of the objective increases n times with a simultaneous increase in the depth of focus. This greatly extends the potentiality of conventional photolithography, essentially giving a chance to advance lithography into the EUV range. It is today difficult to estimate the actual capability of this new approach. The advertising statements of leading lithographic tool producers may be far from reality. However, it has become known that the Intel 0.065 μm process announced in 2006 will likely be based on this approach. Certainly, application of such a technology is much more cost-effective than the production of sophisticated EUV lithographic facilities. Other transparent immersion liquids with refractive indices somewhat higher than that of water may also be used. The predicted values of NA and k_1 are shown on the right of Fig. 12. According to the estimates (which may seem to be somewhat optimistic), a resolution as high as 22–35 nm may be reached by 2014 without radically redesigning the existing facilities. Thus, upgrading of present-day exposers based on ArF excimer lasers makes the development not only of exposers based on 157-nm F_2 lasers but even of EUV nanoexposers aimed at 20–65-nm design rules meaningless.

The elaboration of the idea of immersion nanolithography using excimer lasers leaves only a niche for extremely high resolution near 5–20 nm where EUV nanolithography still remains unrivaled. Such a high resolution can hardly be obtained without considerably cutting the actinic radiation wavelength, i.e., without resorting to the EUV at 13.4 nm. However, this goal seems unrealistic unless EUV nanolithography processes and equipment are refined. Along with an increase in the numerical aperture of the imaging objective to a maximum possible extent, this requires application of high-contrast nonlinear resists. Development of two- or four-mirror objectives provided with image correction schemes becomes a topical problem. Still more advanced tools for image improvement are now under development. As was noted above, a four-mirror objective (Fig. 8) makes it possible to pattern 10-nm-wide features (lines and spacings) with a reasonable contrast even in the light field. It is natural to assume that the only way to further shrink the linewidth (at least twice) is application of nonlinear resists. There is already no doubt that a high-resolution EUV nanoexposer will be created in the near future. However, wide application of such facilities must be preceded by designing ULSI circuits based on new-generation mesoscopic and quantum-size approaches. These new approaches are expected to displace the conventional technologies, such as insulated-gate MIS transistors, which have exhausted their potentialities nearly completely. Therefore, at least at the early stage, high throughput will not be a critical point in ultra-high-resolution EUV nanoexposers.

REFERENCES

1. D. Kim, D. Cha, and S. Lee, Jpn. J. Appl. Phys. **37**, 2728 (1998).
2. N. N. Salashenko, in *Proceedings of the All-Russia Meeting on X-ray Optics, Nizhni Novgorod, 1998*, pp. 53–57.
3. R. Stulen, R. R. Freeman, and T. E. Jewell, *EUV Lithography: A Talk at the International Seminar on EUVL Problems, FTI, St. Petersburg, 1994*.
4. http://www.asm1.com.
5. *Proceedings of the EUVL Source Workshop, Dallas, 2002.*
6. G. O'Sullivan and P. Dunne, in *Proceedings of the EUVL Source Workshop, Dallas, 2002.*
7. A. Andreev, J. Limpouch, and A. B. Iskakov, Phys. Rev. E **65**, 026403 (2002).
8. V. M. Borisov, A. Yu. Vinokhodov, Yu. B. Kiryukhin, *et al.*, in *Proceedings of the All-Russia Workshop on X-ray Optics, Nizhni Novgorod, 2003*, pp. 124–130.
9. D. A. Tichenor, G. D. Cubiak, M. E. Malinovski, R. H. Stulen, *et al.*, Proc. SPIE **2194**, 95 (1994).
10. Y. Horiike, in *Proceedings of the 5th ISTC Scientific Seminar "Nanotechnologies in Physics, Chemistry, and Biotechnology," St. Petersburg, 2002.*
11. U. Dinger, F. Eisert, H. Lasser, *et al.*, Proc. SPIE **4146**, 35 (2000).
12. K. Diefendorff, www.MDRonline.com.
13. W. Ulrich, S. Beiersdörfer, and H.-J. Mann, Proc. SPIE **4146**, 13 (2000).
14. J. E. Bjorkholm, Intel Technology Journal, Issue Q3'98, 1.
15. N. Matsuzawa, S. Irie, E. Yano, *et al.*, Proc. SPIE **4343**, 278 (2001).
16. S. A. Bulgakova, A. Ya. Lopatin, V. I. Luchin, *et al.*, in *Proceedings of the All-Russia Meeting on X-ray Optics, Nizhni Novgorod, 1998*, pp. 124–130.
17. G. Dao, A. Yen, and W. Trybula, Future Fab. Intl. **12** (2002).
18. E. G. Barash, A. Yu. Kabin, V. M. Lyubin, and R. P. Seisyan, Zh. Tekh. Fiz. **62** (3), 106 (1992) [Sov. Phys. Tech. Phys. **37**, 292 (1992)].
19. N. A. Kaliteevskaya and R. P. Seisyan, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 587 (2000) [Semiconductors **34**, 825 (2000)].
20. http://www.ixbt.com.

Translated by V. Isaakyan

THEORETICAL
AND MATHEMATICAL PHYSICS

Interference Bandwidth Compression of the Spectrum of Transition Radiation from a Modulated Electron Flux

Yu. K. Alekseev[†], V. V. Achkasov, D. S. Kulida, and A. P. Sukhorukov

Moscow State University, Vorob'evy Gory, Moscow, 119992 Russia

e-mail: kulida@nm.ru

Received May 26, 2004

Abstract—The possibility of converting the modulated electron flux energy to electromagnetic oscillations (transition radiation) when the flux collides with an unmoved metallic screen is considered. The interference of waves arising at different sites of the screen changes the spectral–angular distribution of the radiation energy density. These changes are analyzed with allowance for a time delay between radiations coming from different parts of the inclined bunch. A plausible mechanism behind the frequency shift of interference maxima, which is observed when the angle of inclination of the screen is changed, is discussed. © 2005 Pleiades Publishing, Inc.

About 60 years ago, Ginzburg and Frank showed that a charge uniformly moving with a slower-than-light velocity may radiate electromagnetic waves. This happens when the charge moves in a medium the characteristics of which vary in space and time [1, 2]. This transition radiation is a purely classical effect, which is completely described by the Maxwell equation.

In practice, the transition radiation is used largely in detectors of charged relativistic particles and high-energy beams. Specifically, it is applied for lepton identification in high-energy physical experiments [3]. Note also that this type of radiation was observed in X-ray tubes as anode "glow" well before its theoretical explanation. This effect was explained only after the theory of transition radiation had been developed.

The theory of transition radiation was elaborated upon in [4, 5] and other works. In particular, Anisimov and Levitskiĭ [6] formulated a theory of transition radiation and surface wave excitation for the case of normal incidence of a modulated electron beam on the boundary of a plasma.

The spectral and angular characteristics of the transition radiation from a periodic train of extended electron bunches differ considerably from those of the transition radiation from a single electron [7]. Interference redistribution of the transition radiation energy may result in its concentration within a narrow frequency and/or angular range; thus, the radiation efficiency at a selected frequency may increase drastically.

In this paper, we consider the transition radiation of an electron flux modulated in the field of a traveling electromagnetic wave (Fig. 1). Such a wave may be a microwave (submillimeter) beam in free space, an *N*th harmonic generated when microwave radiation passes through a nonlinear medium (at terahertz frequencies),

or a focused optical or UV laser beam. Wide (wider than the radiation wavelength) sheet electron beam *1* passes through focused linearly polarized electromagnetic beam *2*, in which the electric field vector is aligned with the beam particle velocity. When passing through the electromagnetic beam, the electrons become velocity-modulated and, continuing to move, group into bunches *3*. The bunches fall on stationary metallic screen *4* and generate transition radiation *5*. We obtained the frequency and energy characteristics of the radiation and its directional patterns. It was also found that the frequency of interference fringes can be controlled by rotating (inclining) the metallic screen.

A distinct feature of electron bunching in the system shown in Fig. 1 is that the flux is not filamentary in its cross section, since it is sufficiently wide. Accordingly,
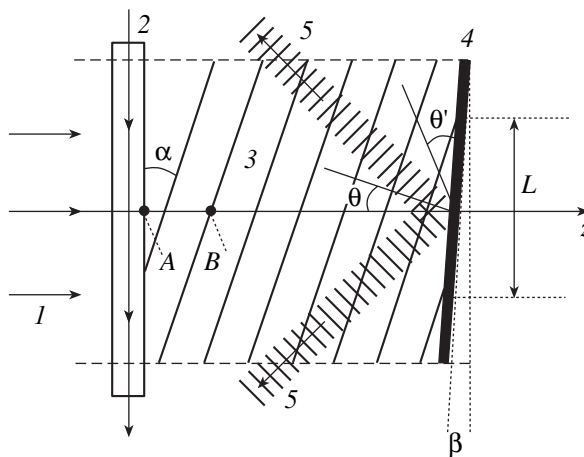

Fig. 1. Bunching of an electron flux and the inclined screen. *1*, electron flux; *2*, electromagnetic beam; *3*, electron bunches; *4*, metallic screen; and *5*, transition radiation of the electron bunches.

[†] Deceased.

the modulation of the electrons that are more distant from the radiation source will lag behind by the time taken for the light to travel the difference in distances. As a result, the electrons group into linear bunches making angle α with the propagation direction of the electromagnetic beam.

Let us calculate this angle. The longitudinal electric field of a monochromatic electromagnetic beam can be written in the form

$$E(x, y, z, t) = E_1(x, y, z)\sin\left(\frac{\omega_0}{c}x - \omega_0 t\right), \quad (1)$$

where E_1 is the field amplitude, ω_0 is the circular frequency, and t is the time.

Let (x_b, z_b) be the coordinates of bunch B (Fig. 1) in the plane $y = 0$ and t_b be the time of arrival of an electron at the bunch. We restrict our consideration to the two-dimensional problem, $\partial/\partial y = 0$. If $\Phi(x_b)$ is the phase of the electron exit velocity at a point with coordinate $x = x_b$ at time instant t_0, then

$$\Phi(x_b) = \frac{\omega_0}{c}x_b - \omega_0 t_0 + \psi_0, \quad (2)$$

where ψ_0 is the phase difference between the exit velocity and electromagnetic field at time t_0.

The velocity of electron A (with coordinates (x_b, z_0)) at the exit from the modulating beam at time t_0 can be written in the form

$$v(x_b, t_0) = v_0 + kE_1(x_b, 0, 0)\sin[\Phi(x_b)], \quad (3)$$

where v_0 is the velocity with which the electron flies into the beam and k is the velocity modulation efficiency of the beam, which depends on the focusing, amplitude, and mode composition of the beam.

Note that, in this work, we assume that the space of interaction, the distribution of the electromagnetic field, and the velocity of electron motion along the x axis are uniform; i.e., $\partial n/\partial x = 0$, $\partial E_1(x, y, z)/\partial x = 0$, and $\partial v_0/\partial x = 0$. Hence, phase shift ψ_0 is also independent of this coordinate.

The time the particle takes to travel the distance from the end of the modulator (z_0) to the bunch (z_b) is

$$t_b - t_0 = \frac{z_b - z_0}{v_0 + kE_1\sin[\Phi(x_b)]}. \quad (4)$$

Using the law of conservation of charge, $I_0 dt_0 = I_b dt_b$, we determine the current of the bunch,

$$I_b = \frac{I_0}{dt_b/dt_0}$$

$$= \frac{I_0}{1 - \dfrac{(z_b - z_0)(-kE_1\omega_0)\cos[\Phi(x_b)]}{\{v_0 + kE_1\sin[\Phi(x_b)]\}^2}}, \quad (5)$$

and then its coordinate.

Here, I_0 is the convective current density of the electron flux flying into the bunch and I_b is the current density in the bunch in cross section z_b. Assuming that the modulation amplitude is small ($kE_1 \ll v_0$), we can write

$$I_b \approx \frac{I_0}{1 + \dfrac{kE_1\omega_0}{v_0}(z_b - z_0)\cos[\Phi(x_b)]}. \quad (6)$$

Let us take the maximal current of the flux to find the coordinate of the bunch. The condition $I_b = I_{b,\,max}$ is fulfilled if $\cos[\Phi(x_b)] = -1$, i.e., if $\Phi(x_b) = (2n + 1)\pi$, where $n = 0, \pm1, \pm2, \ldots$. Then,

$$t_0 = \frac{x_b}{c} - (2n + 1)\frac{\pi}{\omega_0} + \frac{\psi_0}{\omega_0}.$$

Substituting this expression into (4) yields

$$z_b = -\frac{v_0}{c}x_b + z_0 + v_0 t_b + \frac{\pi v_0}{\omega_0}(2n + 1) - \frac{v_0\psi_0}{\omega_0}. \quad (7)$$

Eventually, we obtain the spatial distribution of the electrons in the bunch at fixed time instant t_b (see Fig. 1). Inclination α is found from (7),

$$\tan\alpha = |dz_b/dx_b| = v_0/c. \quad (8)$$

Consider a charge moving with constant velocity $v < c/n$, where c is the speed of light in a vacuum and n is the refractive index. Under these conditions, Cherenkov radiation is absent and the bremsstrahlung of the charged particles at the target can be neglected. In addition, we assume that the electrons leaving the vacuum fall on a nonmagnetic perfectly conducting metallic screen.

The back-radiation energy is then calculated by the formula [2]

$$W^r(\omega, \theta) = \frac{1}{4\pi\varepsilon_0}\frac{q^2 v^2 \sin^2\theta}{\pi^2 c^3\left[1 - \left(\dfrac{v^2}{c^2}\right)\cos^2\theta\right]^2}, \quad (9)$$

where θ is the angle between wavevector $\mathbf{k}$ of the transition radiation and vector $-\mathbf{v}$.

In (9), the angles are taken in the polar coordinate system where the radius is measured from the point of incidence of the electrons on the screen. The rotation (polar) axis of this coordinate system coincides with the normal to the plane of the screen, and the coordinates are specified by angles θ and φ and by polar radius r.

For convenience, the directional diagrams of the transition radiation will subsequently be represented in the polar coordinate system with angles θ' and φ' and polar radius r (angles θ and θ' are shown in Fig. 1). In the new system, the polar axis passes through the points of incidence of the electrons on the screen and is directed upward on the plane of the screen. The two systems are related by the formulas

$$\cos\theta = \sin\theta'\cos\varphi', \quad (10)$$

$$\sin\theta = \sqrt{\sin^2\theta'\sin^2\varphi' + \cos^2\theta'}. \quad (11)$$

In the primed coordinate system, formula (9) takes

the form

$$W^{\mathrm{r}}(\omega, \theta') = \frac{1}{4\pi\varepsilon_0} \frac{q^2 v^2 (\sin^2\theta' \sin^2\varphi' + \cos^2\theta')}{\pi^2 c^3 \left[1 - \left(\frac{v^2}{c^2}\right)\sin^2\theta'\cos^2\varphi'\right]^2}. \quad (12)$$

Consider the transition radiation from a single filamentary bunch under the assumption that its height equals the distance between point of incidence of neighboring bunches and the velocity spread in the bunch can be neglected. Then, for normal incidence of the electrons on the screen, the phase difference for the point of the filamentary bunch that is displaced by x from the initial point is given by

$$\Delta\varphi = 2\pi\frac{\Delta t_z + \Delta t_{\mathrm{d}}}{T_0} = 2\pi\frac{x\tan\alpha}{v}\frac{\omega}{2\pi} + \frac{x\omega\sin\theta}{c}$$
$$= \frac{x}{c}\omega + \frac{x\omega\cos\theta'}{c} = x\frac{\omega}{c}(1 + \cos\theta'), \quad (13)$$

where Δt_z is the time delay due to different times of arrival of the electrons from different parts of the bunch at the screen, Δt_{d} is the time delay due to different distances from the points of incidence of the electrons on the screen to the observation point, and $T_0 = 2\pi/\omega_0$ is the modulating field period.

Let us calculate the amplitude of the field on the screen between two neighboring points of incidence of the electron bunches. The distance between these points is clearly $L = \lambda_0 = 2\pi c/\omega_0$. Then, using (13), we get

$$E(\omega, r, \theta', \varphi')$$
$$= \int_0^{\lambda_0} \Xi(\omega, r, \theta', \varphi')\cos\left[\frac{\omega}{c}x(1 + \cos\theta')\right]dx, \quad (14)$$

where

$$\Xi(\omega, r, \theta', \varphi') = K\frac{v}{cr}\frac{\sqrt{\sin^2\theta'\sin^2\varphi' + \cos^2\theta'}}{\left(1 - \frac{v^2}{c^2}\sin^2\theta'\cos^2\varphi'\right)} \quad (15)$$

is the amplitude of the field of the transition radiation of the bunch per unit screen length at the point r distant from the screen (θ' is the angle between the wavevector of the transition radiation and the plane of the screen, and ω is the frequency of the transition radiation), and K is a proportionality factor that depends on the screen material properties and beam current density.

Substituting (15) into (14) yields

$$E(\omega, r, \theta', \varphi') = K\frac{v}{cr}\frac{\sqrt{\sin^2\theta'\sin^2\varphi' + \cos^2\theta'}}{\left(1 - \frac{v^2}{c^2}\sin^2\theta'\cos^2\varphi'\right)}$$
$$\times \int_0^{\lambda_0} \cos\left[x\frac{\omega}{c}(1 + \cos\theta)\right]dx.$$

Upon integration, we get

$$E(\omega, r, \theta', \varphi') = K\frac{v}{cr\omega}\frac{\sqrt{\sin^2\theta'\sin^2\varphi' + \cos^2\theta'}}{\left(1 - \frac{v^2}{c^2}\sin^2\theta'\cos^2\varphi'\right)}$$
$$\times \frac{1}{(1 + \cos\theta')}\sin\left[2\pi\frac{\omega}{\omega_0}(1 + \cos\theta')\right]. \quad (16)$$

The above formulas allow us to construct the angular and frequency dependences of the transition radiation field amplitude in the plane of incidence of the electronic filaments and also in the plane normal to the plane of incidence. The directional diagram of the transition radiation in the plane of incidence for $\omega = 1.1\omega_0$ is shown in Fig. 2. It is seen that the transition radiation from one filamentary electron bunch is highly smeared over angles. The energy flux density is negligibly small in any direction. Because of this, we will consider an interference way of raising the directionality and, hence, the power density of the radiation.

Calculation of the angular distribution of the transition radiation from a wide (much wider than the modulating beam wavelength) electron flux requires that the interference of rays from equilong ($L = \lambda_0$) areas on the metal surface be taken into account. The expression for the angle of interference maximum is easy to derive,

$$\cos\theta' = \frac{n\omega_0}{\omega}, \quad (17)$$

where n is an integer (order of interference).

Thus, at frequencies $\omega < \omega_0$, the energy of the radiation is negligibly low. In the frequency range $\omega_0 < \omega < 2\omega_0$, there exists one interference maximum of the radiation amplitude.

Substituting (17) into (16) yields

$$E(\omega, r, n, \varphi') = K\frac{vc}{r}\frac{\sqrt{(\omega^2 - n^2\omega_0^2)\sin^2\varphi' + n^2\omega_0^2}}{\omega^2 c^2 - v^2(\omega^2 - n^2\omega_0^2)\cos^2\varphi'}$$
$$\times \frac{\omega^2}{\omega^2 + n^2\omega_0^2}\sin\left[\frac{2\pi}{\omega_0\omega}(\omega^2 + n^2\omega_0^2)\right], \quad (18)$$

the angle of nth interference maximum being given by the expression $\theta' = \arccos(n\omega_0/\omega)$. Note that the value of E is other than zero at any angle φ', indicating that the amplitude of the transition radiation field varies with this angle only slightly.

The 2D directional diagram of the transition radiation from a long train of filamentary bunches in the direction $\varphi' = 0$ with the screen placed vertically is shown in Fig. 3 (up to the ninth order of interference).

Here, the transition radiation concentrates in narrow angular and frequency intervals, since the emissions from different bunches interfere; accordingly, the radiation power density increases significantly.
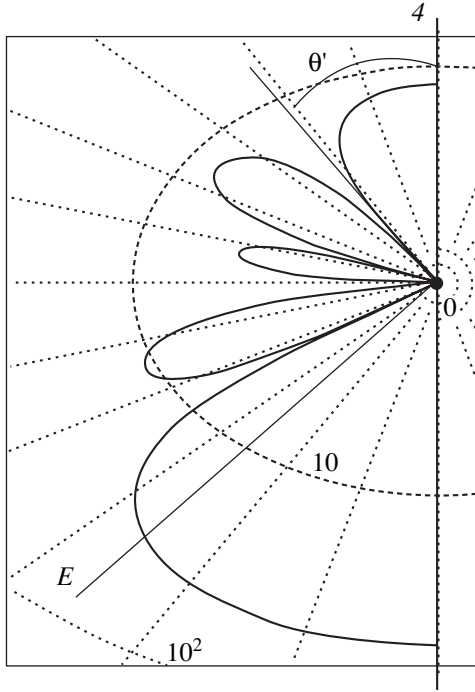
Fig. 2. Directional diagram of the transition radiation from filamentary electron bunches in the plane of incidence at $\omega = 1.1\omega_0$ and $v = 0.5c$. The plot is constructed in polar coordinates (r, θ') using formula (16). Here, θ' is the angle between screen plane *4* and the wavevector of the transition radiation; E is the transition radiation energy flux in the logarithmic scale.

Consider the transition radiation pattern when the modulated electron beam strikes the screen that makes angle β with the modulating electromagnetic beam. In this case, the spacing between the points of incidence of neighboring bunches on the screen is

$$L = \frac{\lambda_0 \sin\alpha\cos\beta}{\sin(\alpha - \beta)}, \qquad (19)$$

where α is the angle between the bunch and electromagnetic beam (Fig. 1).

Thus, if it is necessary to generate radiation at frequencies equal to, or lower than, that of the incident bunch, angle β should be selected so that the radiation intensity maximum is at a desired frequency.

In the case of the inclined screen, the angular direction of the transition radiation maximum differs from that for the vertical screen. Instead of (17), we have

$$\cos\theta' = \frac{n\omega_0}{\omega}\frac{\lambda_0}{L}. \qquad (20)$$

Substituting (19) into (20), we arrive at

$$\cos\theta' = n\frac{\omega_0}{\omega}(1 - \cot\alpha\tan\beta) = n\frac{\omega_0}{\omega}\left(1 - \frac{c}{v}\tan\beta\right). \quad (21)$$
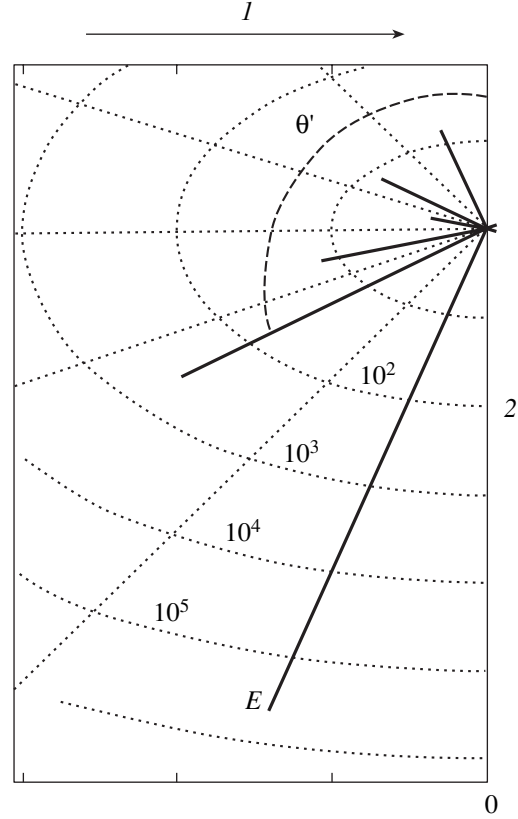
Fig. 3. Two-dimensional directional diagram of the transition radiation from a train of inclined electron bunches for the screen placed vertically ($\omega = 3.1\omega_0$). The diagram is constructed in the polar coordinates using formulas (17) and (18). Along the radius, the transition radiation energy flux in certain directions is plotted in the logarithmic scale. Interference is taken into account up to the ninth order. *1*, electron flux direction; *2*, line of incidence of the electrons on the screen.

From (21), it follows that, if $\beta \longrightarrow \alpha$ (Fig. 1), interference maxima are absent. For $\beta = \alpha - \pi/2$, an interference maximum of the first order will be observed.

Let us trace how the angular direction of the interference fringe varies with the inclination of the screen. In view of the fact that the normal component of the electron velocity changes $\cos\beta$ times when the screen is inclined by angle β, we find from formulas (18) and (21) that

$$E(\omega, r, \theta', \varphi') = K\frac{v\cos\beta}{cr\omega}\frac{\sqrt{\sin^2(\theta')\cos^2\varphi' + \cos^2\theta'}}{\left(1 - \frac{v^2\cos^2\beta}{c^2}\sin^2\theta'\cos^2\varphi'\right)}$$

$$\times \frac{1}{(1 + \cos\theta')}\sin\left[2\pi\frac{\omega}{\omega_0}(1 + \cos\theta')\right], \qquad (22)$$

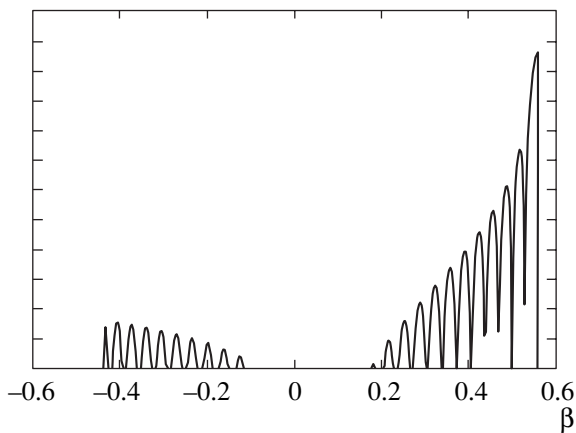where $\theta' = \arccos\left[n\frac{\omega_0}{\omega}(1 - c/v\tan\beta)\right]$.

Fig. 4. Energy flux of the transition radiation from a wide electron bunch vs. angle of inclination β (expression (22)). $\omega/\omega_0 = 7.1$. The bunch potential is 1.5 kV. Modulation is accomplished with a 694-nm ruby laser. The energy flux is shown in the direction of the field amplitude interference maximum.

Figure 4 plots the transition radiation flux density against inclination β of the screen at $\omega/\omega_0 = 7.1$. For simplicity, the plot is constructed for the first order of interference. For all other orders of interference, the pattern is basically the same.

From Fig. 4, it is evident that a major part of the energy concentrates in a narrow frequency interval. It is also clear that the angle of inclination of the screen can be taken such that the power of the transition radiation at a desired frequency rises greatly. Such adjustability is of great value in designing radiators finely tunable in a wide frequency range.

The maximal frequency of the transition radiation is limited by the properties of the metal the screen is made of. As follows from the reference books, in the case of a copper-made screen, the transition radiation disappears at frequencies higher than 2.3×10^{16} s^{-1}, because the properties of the metal change at these frequencies. If the modulation is accomplished using a 694-nm ruby laser ($\omega_0 = 2.7 \times 10^{15}$ s^{-1}), the transition radiation arises at frequencies up to $8\omega_0$. Figures 3 and 4 imply that the total radiation power will concentrate in several narrow angular peaks; accordingly, the transition radiation flux density increases greatly in the directions of these peaks (in comparison with the flux density associated with the unmodulated beam in the absence of interference). Varying the angle of inclination of the screen or the initial velocity of the electron flux, one can control the angular and frequency positions of the amplitude maxima in a wide range.

The data reported in this article indicate the possibility of using the transition radiation in continuously tunable sources of electromagnetic waves. The ways of concentrating the transition radiation energy at selected frequencies and within certain angular directions are considered. Estimation of the power and efficiency of the transition radiation is left beyond the scope of our report. Of particular importance is also taking into account the finiteness of the electron bunch width and the effect of the bunch on the interference pattern. Also, it is of applied interest to estimate the effect of a thermal spread in electron velocities on the width and lifetime of electron bunches. Finally, coherence of the transition radiation (hence, the validity of considering interference) in the case of the rough metallic screen surface deserves special theoretical consideration and experimental verification. It should be emphasized here that the last-named problem cannot be solved analytically. It can be tackled either experimentally or by constructing a probability model of surface irregularity distribution on the screen and then applying the model results in studying the properties of the transition radiation.

REFERENCES

1. V. L. Ginzburg and I. M. Frank, Zh. Éksp. Teor. Fiz. **16**, 15 (1946).
2. V. L. Ginzburg and V. N. Tsytovich, *Transition Radiation and Transition Scattering* (Nauka, Moscow, 1984; Hilger, Bristol, 1990).
3. V. I. Savel'ev, Mat. Model. **14** (11), 93 (2002).
4. G. M. Garibyan, Zh. Éksp. Teor. Fiz. **33**, 527 (1957) [JETP **6**, 411 (1957)].
5. M. I. Ryazanov, Pis'ma Zh. Éksp. Teor. Fiz. **39**, 569 (1984) [JETP Lett. **39**, 698 (1984)].
6. I. A. Anisimov and S. M. Levitskiĭ, Radiotekh. Élektron. (Moscow) **31**, 614 (1986).
7. B. M. Bolotovskiĭ and A. V. Serov, Zh. Tekh. Fiz. **72** (1), 3 (2002) [Tech. Phys. **47**, 1 (2002)].

Translated by V. Isaakyan

THEORETICAL
AND MATHEMATICAL PHYSICS

Entropy, Information, and Complexity of the Steady States of Open Systems Not Satisfying the Principle of Local Equilibrium

V. G. Usychenko

St. Petersburg State Polytechnical University, ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia
e-mail: Usychenko@rphf.spbstu.ru
Received September 13, 2004

Abstract—The entropy of the steady state of an open electronic system is determined, and its relation to the entropy arriving at the environment is established. The notion of system ordering is defined, and quantitative parameters characterizing ordering are introduced. The information and conditional complexity of the system versus the state of thermodynamic chaos are determined. These values are numerically estimated for vacuum, magnetron, and Gunn diodes. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

The theory of space–time structures [1–6] that form in open systems containing a large number of particles dates back to the evolution of thermodynamics. The most comprehensive and consistent thermodynamic theory of structure [2, 3] is valid under local equilibrium conditions, where temperature, pressure, and density at each point of the system are related through the equations of equilibrium state, whereas the entropy is given by the Gibbs formula. This theory has proved itself adequate in many problems of chemical physics, hydrodynamics, and biology; however, it is inapplicable, e.g., to electronic systems.

High-field domains in a Gunn diode [7, 8], vortices and solitons in a magnetron diode [9–11], and wave processes in both Penning cells [12] and a plasma [13] are generated at voltage gradients so high that the principle of local equilibrium fails; moreover, the temperature and entropy are no longer critical parameters of the system under these conditions. Electron motion in vacuum devices is described by time-reversible equations of mechanics [14], and the energy is dissipated not inside the device but at its boundaries, electrodes, which are exposed to particles bombarding them from the outside.

Yet, it is desirable to characterize the degree of ordering in steady-state electronic (and any other) structures using classical notions [1–6, 15] such as entropy, information, and complexity. However, these quantities do not directly follow from the thermodynamic theory of structure [2, 3], since this theory deals with production (variation) of entropy. Being a time derivative, the associated value gives no estimate of the steady-state value of entropy.

The focus of attention in this paper is electronic devices. Interest in these devices is associated with long-range Coulomb forces, which considerably favor electron self-organization, and also with the fact that these devices may operate both under near-equilibrium thermodynamic conditions and under substantially nonlinear conditions, which are separated from the states in the thermodynamic branch by points of instability. This circumstance greatly facilitates verification of the characteristic functions introduced to uniquely describe any operating regime. Among such functions used in this paper are the translational entropy, entropy, information, and complexity of a current steady state.

BASIC CONCEPTIONS

From the theory of electronic structures [14], it follows that the tendency of electrons toward self-organization in the devices shows up irrespective of the number of particles (it may be as small as desired) and a deviation from thermodynamic equilibrium, however slight. In other words, the electronic structures form both on and beyond the thermodynamic branch, being distinguished only by the degree of organization. Let us demonstrate this statement with the devices the synergetic theory of which is presented in [14]. By the structure, we will mean the spatial distribution of electrons, which is specified by their interaction with each other and with external fields.

The objects of study are a vacuum diode (VD), a magnetron diode (MD) and a Gunn diode (GD). The VD is of interest as a device all states of which lie on the thermodynamic brunch. The MD is also a vacuum device, but its steady states are beyond the limits of stability on the thermodynamic branch. Importantly, when the cathode emission changes, the MD demonstrates both regular and turbulent behavior [16]. As for the GD, its steady states depend on the applied voltage and may occur both on the thermodynamic branch (low field

conditions) and beyond a point of instability (high-field domain conditions). In addition, unlike the VD and MD, the GD is a solid-state, rather than vacuum, device and the energy of its particles is dissipated not at the boundaries of the system but in the interior. Let us select quantities that will be used below, starting from the synergetic theory of these devices [14].

The energy conservation law for an electron device operating in the steady-state regime is given by

$$W_N = W_\Sigma + W_{kr}. \tag{1}$$

The left-hand side represents the total internal energy of the system,

$$W_N = -e\beta NU_a + NkT_c. \tag{2}$$

Here, $-e\beta NU_a = -e \int_V n U_v dV$, where $e < 0$ is the electron charge; n is the electron density in elementary volume dV; N is the total number of particles in volume V of the device; and $U_v > 0$ is the vacuum potential (of the vacuum device in the absence of electrons). Averaging factor $\beta \le 1$ expresses the value of integral $\int_V n U_v dV$ in units of NU_a, where U_a is the anode voltage applied from an external source. In essence, coefficient β indicates the fraction of potential U_a per electron on average. Term NkT_c characterizes (up to a factor close to unity) the thermal energy of the particles, which depends on cathode temperature T_c in VDs and MDs, and k is the Boltzmann constant. The quantity

$$W_\Sigma = W_{sc} + \sum_i W_{ksi}, \tag{3}$$

on the right of Eq. (1) determines the energy of the electronic structure, which generally involves the potential energy

$$W_{sc} = -e \int_V n U_{sc} dV > 0$$

of the space charge ($U_{sc} \le 0$ is the space charge potential in elementary volume dV) and the sum of kinetic energies, $\sum_i W_{ksi}$, needed for the structure to move with a transport velocity in the field of external forces. The second term on the right of (1),

$$W_{kr} = 0.5m \int_V n u^2 dV = 0.5mN\langle u^2 \rangle, \tag{4}$$

takes into account the kinetic energy of electron motion relative to the structure. Here, u is the relative velocity of the particles in elementary volume dV and $\langle u^2 \rangle$ is the volume-averaged square of this velocity.

The appearance of particles in the system is described in terms of probability theory, since the site and time of emission, as well as the initial velocity vector, are random quantities for each particle. During self-organization, the particles also gain a coherent compo-

nent of the velocity, which reflects their collective movement in space. A coherent motion of a large number of particles can be allowed for by introducing a collective degree of freedom. However, in the course of self-organization, initially random forms of motion do not disappear, persisting in the relative velocity components.

Along with time-invariant quantities N and W_N, the steady-state operation of an electronic device is characterized by average lifetime $\langle \tau \rangle$ of electrons in the system and energy fluxes incoming to and outgoing from the device.

At a constant anode voltage, $U_a \gg -kT_c/e$, energy flux $dW_0/dt = P_0 = U_a I_a$ (where $I_a = -edN/dt \cong -eN/\langle \tau \rangle$ is the anode current) enters the device from a power supply. Within time $\langle \tau \rangle$, the power source applies energy $W_0 = -eNU_a$ to the device. For the same time, the electrons dissipate energy W_a (dissipation proceeds mainly at the anode in the MD and VD or in the crystal lattice in the GD). All energy W_a is transferred to the environment as heat; accordingly, power $P_a = dW_a/dt$ is the energy flux out of the device. The two forms of energy are related by an expression common to all the devices [14],

$$W_a = \zeta W_0 = -\zeta e NU_a, \tag{5}$$

where $0 < \zeta \le 1$. If the structure is absent or is immobile in the field of external forces, $\zeta = 1$. In this case, energy $W_0 = -eNU_a$ from the power source is completely dissipated, converting to heat. Such a situation is typical of the GD operating in the low field mode and of the VD when anode voltage $U_a \gg -kT_c/e$. If the structure possesses a collective degree of freedom, then, moving in the device space with a transport velocity, it does work

$$A = W_0 - W_a = (1 - \zeta)W_0 = \eta W_0, \tag{6}$$

with an efficiency η. It is clear from (5) and (6) that parameter $\zeta = 1 - \eta < 1$ determines the wasteful fraction of energy W_0.

For the MD operating in the soliton mode and the GD in the domain mode, the values of ζ are given by [14]

$$\zeta_{MD} = 1 + \frac{m}{2eU_a}(\omega\Omega(r_a^2 - r_c^2) - \Omega^2 r_a^2),$$
$$\zeta_{GD} = \left(1 - \frac{N_1}{N}\right)\frac{E_r}{E_a}, \tag{7}$$

where r_c and r_a are the cathode and anode radii of the MD, $\omega = -eB/m$ is the cyclotron frequency, B is the magnetic field induction, M is the electron mass, $\Omega \ll \omega$ is the angular frequency of the soliton (see figure), N_1 is the number of electrons in the domain, $E_a = U_a/l$ is the mean electric field strength in the GD of length l, and E_r is the electric field strength outside the domain.

Let us consider a number of specific examples to clarify the physical meaning of the quantities being used. The structure of the VD is represented by the

space charge, most electrons of which concentrate near the cathode. Coulomb interaction between the particles in the space charge is characterized by potential energy W_{sc}. The space charge as a whole is quiescent; i.e., the velocity coherent component is zero and, thus, $W_{ksi} = 0$, $W_a = W_0$, and $\zeta = 1$. The kinetic energy of any particle, $0.5mu^2$, is specified by its velocity u, which depends on smoothed potential $U = U_v + U_{sc}$ at the particle location. The value of W_{kr} in (4) is defined by the total kinetic energy of all the particles moving independently relative to the quiescent space charge.

The electronic structure of low-emission MDs is represented [11] by the Brillouin "hub," along which one (in the simplest case) soliton propagates. The structure has potential energy W_{sc}, and the wave travels around the cathode with constant angular frequency Ω. The cross section of this structure in the cylindrical coordinate system rotating with frequency Ω is shown in the figure. Moving with transport frequency Ω, the wave as a collective object does two works: work against the electromotive force and work against the centripetal force. The energies

$$W_{ks1} = 0.5m\int_V n(r^2 - r_c^2)\omega\Omega dV,$$

$$W_{ks2} = -0.5m\int_V nr^2\Omega^2 dV, \qquad (8)$$

spent on doing these works are picked up from the anode power supply. In the rotating coordinate system, the mean square of the velocity of the electrons bombarding the anode, $\langle u_a^2 \rangle$, satisfies the equality $\langle u_a^2 \rangle = -\zeta 2eU_a/m$ [14], from which we find parameter ζ given by (7), $\zeta = \zeta_{MD} < 1$.

Energy W_{kr} of the relative motion of the particles is described by formula (4), whose form differs radically from (8). If W_{kr} is a sum of quadratic values, i.e., reflects the fact that the particles move independently, energies W_{ks1} and W_{ks2} are expressed in terms of transport frequency Ω, which characterizes the coherent component of the motion of the same particles doing work coherently. Then, parameter ζ_{MD} (see (7)), which is directly related to the efficiency, depends on the transfer velocity alone.

In high-field GDs, the structure is represented by a domain possessing potential energy W_{sc} (outside the domain, the space charge is absent because of the electroneutrality of the semiconductor). The transfer of the domain to the anode is characterized by drift kinetic energy W_{ks1}. The total energy of the domain, $W_\Sigma = W_{sc} + W_{ks1}$, is not dissipated in the course of motion [14]. The existence of the domain is provided by the work done by the system with efficiency $\eta_{GD} = 1 - \zeta_{GD}$. In this case, only those electrons dissipate energy that are outside the domain. Drifting to the anode independently,
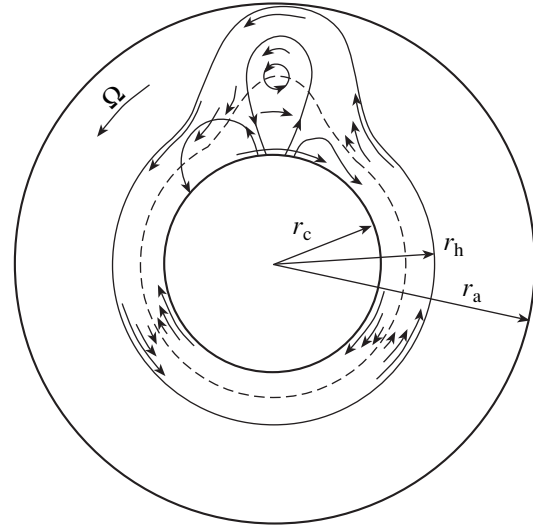
Space charge structure in the MD's central cross section. The lengths and directions of the arrows indicate the magnitudes and directions of the particle velocities; on the separatrix (dashed line), the velocities are equal to zero.

these particles have total kinetic energy W_{kr} (4). The facts noted above are reflected in formula (5) with parameter (7), $\zeta = \zeta_{GD} < 1$.

Thus, in the general case, energy W_Σ of space charge (3), which represents a collective structure, and kinetic energy W_{kr}, which takes into account incoherent particle motions, can be separated out in a device whose steady-state operating conditions are analyzed in an appropriate coordinate system. The sum of these two energies is total energy (1) of the system. There also exists energy flux $dW_a/dt = P_a$ escaping the device in the form of heat. It is important that energy W_a of this flux calculated for lifetime $\langle \tau \rangle$ of the particles is directly related only to energy W_{kr}. This relationship has the form [14]

$$W_{kr} = 0.5mN\langle u^2 \rangle = \theta W_a, \qquad (9)$$

where

$$\theta = \langle u^2 \rangle / \langle u_a^2 \rangle \leq 1 \qquad (10)$$

is a coefficient that characterizes the nonuniformity of the particle kinetic energy spatial distribution.

ENTROPY ENTERING AN ENVIRONMENT

Consider an environment into which energy W_a is dissipated as an infinite thermostat whose temperature T_e cannot be measured by a metering device at hand. In thermodynamics, the generalized coordinate of heat transfer is entropy S; therefore, the process of energy absorption by the environment can be described as

follows:

$$\frac{dW_a}{dt} = P_a = T_e \frac{dS}{dt}.$$

Hence,

$$\frac{dS}{dt} = \frac{1}{T_e} P_a.$$

Thus, total dissipated power P_a is spent to raise the entropy in the environment. For electron average lifetime $\langle \tau \rangle$, N electrons will transfer the entropy outside the device,

$$\Delta_{ex}S = P_a \langle \tau \rangle / T_e = W_a / T_e. \qquad (11)$$

The thermodynamic theory of structure [2, 3] uses the notion of entropy production. For example, for time $\langle \tau \rangle$, a resistor, which a low-field GD is, will produce the entropy

$$\Delta_i S = P_a \langle \tau \rangle / \langle T_i \rangle = W_a / \langle T_i \rangle,$$

where $\langle T_i \rangle$ is the temperature appropriately averaged over the volume of the GD.

Obviously, $\Delta_i S \le \Delta_{ex}S$, the equality being achieved in the limit $W_a \longrightarrow 0$, when $\langle T_i \rangle \longrightarrow T_e$.

In the vacuum devices, the energy of moving electrons is dissipated, being converted to heat, not in the interior but at the inner surfaces of the electrodes. This circumstance, however, is of little significance and expression (11) is equally applicable to vacuum devices as well.

Substituting W_a given by (5) into (11) yields an expression for the entropy transferred outside the device,

$$\Delta_{ex}S = \zeta \Delta_{ex}S_{max}, \qquad (12)$$

where

$$\Delta_{ex}S_{max} = -eNU_a / T_e \qquad (13)$$

is the expression common to all the devices, which determines a maximal amount of entropy gained by the environment iff energy $W_0 = -eNU_a$ picked up from the power source is completely converted to heat.

It is clear from (12) and (13) that entropy $\Delta_{ex}S$ is an extensive value. The ability of a specific device to transfer the energy to the environment depends on parameter ζ. Let us find the numerical values of this parameter for the diodes and gain a deeper insight into its physical meaning.

The space charge of the VD is quiescent in the field of constant external forces: energy $W_0 = -eNU_a$ delivered to the device is completely dissipated at the anode, the electrons travel independently, and $\zeta = 1$.

We now turn to the MD that was considered experimentally and theoretically in [11, 14, 17]. The cathode and anode radii in this MD are $r_c = 2.15$ and $r_a = 4.5$ mm, respectively. For $U_a = 2500$ V, $B = 0.14$ T, $\omega \cong$ $2\pi \times 3.9 \times 10^9$ Hz, and negligibly low cathode emission, the measured [16] and calculated [11] value of the wave frequency is $\Omega \approx 2\pi \times 4.06 \times 10^8$ Hz. Under such conditions, we find by formula (7) that $\zeta = \zeta_{MD} = 3.3 \times 10^{-2}$.

A low-field GD is a resistor where electrons drifting from the cathode to the anode gain and completely dissipate energy $W_0 = -eNU_a$. In this case, $\zeta = 1$. Under the high-field domain conditions, we find by formula (7) that $\zeta = \zeta_{GD} \approx 0.2$, assuming [14] that $N_1/N \ll 1$ and $E_r/E_a \approx 0.2$.

The examples cited above illustrate that, if $\zeta = 1$, cooperative forms of particle motion are absent. The higher the degree of electron cooperation, the smaller ζ.

ENTROPY OF A STEADY STATE

Let us multiply the right- and left-hand parts of (12) by coefficient θ. In view of (1), (9), and (13), we obtain a new quantity,

$$S_{st} = \theta \Delta_{ex}S = \theta \zeta \Delta_{ex}S_{max} = \theta W_a / T_e$$
$$= W_{kr}/T_e = -\theta \zeta eNU_a / T_e, \qquad (14)$$

which will be referred to as the entropy of a steady state of the system. Representing W_{kr} as $W_{kr} = T_e S_{st}$ and substituting this expression into (1) yields the law of conservation of energy in the device in the form

$$W_N = W_\Sigma + T_e S_{st}. \qquad (15)$$

Thus, total internal energy W_N of the system is equal to the sum of energy W_Σ of the structure and energy $T_e S_{st}$, which cannot be converted to the energy of the structure under current equilibrium conditions and is "useless" in this respect. Such an energy is called the bound energy in the thermodynamics of reversible processes. We will apply this term to energy W_{kr}.

The entropy of the steady state,

$$S_{st} = W_{kr}/T_e = \theta \zeta \Delta_{ex}S_{max}, \qquad (16)$$

is proportional to the outgoing entropy. For all the devices, entropy (11) leaving the system is related to heat transfer to the environment and, therefore, it is a thermodynamic quantity. On the other hand, the entropy of the steady state, W_{kr}/T_e, characterizes incoherent motions of particles appearing at random; therefore, it is a statistical quantity in essence and not in form. Thus, formula (16) relates thermodynamic entropy W_a/T_e to statistical entropy W_{kr}/T_e. The different physical meanings of these quantities are due to the difference in their origins. Specifically, entropy S_{st} is related to total internal energy (1) of the system through energy W_{kr}, whereas entropy $\Delta_{ex}S$ is associated with the flux of energy $dW_a/dt \cong P_a$ being released from the system in the form of heat.

Factor $\Delta_{ex}S_{max}$ in (16) indicates that S_{st} is an extensive (generalized) value, whereas the product $\theta \zeta \le 1$

reflects specific features of steady states in each of the devices and modes of operation. Let us find the values of product $\theta\zeta$ for the diodes operating in different modes.

To begin with, consider a VD with a short-circuited anode and cathode. At low emission, when the space charge is absent, so that $W_{sc} = 0$, the electron velocities obey the Maxwellian distribution [17] and are kept constant throughout the path. Therefore, $\langle u^2 \rangle = \langle u_a^2 \rangle$ and coefficient θ reaches a maximum, $\theta = 1$. Thermal electrons produce current I_a. Since $U_a = 0$, entropy $\Delta_{ex}S = 0$, according to (11). Thus, as was expected, the particles, being scattered at the anode, give up to the environment the same amount of energy (entropy) as they have when entering the system from the outside.

Intense emission generates the space charge. The number of electrons with an energy insufficient to get out of the potential dip rises considerably [17]. The larger the number of particles in the diode, the deeper the potential dip, the stronger the inequality $\langle u^2 \rangle < \langle u_a^2 \rangle$, and the lower θ.

Consider now a VD where the anode and cathode are connected through a resistor of resistance R. Current I_a through the resistor generates voltage $U_a < 0$ at the anode. From (11), it then follows that $\Delta_{ex}S < 0$; i.e., the entropy gained by the diode due to incoming electrons will be restored to the environment by the diode itself only partly, since the remaining part of the entropy will be restored by the resistor. Let us determine steady-state entropy $S_{st} = W_{kr}/T_e$ of the diode in this case. The value of W_{kr} will be found from energy conservation law (1), which, in the case considered, can be written as $NkT_c = W_{sc} + W_{kr} + I_a^2 R \langle \tau \rangle$. Eventually, we have

$$S_{st} = (NkT_c - W_{sc} - I_a^2 R \langle \tau \rangle)/T_e. \qquad (17)$$

It is clear from this formula that an increase in energy W_{sc} of the space charge, as well as the insertion of a resistor between the anode and cathode, decreases entropy S_{st}. The electron motion becomes ordered. In practice, this means the reduction of the electron flow noise. In terms of thermodynamics, formula (17) explains the well-known effect of low-frequency noise suppression by means of negative feedback through the space charge potential [17] and load resistance.

Let us now apply voltage $U_a \gg -kT_c/e$ to a VD. At a low emission rate, when $W_{sc} = 0$, we have $\theta = \langle u^2 \rangle/\langle u_a^2 \rangle = 0.5$. As the emission intensity grows, the space charge forms near the cathode and θ, as well as product $\theta\zeta$, diminishes. Entropy S_{st} (16) also decreases, so the degree of ordering is bound to rise. Indeed, it has been shown above that an increase in W_{sc} suppresses the current low-frequency noise [17].

In the GD, $\theta \cong 1$; therefore, $\theta\zeta \cong \zeta \approx 0.2$ in the domain mode. In the low-field mode, the GD behaves as a resistor for which $\theta \cong 1$ and $\zeta = 1$ [14].

Let us now find the value of θ for an MD with parameter ζ known. To this end, the electrons of the wave are ignored and only the electrons of the hub (see figure) will be taken into account. The error of such an approximation is reduced by assuming that the density of the particles in the hub, $n(r) = 1.7 \times 10^{17} \mathrm{m}^{-3}$, is distributed uniformly and equals the density at the cathode (actually, it declines gradually [11] over the radius of the hub, $r_h = 2.75$ mm, and is 17% lower at its outer surface).

The electrons in the hub are assumed to circle around the cathode [11] with radius-dependent azimuth velocity $u = r\dot{\psi} = 0.5r\omega(1 - r_c^2/r^2) - r\Omega$. Here, $\dot{\psi}$ is the angular frequency of the particle. Under this assumption, the number of electrons in the volume of the hub is

$$N = nl\pi(r_b^2 - r_c^2) \approx 1.6 \times 10^{10}$$

(where $l = 10$ mm is the MD length) and the energy of their relative motion,

$$W_{kr} \approx \pi mnl \int_{r_c}^{r_b} r^3 \dot{\psi}^2(r)dr \approx 8.4 \times 10^{-8} \text{ J.}$$

From (9) with regard to (5) and (7), we obtain $\theta\zeta = W_{kr}/(eNU_a) \approx 1.3 \times 10^{-2}$; hence, knowing $\zeta = \zeta_{MD}$, we find that $\theta \approx 0.4$.

It follows from the above examples that, if $\theta = 1$, the electron velocities obey the Maxwellian distribution. The smaller θ, the greater the deviation of the electron velocity distribution from the Maxwellian law.

As the emission from the cathode grows, the electron density in the MD hub remains unchanged, while the number of waves and the electron density in the waves increase [11]. Accordingly, the values of θ and ζ change. Eventually, a turbulent regime sets in [11, 16]. Measurements show (see, e.g., [18–20]) that, in fully developed turbulence, the electron velocity distribution is Maxwellian ($\theta = 1$) with temperature $T_{trb} \gg T_c$. Supposing that the formula $W_{kr} = 1.5kT_{trb}N$ is also applicable in this regime, along with the formula $W_{kr} = -\theta\zeta eNU_a$, we obtain $(\theta\zeta)_{trb} = -3kT_{trb}/2eU_a$. In the MD considered, fully developed turbulence is characterized by temperature $T_{trb} \approx 2 \times 10^6$ K [21], which yields $(\theta\zeta)_{trb} = \zeta_{trb} \approx 0.14$ at $\theta = 1$.

Among the diodes under consideration, only a low-emission VD, $U_a \ll -kT_c/e$, and a low-field GD exhibit the Maxwellian distribution of electron velocities and the maximal value of $\theta\zeta$, $\theta\zeta \cong 1$. Such states in a system of particles can be called "electronic thermodynamic chaos." In the fully developed turbulent mode of MD operation, we have $\theta_{trb} = 1$ and $\zeta_{trb} \approx 0.14$. The value $\zeta_{trb} < 1$ testifies that turbulence is a form of coherent

particle motion that differs radically from thermodynamic chaos ($\zeta = 1$). Indeed, experimental data [16, 21] show that, when the MD operates in the fully developed turbulent mode, a statistically averaged set of approximately 10^6 jointly traveling electrons, rather than a single electron, should be viewed as a charge unit. The results obtained numerically support the concepts of turbulence, which are accepted in contemporary thermodynamics [22, 23] and physics [24, 25].

It is clear from (14) that the entropy

$$S_{st} = W_{kr}/T_e = -\theta\zeta eNU_a/T_e \qquad (18)$$

is a measure of incoherent particle motions persisting in a system the rate of which is characterized by bound energy W_{kr}. Product $\theta\zeta$ specifies the degree of disorder in the system. In a completely ordered system, $W_{kr} = 0$ and $\theta\zeta = 0$. In this case, the particles are fixed relative to each other and occupy sites where the resultant force is zero. However, such a system is isolated rather than open. Indeed, through an open system, a particle flux flows and, hence, condition (9), $W_{kr} = \theta W_a > 0$, means the fundamental inaccessibility of absolute order. It has been mentioned above that the collective structures in the VD and MD exchange energy and matter with the environment owing to just relative velocities of the particles. Therefore, W_{kr} also characterizes the energy of exchange processes occurring inside the system.

Let us draw intermediate conclusions. In a completely disordered system, the particles are in the state of thermodynamic chaos. This state features the absolute absence of correlations and the Maxwellian distribution of electronic velocities. Parameters ζ and θ allow for the deviation of a specific state of the system from thermodynamic chaos. Specifically, parameter ζ characterizes the coherency of particle motion: when correlations are totally absent, $\zeta = 1$, while an increase in the number of jointly traveling particles is accompanied by a decrease in ζ. Coefficient θ shows to what extent the electron velocities obey the Maxwellian distribution: for the complete correspondence, $\theta = 1$. The greater the deviation, the smaller θ. Based on the aforesaid, parameter ζ can be referred to as "the parameter of spatial disorder of particles" and coefficient θ as "the coefficient of velocity disorder of particles (or their clusters in the turbulent regime).

ENTROPY AND TEMPERATURE OF A POWER SOURCE

Formula (6) represents the first principle of thermodynamics for electronic devices. If energy $W_0 = -eNU_a$ delivered to the device is completely spent on doing work A and transferring heat W_a to the environment (i.e., any other energy losses, e.g., those due to electromagnetic radiation, are absent), the Carnot theorem on maximum work [26] is valid. Entropy $\Delta_{ps}S = W_0/T_{ps}$ transferred from the power source to the device for time $\langle\tau\rangle$ (here, T_{ps} is the temperature of the power source)

will be equal to entropy $\Delta_{ex}S = W_a/T_e$ transferred from the device to the environment. Equating these entropies, we obtain an expression for the effective temperature of the power source,

$$T_{ps} = T_e\frac{W_0}{W_a}. \qquad (19)$$

Ratio $T_{ps}/T_e = W_0/W_a = 1/\zeta$ demonstrates the efficiency of power source energy usage by the device.

INFORMATION AND COMPLEXITY

We assume that the Carnot theorem holds true at any change in the operating mode of the device. Then, in the absence of hysteresis, all processes taking place in the device can be considered [26] as reversible and entropy S_{st} (14) of a current steady state can be regarded as being a function of this state. This means that the entropies of any initial state and any final state are independent of the route which the transition between them follows. Imagine, for example, that the anode voltage, magnetic induction, and number of electrons are changed in the MD. It is easy to check that the entropy of the final state will be independent of the sequence of these changes. Hence, it follows that the difference between the initial and final entropy values will also be a function of state.

Let us introduce the entropy of the reference (standard) state of the system as follows:

$$S_{st}^0 = W_0/T_e = \Delta_{ex}S_{max} = -eNU_a/T_e. \qquad (20)$$

This expression coincides with (13) and gives the entropy of a steady state for which $\theta\zeta = 1$ (see (14)). Taking the entropy of the standard state, S_{st}^0, as the initial value and entropy S_{st} of a current steady state as the final value, we arrive at a new function of state,

$$I = S_{st}^0 - S_{st}, \qquad (21)$$

which will be referred to as the thermodynamic information.

Using the GD as an example, let us have a closer look at the evolution of processes taking place in the device after switching on the anode voltage source. At the instant voltage U_a is applied, product $\theta\zeta = 1$ and the standard state entropy is described by the relationship $S_{st}^0 = \Delta_{ex}S_{max}$, according to (20) and (14). After a time, a high-field domain forms and the steady-state regime is established, which is characterized by product $\theta\zeta \approx 0.2$ and entropy $S_{st} = \theta\zeta\Delta_{ex}S_{max} = 0.2\Delta_{ex}S_{max}$.

Let us write formula (21) in expanded form,

$$I = (1 - \theta\zeta)\Delta_{ex}S_{max} = -(1 - \theta\zeta)eNU_a/T_e. \qquad (22)$$

Substantiating the principle of negentropy of information, Brillouin [27] put forward the idea that a system in the state characterized by a maximal value of

entropy S_0 possesses bound information $I_{b0} = 0$. If the negentropy flux reduces the entropy of the system to final value S_1, the bound information is $I_{b1} = S_0 - S_1$. In essence and in form, the Brillouin bound information coincides with thermodynamic information given by (21) and (22). Up to a constant factor the value of which is insignificant in the given case, bound information I_{b1} equals the maximal value of the Shannon's information entropy. Thus, Eq. (21) relates thermodynamic entropy $S_{st}^0 = -eNU_a/T_e$, statistical entropy $S_{st} = W_{kr}/T_e$, and information entropy I. It is worthy to note that statistical entropy equals thermodynamic entropy only if I is zero.

Thermodynamic information has its own energy measure,

$$W_I = -eNU_a - W_{kr} = -(1 - \theta\zeta)eNU_a, \qquad (23)$$

which can be obtained from (22) by multiplying by the environmental temperature. Information energy W_I is contained, e.g., in the food of living organisms. Information I can be lost and the information energy can be released when the structure breaks. Such an energy release accompanied by the formation of molecular chaos occurs, e.g., in the course of organic substance burning.

Information I is an extensive quantity, characterizing a system as a macroscopic object. At the same time, it contains an intensive (local) component,

$$\iota = 1 - \theta\zeta, \qquad (24)$$

which characterizes specific features of the structure at the microlevel.

Thermodynamic chaos is the simplest state of a system, since it can be completely described by the only intensive parameter, temperature. Any deviation from this state raises the complexity of the system, and additional parameters have to be introduced for its description. The complexity of one object, X, relative to that of another object, Y, is called conditional complexity $K(X/Y)$—the term introduced by Kolmogorov [28]. From (24), it follows that the conditional complexity of a system can be measured with the parameter

$$\kappa(x/y) = \iota = (1 - \theta\zeta), \qquad (25)$$

which characterizes the structure of the system at the microscopic level. Complexity $\kappa(x/y)$ is independent of the dimension of the system; therefore, it is not surprising that the human genome and the genome of a tiny nematode are comparable in complexity. From (25), it follows that the Kolmogorov conditional complexity is a thermodynamic information of the system that is contained in the system itself.

Numerical values of all the parameters discussed above are listed in the table.

Judging by the value $(1 - \theta\zeta) = 0$, the least complex systems are a VD for which $-eU_a \ll kT_c$ and a low-field GD. The most complex system, i.e., that requiring the

Table

Param-eter	VD, $-eU_a \ll kT_c$; low-field GD	D, $-eU_a \gg kT_c$	MD, regular mode	MD, turbulent mode	GD, domain mode
0	1	≤ 0.5	0.4	1	1
ξ	1	1	0.033	0.14	0.2
$\theta\xi$	1	≤ 0.5	0.013	0.14	0.2
$1 - \theta\xi$	0	≤ 0.5	0.987	0.86	0.8

largest number of phenomenological parameters for its description, is an MD operating in the regular mode, $(1 - \theta\zeta) = 0.987$. The turbulent regime is less complex: $(1 - \theta\zeta) = 0.86$. Indeed [16, 21], only two parameters are necessary for its description: temperature T_{trb} and average number N_{trb} of particles per coherent "fragment."

Thus, the starting point for measuring the degree of ordering in a system is the state of thermodynamic chaos, which is quantitatively measured in terms of the entropy of the standard state (see (20)). Thermodynamic information (as well as complexity) is a measure of deviation of a system's particular steady state from thermodynamic chaos.

PRINCIPLES OF SELF-ORGANIZATION

Let us turn to formula (23). Given N and U_a, energy W_I is determined by energy W_{kr}. According to the principle of minimization of integral Lagrangean [14], the self-organization of particles in open systems is accompanied by a decrease in energy W_{kr}. In the steady state, W_{kr} reaches a minimal value. According to (18), the entropy of the steady state also reaches a minimum. In this case, information I is maximal, as follows from (23) and (22). Thus, the principle of Lagrangean minimization implies that a system tends to the steady state, which has a minimal entropy and a maximal information and complexity.

Having introduced quantities $K\ln R_0$ and $K\ln R_1$, where K is a constant and R_0 and R_1 are equiprobable samplings in the initial and final states, Haken [29] writes the information in the form $I = K\ln R_0 - K\ln R_1 = K\ln R$ and, using an alphabet containing a large number of symbols, passes to the information entropy

$$i = -K\sum_j p_j \ln p_j, \quad \sum_j^M p_j = 1$$

in Shannon's formulation (up to coefficient K). Then, he shows that, in a many-particle physical system, the value of i reaches a maximum in the steady state. If K is the Boltzmann constant, the maximum of the Haken information entropy corresponds to the maximum of thermodynamic information (22) and to the maximum

of Brillouin bound information I_{b1}. These three maxima refer to that state of the system in which energy W_{kr} and entropy S_{st} reach maximal values. Thus, the principles of self-organization of matter mentioned above are consistent with each other. In states close to thermodynamic equilibrium, we have $\theta\zeta \cong 1$, $W_{kr} \cong W_a$, and $S_{st} = S_{st}^0 \cong \Delta_{ex}S \cong \Delta_i S$. Since W_a is the dissipated energy and $\Delta_i S$ is the production of entropy, minimization of these two quantities means fulfillment of the Onsager principle of least dissipated energy and the Prigogine entropy production minimum principle.

CONCLUSIONS

The description of steady-state regimes that are far from thermodynamic equilibrium requires two energy relationships to be invoked. The law of conservation of total internal energy (see (1)) defines the shape, volume, internal structure, and performance of a system, as well as its steady state entropy and complexity. The law of conservation of energy (see (6)) defines the rate of energy, matter, and entropy exchange between the system and the environment.

The steady states of open systems lie between two extreme cases. One of them is thermodynamic chaos, which is described by the probability law, the Maxwellian distribution. In this state, the entropy of the system is maximal, whereas the information and complexity are zero. The other is a perfectly ordered state, which is completely described by deterministic laws. Here, the entropy is zero, whereas the information and complexity are maximal. Thermodynamic chaos is the simplest state, which is most frequently encountered in nature. Therefore, any process describable by deterministic functions is "perceived" by nature as an unusual event that carries information about something other than from thermodynamic chaos.

Absolute order in an open system is an idealization, because it is impossible to reach the state that is characterized by $W_{kr} = 0$ and $\theta\zeta = 0$. The inaccessibility of absolute order means that a complete and exact description of an open system is basically impossible. It has been mentioned above that W_{kr} covers the random components of particle motion and characterizes the energy of "exchange processes." Thus, we cannot gain insight into exchange processes so much deep that all elementary events involving individual particles be exactly known. The Maxwell demon alone could observe all these events simultaneously. However, to do this would require a light source with an energy greater than the energy of the system [27]. Such a severe intervention to the system would drastically alter it. It only remains for us to invoke probabilistic methods for comprehensive description of elementary processes.

REFERENCES

1. E. Schrödinger, *What is Life? The Physical Aspect of the Living Cell; with Mind and Matter; and Autobiographical Sketches* (Cambridge Univ. Press, Cambridge, 1992; RDKh Izhevsk, 1999).

2. I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes*, 3rd ed. (Interscience, New York, 1968; RKhD, Izhevsk, 2001).

3. P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations* (Wiley, New York, 1971; Mir, Moscow, 1973).

4. W. Ebeling, *Strukturbildung bei Irreversiblen Prozessen* (Teubner, Leipzig, 1976; Mir, Moscow, 1979).

5. G. Nicolis and I. Prigogine, *Exploring Complexity* (Freeman, New York, 1989; Mir, Moscow, 1990).

6. W. Ebeling, A. Engel, R. Feistel, *Physik der Evolutionsprozesse* (Akademie-Verlag, Berlin, 1990; Editorial URSS, Moscow, 2001).

7. M. E. Levinshteĭn, Yu. K. Pozhela, and M. S. Shur, *Gunn Effect* (Sov. Radio, Moscow, 1975) [in Russian].

8. H. Haken, *Synergetics: An Introduction* (Springer, Berlin, 1977; Mir, Moscow, 1980).

9. N. A. Kervalishvili, Fiz. Plazmy **15**, 174 (1989) [Sov. J. Plasma Phys. **15**, 211 (1989)].

10. A. V. Agafonov, V. M. Fedorov, and V. P. Tarakanov, Preprint No. 37, FIRAN (Lebedev Physical Institute, Moscow, 1997).

11. V. G. Usychenko, Radiotekh. Élektron. (Moscow) **46**, 1489 (2001).

12. N. A. Kervalishvili, Fiz. Plazmy **15**, 362 (1989) [Sov. J. Plasma Phys. **15**, 436 (1989)].

13. S. V. Korobtsev, D. D. Medvedev, and V. D. Rusanov, Fiz. Plazmy **19**, 567 (1993) [Plasma Phys. Rep. **19**, 291 (1993)].

14. V. G. Usychenko, Zh. Tekh. Fiz. **74** (11), 38 (2004) [Tech. Phys. **49**, 1431 (2004)].

15. D. S. Chernavskiĭ, *Synergetics and Information: A Dynamic Theory of Information* (Nauka, Moscow, 2001) [in Russian].

16. A. V. Smirnov and V. G. Usychenko, Radiotekh. Élektron. (Moscow) **36**, 151 (1991).

17. *Noise in Electron Devices*, Ed. by L. D. Smullin and G. A. Hause (MIT, Cambridge, 1959; Energiya, Moscow, 1964).

18. I. V. Vigdorchik, Zh. Tekh. Fiz. **6**, 1657 (1936).

19. E. C. Linder, Proc. IRE **26**, 344 (1938).

20. G. G. Sominskiĭ, Zh. Tekh. Fiz. **38**, 663 (1968) [Sov. Phys. Tech. Phys. **13**, 497 (1968)].

21. V. G. Usychenko, Radiotekh. Élektron. (Moscow) **44**, 746 (1999).

22. L. G. Loĭtsyanskiĭ, *Fluid Mechanics* (Nauka, Moscow, 1987) [in Russian].

23. O. M. Belotserkovskiĭ, A. M. Oparin, and V. M. Chechetkin, *Turbulence: A New Approach* (Nauka, Moscow, 2002) [in Russian].

24. I. Prigogine and I. Stengers, *Order out of Chaos: Man's New Dialogue with Nature* (Heinemann, London, 1984; Progress, Moscow, 1986).

25. Yu. L. Klimontovich, Pis'ma Zh. Tekh. Fiz. **10** (2), 80 (1984) [Sov. Tech. Phys. Lett. **10**, 136 (1984)].

26. D. Kondepudi and I. Prigogine, *Modern Thermodynamics: From Heat Engines to Dissipative Structures* (Wiley, Chichester, 1998; Mir, Moscow, 2002).

27. L. Brillouin, *Science and Information Theory* (Academic, New York, 1962; Fizmatgiz, Moscow, 1960).

28. A. N. Kolmogorov, *Information Theory and the Theory of Algorithms* (Nauka, Moscow, 1987; Kluwer, Dordrecht, 1993).

29. H. Haken, *Information and Self-Organization: A Macroscopic Approach to Complex Systems* (Springer, Berlin, 1988; Mir, Moscow, 1991).

Translated by M. Lebedev

THEORETICAL
AND MATHEMATICAL PHYSICS

Fractal Signature Methods for Profiling of Processed Surfaces

A. A. Potapov*, V. V. Bulavkin**, V. A. German*, and O. F. Vyacheslavova***

* Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Mokhovaya ul. 11, Moscow, 125009 Russia
e-mail: potapov@mail.cplire.ru
** Tekhnomash State Unitary Research Enterprise, Moscow, 127018 Russia
*** Moscow Automotive Institute (Technical University), Moscow, 105839 Russia
Received September 20, 2004

Abstract—A new approach to surface profiling of structural materials that evolves from the concept of fractal signature is put forward. This approach has been developed and advantageously applied for acquisition of low-contrast targets. It is based on the fractal theory, and fractal signatures and fractal dimensions (which are intimately related to both the object's topology and evolutionary processes in dynamic systems) are used as estimating parameters. The experimental data obtained prove the existence of fractal clusters on the processed surface microrelief. Quantitative characterization of these clusters is given. © 2005 Pleiades Publishing, Inc.

INTRODUCTION

Reliable and long-term functioning of a product to a great extent depends on the surface condition. Surface quality control is therefore a key problem in materials science. It is common knowledge that surface properties, such as corrosion and wear resistance, as well as tribological and reflection properties, have a dramatic impact on the performance of a product as a whole and of its parts [1].

Although this postulate has been long known and is beyond question, surface conditioning, specifically, roughness control, is now assuming an even greater significance in light of emerging material processing technologies. Nanotechnology, where roughness is considered as an inherent property of a structure, and not as a "response" (secondary characteristic) of the surface layer structure to an external physical action (e.g., cutting), is a most vivid example, since the thickness of nanometer layers is comparable to the electron free path in this case.

The surface quality is routinely characterized by roughness (specifically, average roughness R_a, peak-to-peak height R_{max}, average roughness asperity spacing S_m, etc.) and the physicomechanical properties of the surface layer [1]. Surface roughness is considered as a static object arising as a response to some action (physical processing). In other words, roughness may be viewed as the trace of a tool on a surface being processed that is represented by geometrical images: globes, cones, cylinders, and so on. The roughness value is estimated indirectly through the parameters of processing (influencing factors). In machining, these are cutting rate, depth of cut, and others.

Evidently, it is impossible to reliably predict the behavior of a given surface during use if appropriate techniques that characterize its topography and estimate the geometrical characteristics adequately describing the surface profile formation are lacking. Estimates of this kind are of particular value in forecasting the behavior of sophisticated systems (i.e., missiles or space vehicles) where operating reliability and safety are critical issues.

In view of the aforesaid, it is necessary to work up alternative approaches to estimating the surface roughness (at least for a certain group of surfaces). An avenue for tackling this problem may be found in the fractal theory: specifically, fractal (fractional) dimension D and fractal signatures may be taken as estimating parameters [2]. This work is aimed at checking this assumption. It is hoped that such an approach could make it possible to uniquely estimate surface roughness.

CATEGORIZATION OF RELIEF FEATURES ON SURFACES PROCESSED BY MODERN TECHNIQUES

Extensive application of surface processing techniques using concentrated energy fluxes (laser, plasma, and electroerosion techniques) and advances in nanotechnology (chemical assembling, sol–gel methods, CVD of metals, atomic layer epitaxy, etc.) make description and estimation of roughness by means of profilometry difficult. Roughness features and their distribution over a surface thus processed are far from being a periodic sequence of "valleys" and "peaks" describable in terms of Euclidean geometry (Table 1)

[1]—a wide-spread belief related to machining. When the material is exposed to intense energy fluxes; high pressure; fast-moving gaseous media; or a combination of mechanical, ultrasonic, etc., vibrations, roughness features take the "strange" shapes of mushrooms, ridges, grapes, splats, whiskers, etc. [1, 3, 4]. The distribution of these features over the surface is irregular and has different probability densities.

These features are difficult, if at all possible, to describe in terms of classical geometrical parameters, and the surface topography cannot be adequately described by the Abbot–Firestone reference curve [3]. It is worth noting here that the idealized fractal surface assumes roughness features of any scale. A model where the feature height is assumed to be fractal inevitably becomes nondifferentiable (such notions as the normal to the scattered wave front and the ray path, as well as various effects of geometrical optics, are thus excluded from consideration), and a model where the feature slope is fractal allows for only the first derivative. Slopes, curvatures, and distributions of extrema are not therefore critical surface parameters for the fractal relief, as has been usually believed (for details, see [2]).
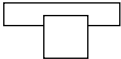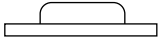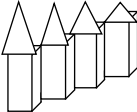
New technologies where an entire surface being processed is exposed to an external action (group surface-conditioning techniques) generate clusters, the evolution of which was outlined in [5]. Specific features of the new processes dictate the need for revising the available approaches to predicting the properties of clusters and parametrically characterizing them.

PROCESSES IN MATERIALS SCIENCE AND RADAR AS EXAMPLES OF THE SYNERGETIC PARADIGM OF OPEN PHYSICOTECHNICAL SYSTEMS

The fractal theory is being advantageously used in materials science to obtain materials with tailored properties [6, 7]. This theory is naturally related to the concept of synergetics as self-organization of structures. The fractal theory may be viewed as a basis for quantitative description of dissipative structures forming under highly nonequilibrium conditions. Elaborating upon such an idea, researchers at the Institute of Metallurgy and Materials Science, Russian Academy of Sciences, have formulated a new branch of materials science, fractal materials science [6, 7]. Now, it becomes possible to esta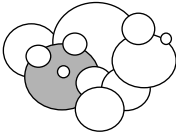blish a correlation between the composition, fractal structure, and properties of a material, which is of great importance in surface conditioning, e.g., in nanotechnology, where it is impossible to distinguish between the bulk and surface structures of the material.

If by a dynamic system (DS) is meant an object or process the state or a set of parameters of which is uniquely defined at a given time and a determinate evolution operator is specified, such a concept of dynamic

Table 1

Surface relief features

Mushrooms	
T-shaped	
Pyramids	
Splats	
Grapes	
Circular lunes	
Globes	
Globes/whiskers	
Ridges	
Moire	

system may be extended to objects of any nature. The DS properties are defined by invariants, such as Lyapunov indices, fractal dimension of the strange attractor, and entropy. Knowing these invariants, one can determine the number of independent variables and find a finite dimension of a phenomenon. After transients have faded out, the phase space of the system contains a limiting set of points that attract phase trajectories (the so-called attractor). The existence of the attractor is associated with a contraction of the DS space volume under the action of the evolution operator.

The attracting set in the DS phase space with aperiodic steady-state oscillations has been called the strange attractor [8]. The strange attractor always has fractal (fractional) dimension D. An important feature of chaotic motion is that initially it is extremely sensitive to small changes. This means that closely spaced trajectories in the phase space diverge exponentially in time. Quantitatively, the rate of divergence is given in

terms of the Lyapunov indices, which exhaustively characterize the complex chaotic behavior and structure of the attractor in the DS phase space. Of most significance is Lyapunov index λ_1, the positiveness of which indicates chaos present in a DS. Remarkably, the geometry and dynamics of the strange attractor are intimately related: the Lyapunov indices describe the geometry of the attractor, and the values of the indices can be estimated by measuring fractal dimension D.

The basic method of research in metallurgy, materials science, and materials processing remains to be the thermodynamic method, which allows scientists to predict the general behavior of a system irrespective of mechanisms behind processes involved. When in thermodynamic equilibrium, a system meets the conditions

$$S(U, V, n_i) = \max; \quad dS = 0, \tag{1}$$

where U is the internal energy, S is the entropy, V is the volume, and n_i is the number of moles of components ($i = 1, 2, \ldots$).

From (1), it follows that thermodynamic equilibrium corresponds to a maximal degree of disorder and the impetus to equilibrium processes is the tendency of the system to minimize its free energy. However, the surrounding world is a rigid hierarchical system. This comes into conflict with the second principle of thermodynamics, which states that the system's entropy rises with time. This contradiction was removed with the evolution of the idea of synergetics. The term "synergetics" covers such areas as the theory of self-organization (or the theory of dissipative systems), theory of open systems, theory of nonequilibrium systems, information dynamics, dynamic theory of shaping, etc. Synergetics relies on the principles of (i) minimal entropy production, (ii) current and local equilibrium, (iii) mosaic nonequilibrium thermodynamics, (iv) least constraint, (v) subordination, and (vi) "nonequilibrium as a source of order" (see, e.g., [7, 9–14]).

These principles may be extended to both living and inorganic nature. The simplest evolutionary equation has the form

$$\dot{q} = \alpha q, \tag{2}$$

where $\dot{q}$ is the rate of production of a substance, q is its concentration, and α is the order parameter.

Synergetic systems are stochastic: their time evolution (behavior) cannot be predicted with an absolute accuracy. Therefore, Eq. (2) must be augmented by term $f(t)$ taking into account force fluctuations,

$$\dot{q} = \alpha q + f(t). \tag{3}$$

Surface roughness associated with physicochemical methods of processing differs from roughness due to cutting. In the latter case, it is the trace of a tool on the surface, which is sequentially "swept" in space and time. In the case of physicochemical methods, the working medium itself (liquid, metal vapors, and so on) serves as a tool. The components of the working medium, which are activated by high temperature, high pressure, or fields of various type, intensely interact with the entire surface almost simultaneously at each point of their location in the working space. This generates the so-called effect of group action on the surface. Thus, the surface processed by physicochemical methods is the result of action of a set of variously activated elementary processes and can be viewed as a synergetic system.

Let us now turn to radiophysics, the area of physics dealing with waves and oscillations in a wide electromagnetic range. Radiophysics has given rise to a number of fields of engineering, radar among them. Extensive research is now under way worldwide aimed at making objects radioparent using advanced materials offering an extremely low reflection coefficient in the rf range (so-called Stealth technology). According to [15], in 1980, the F-15 fighter had a radar contact (σ) of about 10 m^2; subsequently, σ was decreased to 1.0–1.5 m^2. For comparison, F-22 and JSF fifth-generation supersonic fighters have σ on the order of 0.3 m^2. In light of these advances, radar approaches need to be updated. Standard filtering methods are appropriate if arriving radar signals are poisoned by Gaussian noise alone. Actually, however, noise is almost always non-Gaussian with extended tails and may have a high intensity in different spectral ranges. Therefore, under real conditions, the use of classical filtering algorithms may be infeasible at low signal-to-noise ratio q_0^2.

Present-day radar systems combined with a probing channel should be viewed as complex nonequilibrium systems open to energy, entropy, and information fluxes. Elaborating upon this idea, scientists at the Institute of Radio Engineering and Electronics, Russian Academy of Sciences, formulated new information technologies as applied to radiophysics and radar. These fresh insights are based on using texture (mid-1980s) and fractal (mid-1990s) dimensions starting from the DS principles (see above). In essence, a new avenue of research has emerged: application of the DS theory and fractal topology for increasing the information content of radio engineering systems. From the results of these pioneering investigations, the notions of texture and fractal signatures were introduced and this line of inquire was briefly named "fractal radar" and "fractal radiophysics" [2].

Diagram 1 shows the basic problems involved in these investigations and the years the associated works began. Generally, any radio system, together with a medium of radio wave propagation and objects being probed, is natural to consider as open DSs with strange attractors and bifurcation points, which define the pattern of radiophysical processes occurring in such a space–time continuum. Radiophysical phenomena are best to treat in terms of nonlinear dynamics and fractal topology and considered as taking place in a dissipative environment open to information fluxes. Below, we will

discuss the basic notions of fractal analysis as applied to radar.

ON THE NOTION OF FRACTAL IN RADIOPHYSICS AND RADAR

In general form, a radar image (RI) can be represented as set X_k of pels whose values are proportional to the effective scattering area (ESA) of a kth resolving element of the radar station. Figure 1a demonstrates the RI of a locality taken by a helicopter-borne radar at a wavelength of 8.6 mm. Figure 1b shows the RI of the same locality taken at a wavelength of 30 cm. Both images are two-dimensional, the gray intensity being proportional to the ESA. Suppose that surfaces (Fig. 1c) with feature heights h proportional to the gray intensity are constructed for either RI and the surface area is to be measured. On the RI taken at 30 cm, the surface area will be larger, since the number of distinguishable locality features increases with decreasing wavelength. Here, the probing electromagnetic wave serves as a sort of "ruler." As the wavelength decreases, a progressively finer structure of space–time signals, or wave fields, begins to emerge.

If the RI at hand were obtained in a still shorter wave range, the surface area would grow still further; i.e., decreasing the wavelength, we increase the surface area. The obvious question arises: what is the true surface area of the locality imaged? If the surface were covered by objects of simple shape (for example, by rectangular hills; Fig. 1d) and their size far exceeded the wavelength, the surface areas on the RIs of the objects would be nearly the same. Then, we could answer this question, having calculated the number of resolving elements covering the objects. In this case, surface area S would be expressed as

$$S \equiv S(\lambda) = N(\lambda)\delta(\lambda), \qquad (4)$$

where $\delta(\lambda)$ is the area of a resolving element of the radar, $N(\lambda)$ is the number of resolving elements covering the object, and λ is the radar wavelength. It was noted above that $S(\lambda) = $ const for a simple-shape object (Fig. 1d).

For the images in Figs. 1a and 1b, the dependence $S(\lambda) = f(\lambda)$ can be constructed by putting $\delta(\lambda) = K(\lambda)$, where K is a known function, and then constructing the dependence $S(\lambda) = f(\delta)$. It turns out that measured surface area S is fitted well by the formula

$$S(\lambda) = k\lambda^{-D}. \qquad (5)$$

Since $\log(S(\lambda)) = \log k - D \log \lambda$ (k and D are constants), we can calculate parameter D. The dependence $\log S(\lambda) = f(\log \delta)$, which defines the fractal signature of an RI (Figs. 1a and 1b), is shown in Fig. 1e. It characterizes the spatial fractal cepstrum of an RI (the notion of fractal cepstrum was introduced in 1997 [2]). Fractional parameter D is called the Hausdorff–Bezikovich, or fractal, dimension. On the RIs of objects with
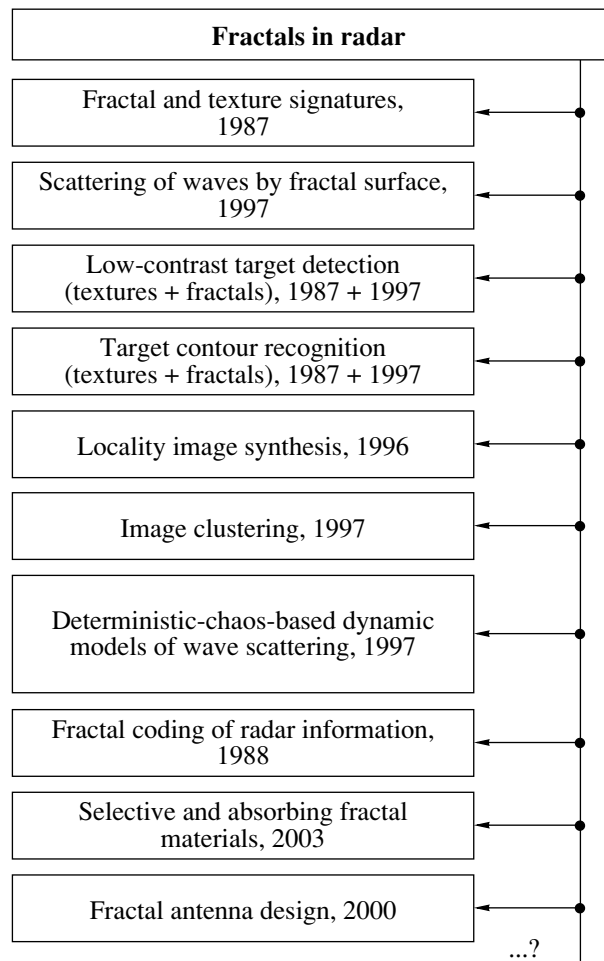
Diagram 1. Advances in information technology.

a simple geometry (rectangles, circles, and smooth curves), this dimension coincides with the topological one (i.e., equals two for two-dimensional RIs) and is specified by the slope of straight lines (5) in the log–log coordinates. However, for most RIs taken from real mantles and meteoric objects, the value of D exceeds a topological dimension of two, highlighting their complicated and irregular structure.

The fractal theory, dealing with geometry and the theory of dimension, is intimately related to the theory of dynamic chaos. It considers quantitative indices, such as fractional, instead of integer, dimension D and fractal signatures. Fractal fractional dimensions and signatures not only characterize the object's topology but also, being related to their properties of dynamic systems, reflect their evolution. The geometry of chaos is representable in terms of the fractal theory and nonlinearity. It appears, however, that fractals are the language not only of chaos but also of the nature. The contours of all natural objects are essentially dynamic processes combining stability and chaos that have suddenly "congealed" in their physical forms. In nature, ordering and disordering are basic interrelated

(a)   (b)   (c)

(d)

(e)

$\ln S$

$D = 2.440393$

$\ln \delta$

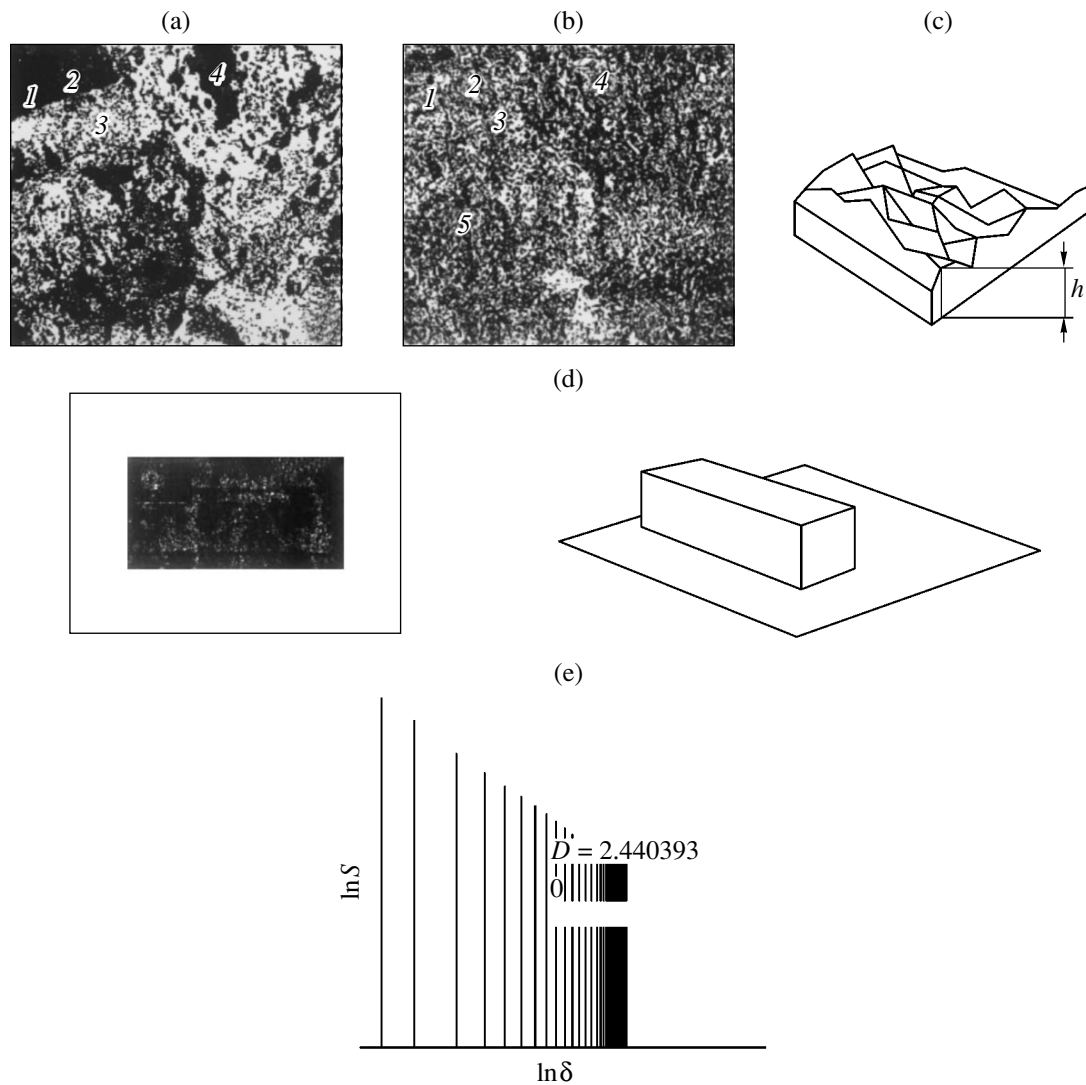**Fig. 1.** (a–d) Examples illustrating the essence of fractal processing of images and (e) spatial fractal cepstrum of the RI.

trends, especially as far as DS evolution is concerned. Bifurcation is associated with catastrophic variations or conflicts. Nonlinearity may suddenly and unpredictably change the course of processes. Between bifurcation points, an open system obeys deterministic laws; near them, the behavior of the system is governed only by fluctuations. Therefore, the effect of fluctuations in the vicinity of bifurcation points is significant. Chaotic processes beginning at the microlevel may "fly up" to the macrolevel and influence the entire system. The ability of an open system to endure under determinate chaos favors information processes irrespective of initial conditions. This statement applies to any engineering, ecological, biologic, economic, and social open system [16].

The fractal–texture relation is a basic point in fractal geometry. The relevant results are summarized elsewhere (see, e.g., [2, 17–34] and Refs. therein). Diagram 2 [27] categorizes mathematical and physical fractals

encountered in radar and radiophysical applications. To complete the discussion, it is appropriate to mention definitions also introduced in [27]. In formalized mathematical terms, the fractal is a functional map (or a set) that results from an infinite recursive process. The properties of the fractal are as follows: (i) self-similarity or invariance under scale (unlimited scaling), which means that fractals are, on average, the same on small and large scales; (ii) fractional (Hausdorff) dimension, which is strictly larger than the topological dimension; and (iii) nondifferentiability (hence, fractional derivatives and integrals). Physically, a fractal is defined as a highly irregular geometrical object (a line or the surface of a body) that possesses the property of self-similarity on a limited scale. Note also that D_0 usually designates the topological dimension of the space where a fractal of dimension D is considered.

Table 2. Classification of physicochemical processing methods

Class A	Class B	Class C
Electroerosion dimensional processing	Liquid-phase (chemical and electro-chemical) metallization	Chemical assembling
Laser processing	Metallization by liquid metal spraying	Sol–gel processes
Electron-beam processing	Plasma evaporation	Atomic layer epitaxy
Ultrasonic processing	Physical condensation	CVD processes
Electrochemical dimensional processing	Microarc oxidation	Chemical synthesis combined with various physical processes
Plasma cutting	Electroerosion synthesis of coatings (hardening)	Laser processing of films
High-frequency mechanical processing	Gas-flame hardening	Free-abrasive processing in special media
Combined methods	Laser-, ion-, and electron-beam hardening	
Laser scribing	Ion implantation	……………………

THERMODYNAMIC NONEQUILIBRIUM OF SURFACE PROCESSING AND FRACTALITY OF SURFACES PROCESSED

A better insight into the process of surface relief formation (specifically, into a roughening mechanism) that is described in terms of fractal analysis may be gained through considering advanced surface processing techniques. They differ widely in ultimate goals and underlying physical mechanisms. This is not surprising: the rapidly burgeoning machine and instrument building industries are now producing sophisticated intricate small-size mechanisms and devices with severe requirements placed upon the surface layer properties and reliability.

In terms of plausible surface roughening mechanisms, the variety of physicochemical processing methods may be subdivided into three basic classes (depending on the aim of application; see Table 2): class A, methods that condition the surface by removing surface layers of the material with its bulk remaining intact; group B, those that apply (deposit) another material on the surface with subsequent diffusion of the coating material into surface layers, which modifies the surface properties (the bulk of the host material remains unchanged); and class C, methods with which structures are formed that may serve both as a matrix (host material) and as a surface (owing to ultrasmall sizes of these structures, as in nanotechnology).

Considering the physics of the methods belonging to class A and class B allows us to infer that many of them (electroerosion, electron-beam, and laser processings; microarc oxidation; and gas-flame, laser, ion, etc., hardenings) share a number of processes. Among them are plasma processes; that is, generation and loss of charged particles in a definite space region that constitute a charge-carrier gas. Strictly speaking, the very fact that a plasma is confined in space entails its thermodynamic nonequilibrium [35]. The nonequilibrium of a plasma shows up in dissipative (collision or collision-free) wave processes, such as absorption of waves by charge carriers with subsequent decay of the waves.

Along with the nonlinear wave phenomena mentioned above, plasmas contain other regular nonlinear waves, solitons. Sometimes, solitons are arranged into structures in plasmas and we may speak of self-organization of nonlinear waves (synergetics); sometimes, soliton chaos may be observed. The case in point is thus dynamic processes causing the formation of dissipative and self-organizing structures (both in the bulk and
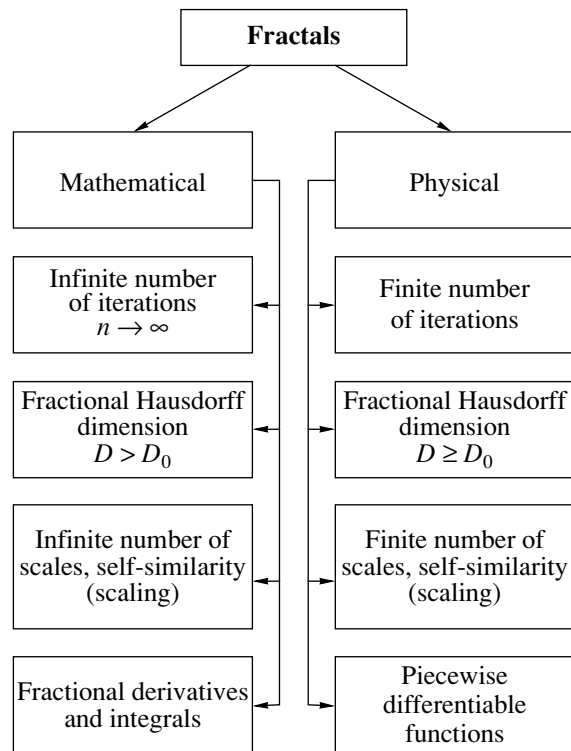

Diagram 2. Categorization and structure of fractal sets (signatures).

directly on the surface), which show up as "unusual" features on the surface (Table 1).

The processes attendant to continuous-wave laser processing are characterized by a low-threshold optical breakdown (lasting 10^{-8} s) in the zone being processed, which generates a plasma cloud (for a time of 10^{-6}–10^{-5} s) [36]. Interaction between the material and heat fluxes with a density of 10^2–10^5 W/m^2 in the active (N_2, CO_2, H_2, CO, or CH_4) or inert (Ar, He, or Xe) environment leads to the formation of a heat source on the surface, which contributes to the specific morphology due to various instabilities.

Under pulsed laser processing (the pulse width on the order of 10^{-9}–10^{-6} s), the material experiences both thermal and dynamic breakdown when subjected to a thermal shock lasting from 10^{-10} to 10^{-7} s [37]. A set of breakdown centers arising during the dynamic breakdown may be considered as a fractal cluster [38]. It was noted that the pulsed heating loads the material dynamically: the one-site mechanism of failure changes gradually to the typical multisite mechanism. The kinetics of dynamic failure may be viewed as being governed by the process of failure center concentration at different scale levels (this process being of percolation nature).

The essence of electroerosion is the breakdown of an insulating medium in an applied electric field for a time of 10^{-8}–10^{-7} s followed by the avalanche-like growth of the number of charge carriers (the formation of a plasma) [39]. This process is also synergetic, since it combines a variety of concurrent dynamic processes. These are (i) bombardment of the surface by charged particles ($W_a^{(1)}$ and $W_c^{(1)}$), (ii) gaskinetic (thermal) bombardment by the particles constituting the discharge channel ($W_a^{(2)}$ and $W_c^{(2)}$), (iii) deceleration of the torches of the vapors arriving at the anode from cathode and vice versa ($W_a^{(3)}$ and $W_c^{(3)}$), (iv) thermal radiation ($W_a^{(4)}$ and $W_c^{(4)}$, and (v) the action of the volume heat source ($W_a^{(5)}$ and $W_c^{(5)}$); that is,

$$W_a = W_a^{(1)} + W_a^{(2)} + W_a^{(3)} + W_a^{(4)} + W_a^{(5)},$$
$$W_c = W_c^{(1)} + W_c^{(2)} + W_c^{(3)} + W_c^{(4)} + W_c^{(5)}, \tag{6}$$

where W is the energy of the related process.

In the case of laser hardening of alloys [40], it was found that concentrated radiation fluxes incident on the surface may initiate either fluctuation (disordered) nucleation and growth of a new phase (indicating that the process is chaotic) or a cooperative (ordered) shift of atoms by a small distance. The latter process may be associated with self-organization, which is consistent with the idea of synergetics in materials science.

It is known that electrode processes in electrolytes are responsible for material removal in dimensional electrochemical processing. These processes, which cause all the changes on the metal surface when its potential is nonequilibrium, may be viewed as ordinary heterogeneous chemical reactions initiated at the liquid–solid interface by a potential difference between the electrodes [41]. Anodic dissolution is a multistage process including the concurrent processes of the formation of the double electric layer (within the dense part of which electrochemical reactions proceed), supply and removal of reacting particles through the diffuse part of the double layer, and formation of new substances (reaction products).

When the current passes through an electrochemical system, electrode potentials φ_a and φ_c become offset from their equilibrium values (the phenomenon of electrode polarization) by $\Delta\varphi$. Essentially, the polarization means that the electrochemical process is irreversible. It may proceed at a reasonable rate only if the polarization (overpotential) is high, i.e., when the total current is roughly equal to the current of the dominating process. Then, the potential shift (concentration overpotential) is given by the Tafel equation

$$-\Delta\varphi = a + b\ln J_c, \tag{7}$$

where a and b are constants and J_c is the cathode current.

The fractality of the steel corrosion kinetics was observed in [42], where the corrosion of structural materials in sodium was studied. It was found that the dissolution of a film coating the metal proceeds via the coalescence of individual corrosion pits. Remarkably, the extension of the corrosion area does not change the fractality of the boundary of corrosion pit traces. The empirical expression for the corrosion zone circumference appears as

$$L(l) \approx l^{-D}, \tag{8}$$

where $D \approx 1.20$ is the fractal dimension of the circumference.

Considering the processes involved in dimensional electrochemical processing from the synergetic standpoint, we note that they all are dynamic, obey the principle of current and local equilibrium and the principle of mosaic nonequilibrium thermodynamics, and correlate well with evolutionary equation (2).

As for CVD processes (including metallorganic CVD (MOCVD) and others), as well as those attendant to ion bombardment of the solid surface, it was found [4, 43] that they typically result in the formation of cones, facets, and/or layered columnar structures on the surface and also favor the growth of large crystals and needle crystals (whiskers). Such features of the relief are attributed to the nonlinear kinetics under highly nonequilibrium conditions (for example, when the temperature varies from 150 to 1000°C, which is typical of CVD processes). Lattice imperfections in the material make a significant contribution to the formation of these features, since the imperfections influence the density of the features and the uniformity of their dis-
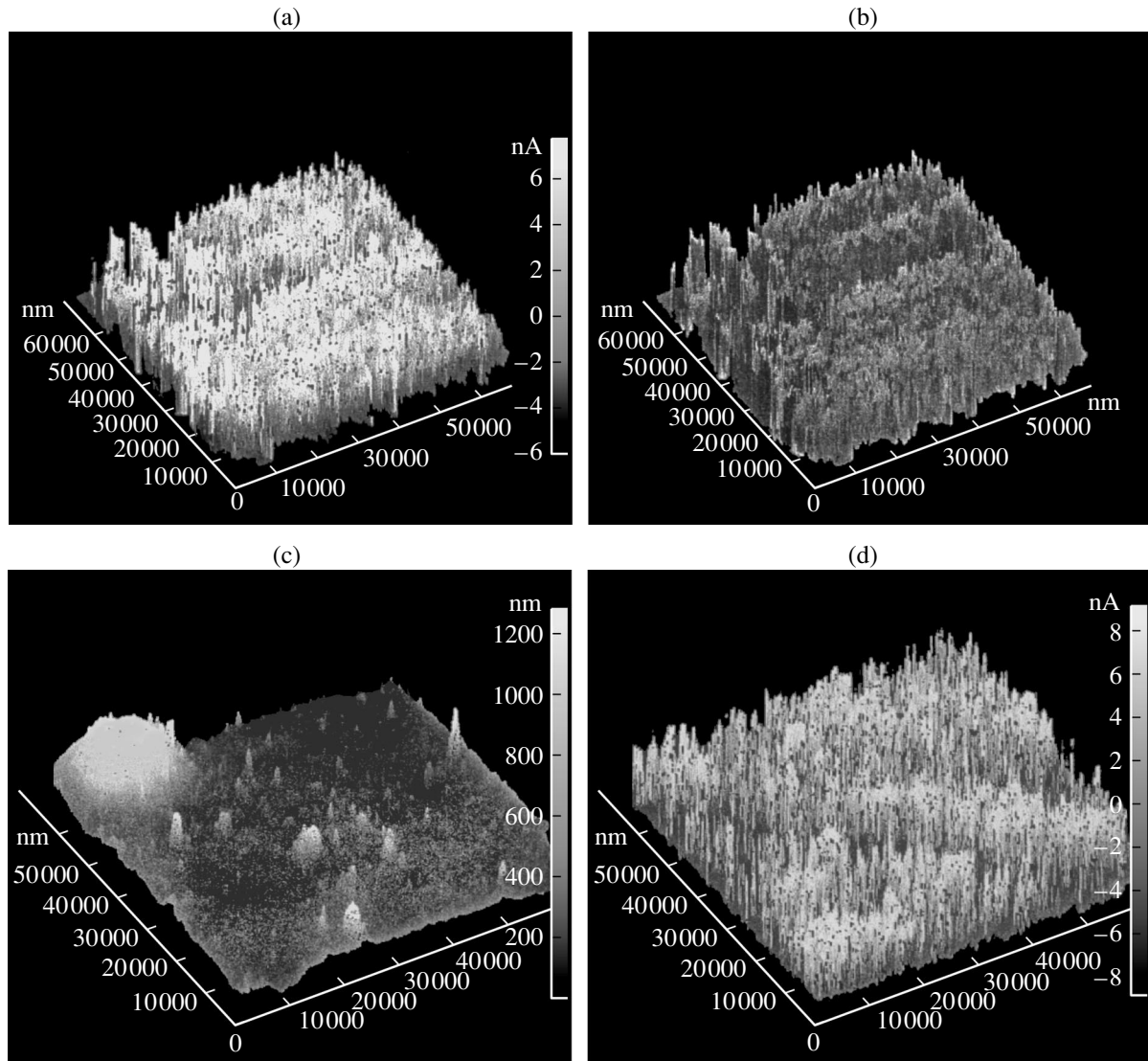
Fig. 2. SPM (3D) images of the surfaces (a–e, g, h, i) subjected to microarc oxidation (regimes 1, 2, 15, 12, 5, 9, and 7, respectively), (f, k, l) coated by means of erosion synthesis (regimes 1, 3, and 5, respectively), (m) subjected to diamond grinding, and (n) subjected to diamond turning followed by electrochemical etching.

tribution over the area being processed. In addition, they have an effect on the onset of dissolution and deposition [45].

Next, it was found [44] that CVD is a multistage process including mass-and-heat transfer, adsorption and desorption, the reaction of decomposition of metallorganic compounds, formation of the solid phase, and crystallization. The shape of surface features depends on whether thermodynamically stable or unstable conditions prevail.

Thus, most of the processes involved in physicochemical processing are consistent with the principles of synergetics and can be referred to so-called nonequilibrium techniques. Processes of such a kind cause, as a rule, the formation of dissipative structures, which require that the energy (electrical, thermal, or light) be

constantly delivered from the outside. The energy can be delivered by special devices, such as power supplies, electron guns, pump systems, etc. Thus, highly nonequilibrium processes should be described in terms of the fractal theory; specifically, the surface roughness should be quantitatively characterized by introducing fractal (fractional) dimension D or fractal signatures.

The roughening mechanisms involved in the processes mentioned above differ radically from those encountered in cutting. The physicochemical processes imply application of various fields (electromagnetic, thermal, and others), as well as high temperatures and pressures, which simultaneously act on the surface in a complicated manner. The methods used in these processes can be classed as methods of group action: conditioning of the surface starts almost instantly and
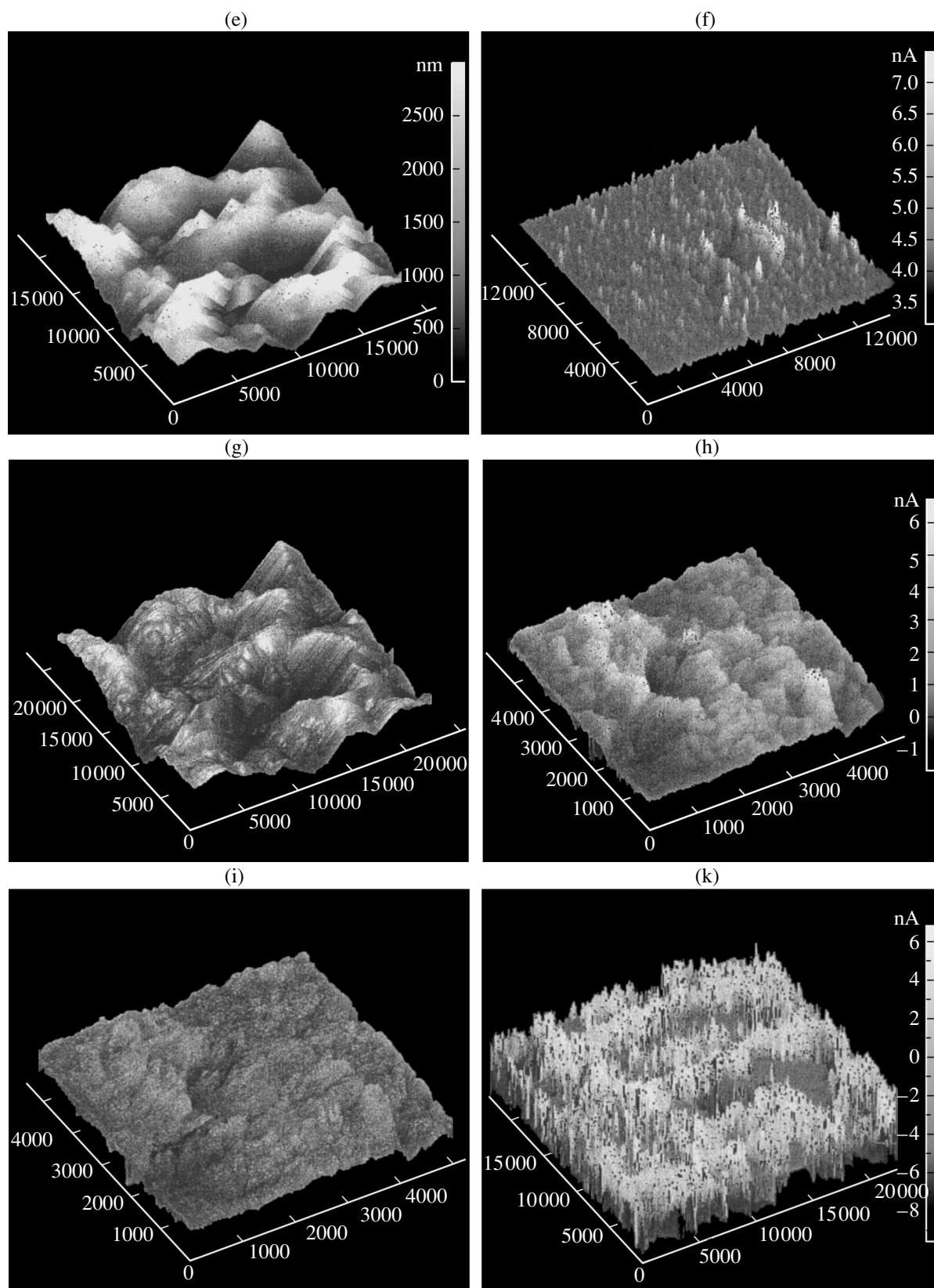
(e)

(f)

(g)
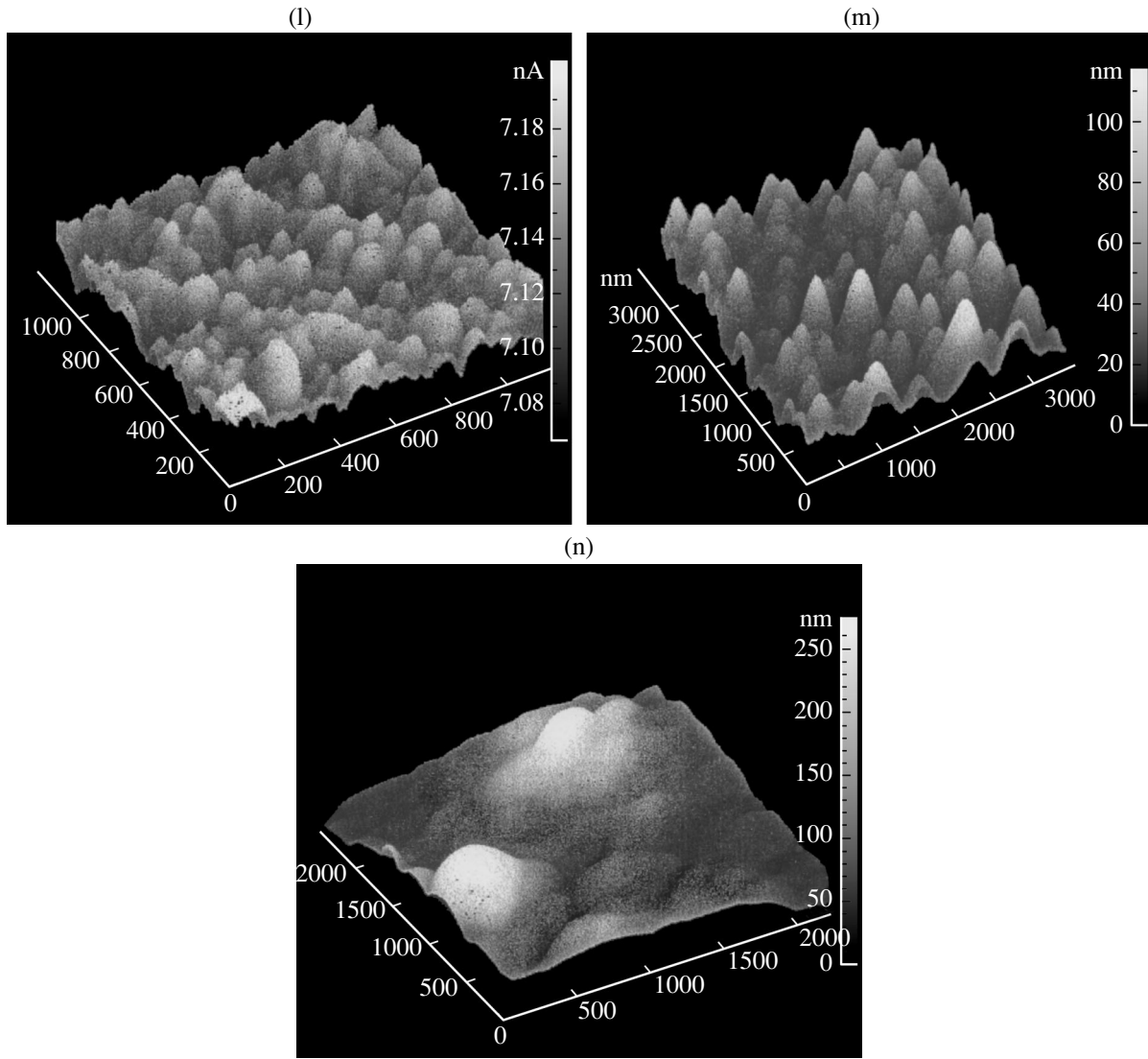
(h)

(i)

(k)

Fig. 2. (Contd.)

(l)

(m)

(n)

Fig. 2. (Contd.)

simultaneously at many "seeding" sites chaotically distributed over the surface.

For the surface relief, both the environment and the inner material structure are responsible. The environment may be electrolyte solutions, insulating media, inert and chemically active gases, metal vapors, etc. External parameters are the temperature, flow velocity of the electrolyte, pressure, and applied fields. As was noted above, lattice imperfections in the material influence the surface relief formation. According to the theory of fractal materials science [6], lattice imperfections both in the volume and on the surface are due to the mechanisms of self-organization of dissipative structures, which are spontaneously reconfigured near bifurcation points and are attributed to the fractal properties of the material. These mechanisms cause both microdefects (vacancies, dislocations, slip bands, grain boundaries, and atomic clusters) and macrodefects

Table 3. Fractal dimensions D and D_s of the surfaces processed

Processing	Regime no.	Fractal dimension D	Fractal dimension D_s	Variance Σ^2
Diamond grinding	5	2.117	1.985	0.024
	8	2.288	2.022	0.026
Microarc oxidation	1	2.102	2.065	0.029
	2	2.279	2.179	0.072
	16	2.420	2.215	0.070
	23	2.392	2.157	0.075
Electroerosion synthesis of coatings	1	2.445	2.237	0.077
	2	2.303	2.179	0.032
	10	2.536	2.136	0.498
Diamond turning followed by electrochemical etching	1	2.085	1.971	0.025
	4	2.097	1.978	0.038
	8	2.174	2.132	0.078

(a)

(b)

(c)

(d)

(e)

(f)

(g) (h)

(i)
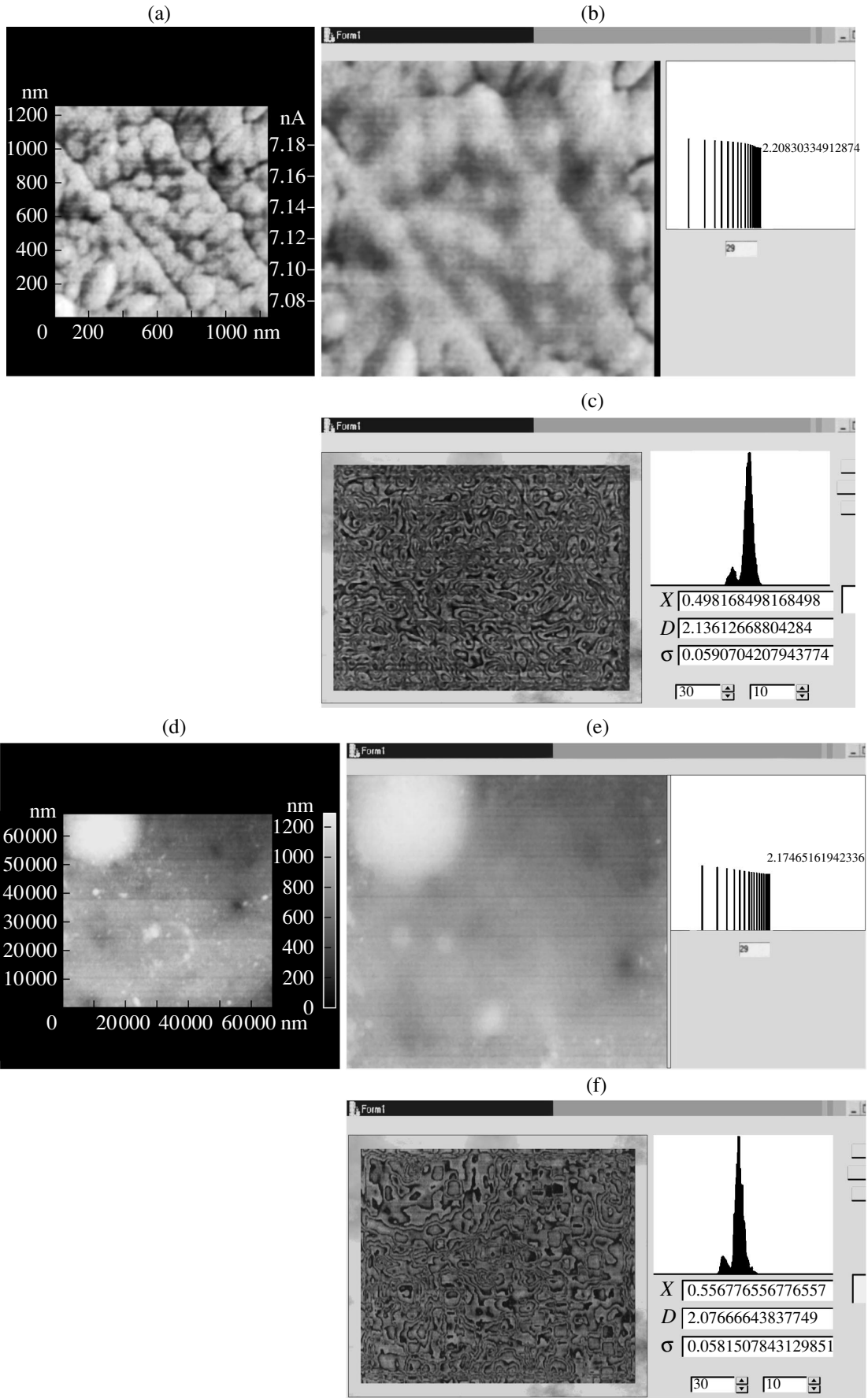
Fig. 3. (Contd.)

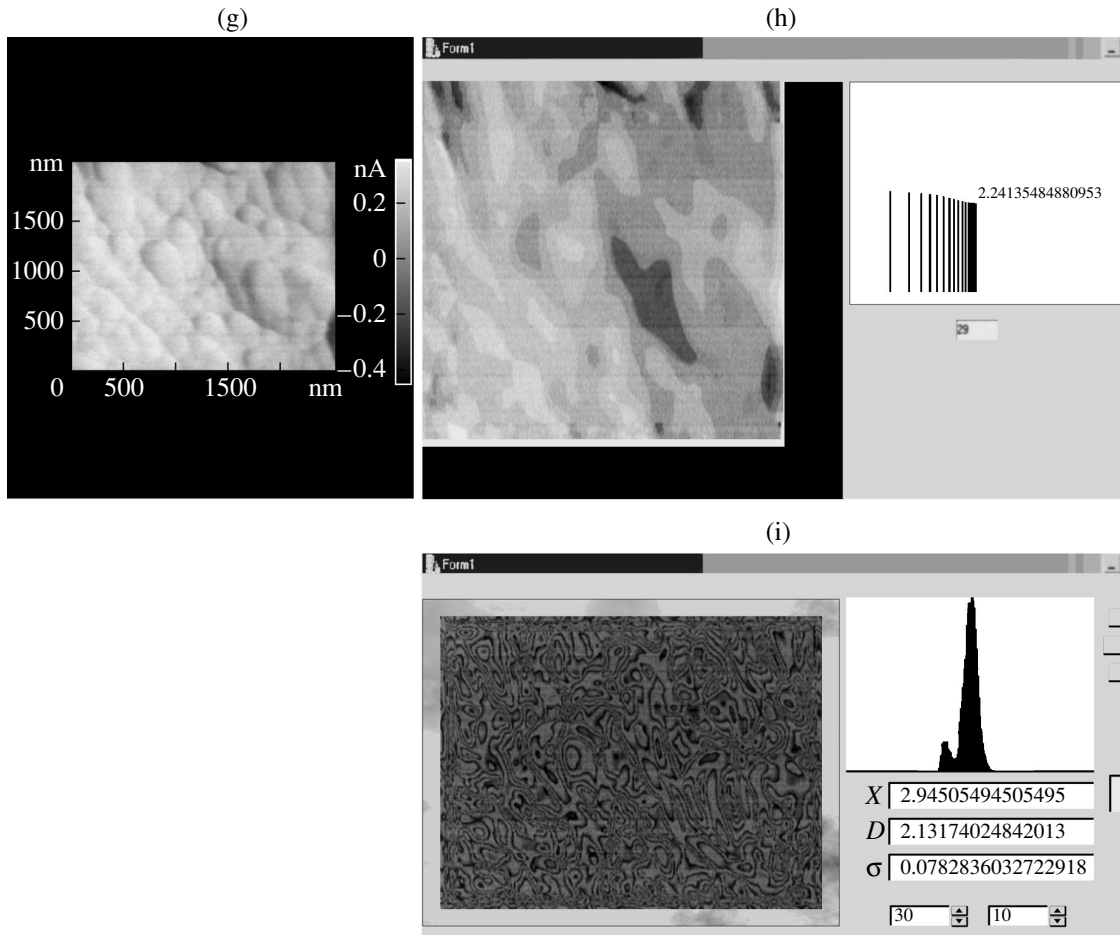(mosaic blocks and dislocation ensembles) that were generated at previous stages of processing (casting, rolling, annealing, etc.) to emerge on the surface. The defects emerging on the surface specify the formation of the above-mentioned surface features at further processing. For example, the nucleation of atomic and/or molecular clusters that then coalesce to form a continuous coating and dissolution (pitting) start at sites energetically more favorable for interaction with the working medium under processing conditions.

As processing goes on, the fractal properties of the material and processing medium (e.g., electromagnetic fields) superpose to form the final surface relief, specifically, the roughness. Under these conditions, the surface should be considered as an expanding stochastic cluster system offering the property of self-similarity (scaling).

EXPERIMENTAL INVESTIGATION OF THE SURFACE FRACTALITY

To verify the approach suggested, we studied the surfaces of AK-12M and AD-16 alloys and 45 steel processed by diamond grinding, microarc oxidation, electroerosion synthesis (hardening), and diamond turning with subsequent electrochemical etching.

Prior to processing, the surfaces were examined under a scanning probing microscope (NT-MDT Co., Zelenograd, Moscow). The images of the surfaces are represented in Fig. 2 in the $512 \times 655 \times 256$ bmp format.

Fractal dimension D and the fractal signatures (fractal cepstrum) of the surfaces were measured with a technique developed at the Institute of Radio Electronics and Electronics, Russian Academy of Sciences [2, 17–34]. Original techniques for measuring the fractal dimension and fractal signatures of multidimensional

Fig. 3. Fractal analysis of the surfaces subjected to (a) diamond turning, (d) diamond turning followed by etching, and (g) diamond turning followed by oxidation; (b, e, h) fractal signatures of the surface images (fractal cepstrum); and (c, f, i) field (on the left) and experimental distribution of local fractal dimensions D (on the right).

(a)

(b)

(c)

(d)

(e)

(f)
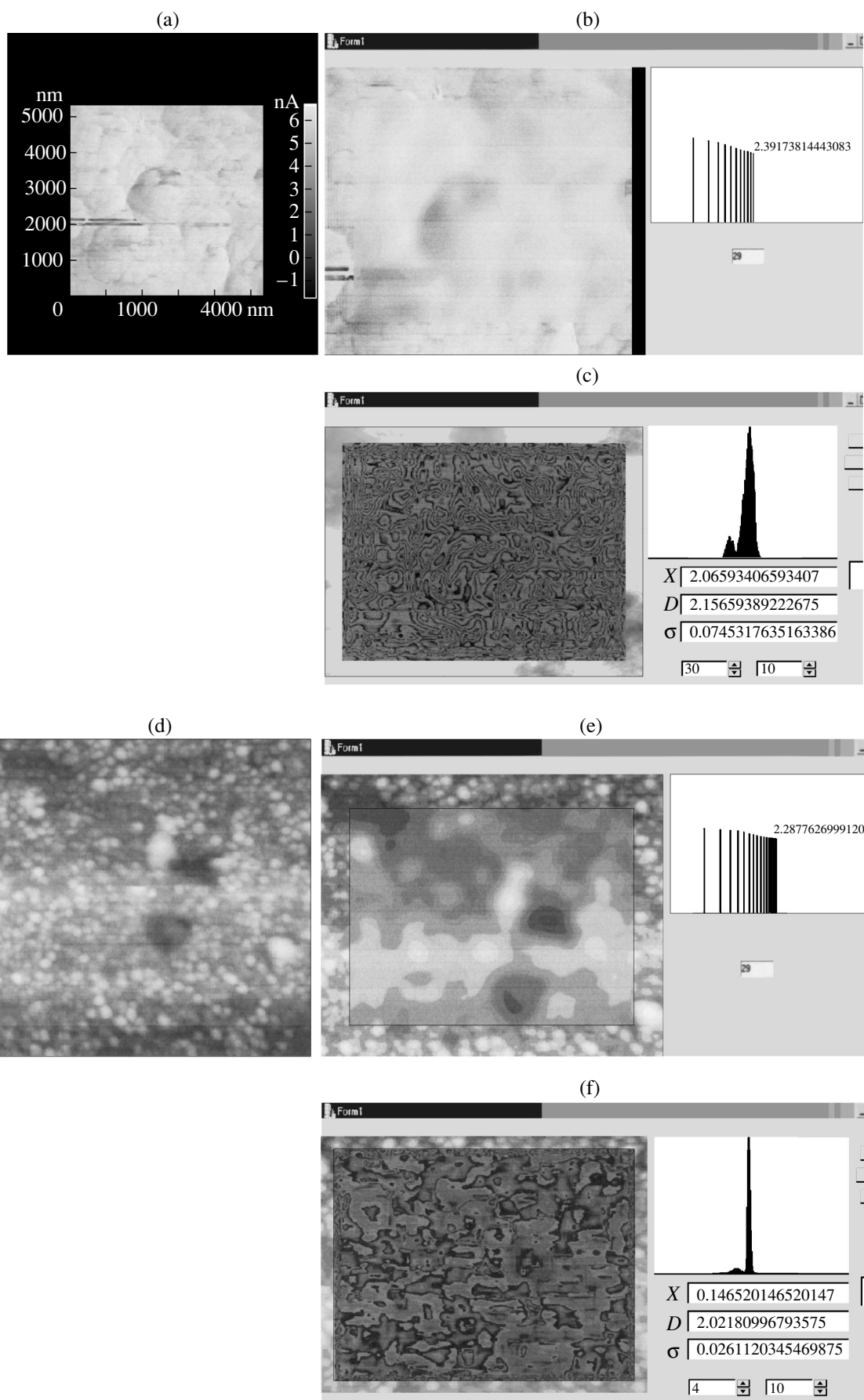
Fig. 4. (Contd.)

stochastic signals are comprehensively discussed in [27, 34]. The experimental data are presented in Figs. 3 and 4. Some of the results were reported in [46]. The fractal signatures (Figs. 3b, 3e, 3h, 4b, 4e, and 4h) show the mean value of D and its rms deviation σ. Setting desired range ΔD of local fractal dimensions in the clustering mode, we see a field of fractal dimensions on the monitor where all $D_i \in \Delta D$ are shown black.

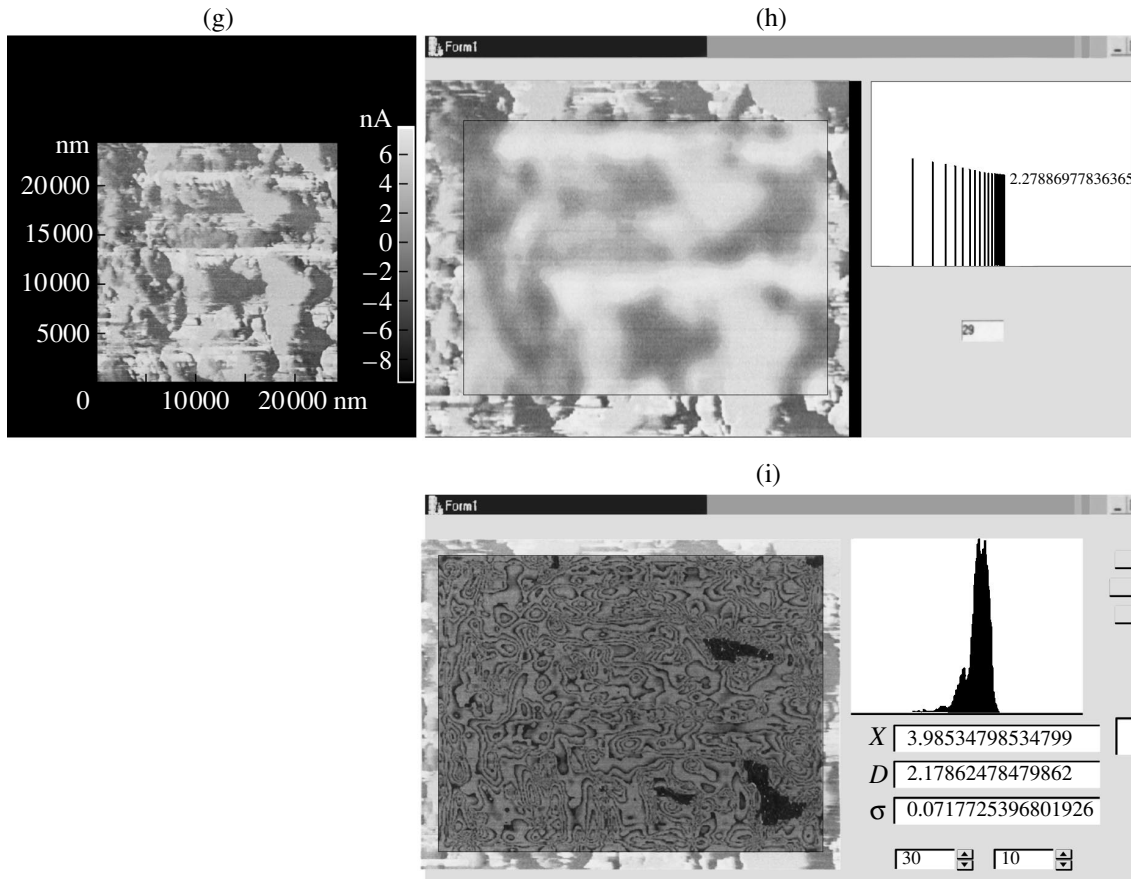When implemented in software, the techniques for measuring fractal characteristics display experimental distributions of local fractal dimensions directly on the monitor Figs. 3c, 3f, 3i, 4c, 4f, 4i). The color of each point on the sample image is in proportion to D_i. Measurements are represented in a wide palette of pseudocolors. On the 2D images of the surface (the left-hand side of Figs. 3c, 3f, 3i, 4c, 4f, 4i), the pseudocolors mark all clusters with given D.

In the case of diamond turning (Fig. 3), experimental spatial distributions of D always have extended "heavy" tails (Paretians) [2, 24, 26, 27, 29, 31, 33], which indicates the hierarchic organization of the

microrelief. Subsequent procedures (etching or oxidation) considerably widen the range of average fractal dimensions. Regime 2 of microarc oxidation (Figs. 4a–4c) give D smaller than that in regime 7 (Figs. 4d–4f). Diamond grinding gives a wider range of local fractal dimensions, as clearly follows from the experimental data (Figs. 4g–4i).

Thus, the numerical investigation of the surfaces processed lets us argue that there are surface areas with fractal properties. This statement is corroborated by the fractal signatures and probability distributions with heavy tails (Figs. 3 and 4), as well as by fractional values of the dimension both on the images (D) and in the spectrum (D_s) (see Table 3). It should be noted that the fractal dimension of the image must satisfy the conditions $D, D_s \geq 2$ [2, 34]. Values $D_s < 2$ are due to an algorithmic inaccuracy and must be considered, in our case, as a fast approach of D_s to the topological dimension of the surface.

Fig. 4. Fractal analysis of the surfaces subjected to (a, d) microarc oxidation and (g) diamond grinding; (b, e, h) fractal signature of the surface images (fractal cepstrum); and (c, f, i) field (on the left) and experimental distribution of local fractal dimensions D (on the right).

CONCLUSIONS

Based on experimental data obtained in earlier experiments, a universal and illustrative method of estimating the surface quality using different fractal characteristics is suggested. The existence of fractal areas on the surfaces processed counts in favor of the idea that the surface forms (at the micro-, mesa-, and macrolevels) by the cluster mechanism, which is related to dynamic nonequilibrium processes. Therefore, the development of a fractal synergetic model of surface relief (roughness) formation that takes into account self-organization phenomena inherent in group processing techniques seems to be of special interest. The surface roughness and surface layer should be considered as a single hierarchic fractal or multifractal system. Such a model would make it possible not only to predict but also to control roughening mechanisms and physicomechanical properties of the surface layer. In this work, only one approach to tackling the challenging problem (estimating the roughness of surfaces processed by nonequilibrium techniques with the aim of controlling the geometric parameters of the roughness and providing desired operational performance of products) is suggested. Effective numerical methods of measuring fractal characteristics (fractal signatures) that were tried experimentally [2, 17–34] may find wide application in physical, engineering, astronomical, biological, and medical research.

REFERENCES

1. N. B. Demkin and E. V. Ryzhov, *Surface Condition and Contact of Machine Parts* (Mashinostroenie, Moscow, 1981) [in Russian].

2. A. A. Potapov, *Fractals in Radiophysics and Radar* (Logos, Moscow, 2002) [in Russian].

3. ISO Standard *4287-98: Geometrical Characteristics of Products. Surface Structure: Profiling Method. Terms, Definitions, and Structure Parameters* (1999).

4. V. T. Syrkin, *CVD Method* (Nauka, Moscow, 2000) [in Russian].

5. A. N. Kolmogorov, Izv. Akad. Nauk SSSR, Ser. Mat., No. 3, 355 (1937); *Selected Works*, Vol. 2: *Probability Theory and Mathematical Statistics* (Nauka, Moscow, 1986; Kluwer, Dordrecht, 1992), pp. 178–182.

6. V. S. Ivanova, A. S. Balankin, I. Zh. Bunin, and A. A. Oksagoev, *Synergetics and Fractals in Materials Science* (Nauka, Moscow, 1994) [in Russian].

7. V. S. Ivanova and V. U. Novikov, Nelineinyi Mir **2** (3), 197 (2004).

8. *Strange Attractors: A Collection of Articles*, Ed. by Ya. G. Sinaĭ and L. P. Shil'nikov (Mir, Moscow, 1981) [in Russian].

9. P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations* (Wiley, New York, 1971; Mir, Moscow, 1973).

10. H. Haken, *Synergetics: An Introduction* (Springer, Berlin, 1977; Mir, Moscow, 1980).

11. W. Ebeling, *Strukturbildung bei Irreversiblen Prozessen* (Teubner, Leipzig, 1976; Mir, Moscow, 1979).

12. G. Nicolis and I. Prigogine, *Self-Organization in Non-Equilibrium Systems* (Wiley, New York, 1977; Mir, Moscow, 1979).

13. Yu. L. Klimontovich, *Statistical Theory of Open Systems* (Yanus, Moscow, 1995–2001; Kluwer Academic, Dordrecht, 1995).

14. V. V. Kadomtsev, *Dynamics and Information*, 2nd ed. (UFN, Moscow, 1999) [in Russian].

15. A. N. Lagar'kov and M. A. Pogosyan, Vestn. Ross. Akad. Nauk **73**, 779 (2003).

16. A. A. Potapov, Nelineinyi Mir **1** (1–2), 3 (2003).

17. B. V. Bunkin, A. P. Reutov, A. A. Potapov, *et al.*, in *Challenging Problems in Radar*, Ed. by A. V. Sokolov (Radiotekhnika, Moscow, 2003) [in Russian].

18. S. A. Podosenov, A. A. Potapov, and A. A. Sokolov, *Pulsed Electrodynamics of Broadband Radio Systems and Fields of Coupled Structures*, Ed. by A. A. Potapov (Radiotekhnika, Moscow, 2003) [in Russian].

19. A. A. Potapov, Zarubezhn. Radioélektronika, No. 6, 3 (2000).

20. A. A. Potapov and V. A. German, Radiotekh. Élektron. (Moscow) **45**, 946 (2000).

21. A. A. Potapov and V. A. German, in *Proceedings of the 1st International Workshop on Mathematical Simulation of Physical Processes in Inhomogeneous Media, Mexico, 2001*, pp. 44–46.

22. A. A. Potapov and V. A. German, in *Proceedings of the 4th Symposium on Physics and Technology of Millimeter and Submillimeter Waves, Kharkov, 2001,* Vol. 1, pp. 268–270.

23. A. A. Potapov and V. A. German, Pis'ma Zh. Tekh. Fiz. **28** (14), 19 (2002) [Tech. Phys. Lett. **28**, 586 (2002)].

24. A. P. Reutov, A. A. Potapov, and V. A. German, Nelineinyi Mir **1** (1–2), 12 (2003).

25. A. A. Potapov, Elektromagn. Volny Elektron. Sistemy **8**, 26 (2003).

26. A. A. Potapov, Radiotekhnika, No. 8, 55 (2003).

27. A. A. Potapov, Radiotekh. Élektron. (Moscow) **48**, 1101 (2003).

28. A. A. Potapov, in *Proceedings of the 1st Scientific Conference "Radio-Optic Technologies in Instrument Making Industry," Sochi, 2003*, pp. 58–61.

29. A. A. Potapov, in *Proceedings of the 3rd Extended Seminar "Artificial Intellect and High-Speed Computing in Aerospace Studies," Pereslavl-Zalesskii, 2003*, pp. 76–90.

30. A. A. Potapov, Nelineinyi Mir **2** (1), 4 (2004).

31. A. A. Potapov and V. A. German, in *Proceedings of the 10th International Scientific Conference "Radar, Navigation, and Communication," Voronezh, 2004*, pp. 318–341.

32. A. A. Potapov, in *Proceedings of the 1st International Scientific Conference "Strategies of Fast Development of Russia: The Unity of Self-Organization and Management", Moscow, 2004*, Vol. 3, Chap. 3, pp. 129–135.

33. A. A. Potapov, in *Proceedings of the 2nd IEEE International Conference on Circuits and Systems for Communications (ICCSC-2004), Moscow, 2004*.

34. A. A. Potapov and V. A. German, Radiotekh. Élektron. (Moscow) **49** (12) 2004 (in press).

35. A. F. Alexandrov and A. A. Rukhadze, *Lectures on Electrodynamics of Plasmalike Media* (Mosk. Gos. Univ., Moscow, 2002), Part 2 [in Rissian].

36. A. A. Uglov, in *Physicochemical Treatment of Materials by Concentrated Energy Fluxes* (Nauka, Moscow, 1989) [in Russian].

37. A. Ya. Uchaev, E. K. Bonyushkin, S. A. Novikov, and N. I. Zavada, *Spall of Metals under Fast Volume Heating Conditions* (TsNIIatominform, Moscow, 1999) [in Russian].

38. E. K. Bonyushkin, N. I. Zavada, L. A. Platonova, *et al.*, in *Nonlinear Dynamics, Fractals, and Neuron Networks in Systems and Processes*, Ed. by Yu. G. Kabaldin (Vladivostok, 2002), pp. 123–171 [in Russian].

39. B. N. Zolotykh, *Physical Foundations of Electrical and Electrochemical Processing* (MGU, Moscow, 1975), Chap. 1 [in Russian].

40. A. N. Kokora and E. N. Sobol', in *Physicochemical Treatment of Materials by Concentrated Energy Flows* (Nauka, Moscow, 1989), pp. 54–65 [in Russian].

41. F. V. Sedykin, *Dimensional Electrochemical Processing of Machine Parts* (Mashinostroenie, Moscow, 1976) [in Russian].

42. N. I. Zasorin, V. V. Kuzin, V. A. Morozov, and A. L. Shimnevich, Preprint FTI (Obninsk, 1996).

43. *Surface Modification by Laser, Ion and Electron Beams*, Ed. by J. M. Poate, G. Foti, and D. C. Jacobson (Plenum, NATO Series, New York, 1983; Mashinostroenie, Moscow, 1987).

44. B. G. Gribov, G. A. Domracheev, B. V. Zhuk, B. S. Kaverin, B. I. Kozyrkin, V. V. Mel'nikov, and O. N. Suvorova, *Growth of Films and Coatings by MOCVD* (Nauka, Moscow, 1991) [in Russian].

45. V. I. Trofimov and V. A. Osadchenko, *Growth and Morphology of Thin Films* (Énergoatomizdat, Moscow, 1993) [in Russian].

46. V. V. Bulavkin, A. A. Potapov, O. F. Vyacheslavova, and V. A. German, in *Proceedings of the International Symposium on Fractals and Applied Synergetics, Moscow, 2003*, pp. 250–253.

Translated by V. Isaakyan

GASES
AND LIQUIDS

Simulation of the Electrohydrodynamic Flow Pattern in an Asymmetric System of Electrodes

Yu. K. Stishkov, A. V. Buyanov, and A. S. Lazarev

Research Institute of Radiophysics, St. Petersburg State University,
Ul'yanovskaya ul. 1, Petrodvorets, St. Petersburg, 198504 Russia

e-mail: pelab@niirf.spbu.ru

Received August 10, 2004

Abstract—Computer-aided simulation of the electrohydrodynamic flow pattern in an asymmetric system of electrodes (configured as "narrow depression under plane" or "blade above plane") is carried out. The objective of the simulation is to reveal a correspondence between the kinematic flow pattern and the distribution of driving (Coulomb) forces. The kinematic pattern of the flow simulated is compared with the results of experimental data processing. Good agreement between the simulation results and experimental data is found. © 2005 Pleiades Publishing, Inc.

INTRODUCTION

The features of electrohydrodynamic flows have been described at length in a variety of works (see, e.g., [1, 2]). The range of existence of these flows is bounded from below by a threshold voltage. At voltages slightly exceeding the threshold, the flows are weakly developed; i.e., they are localized near the active electrode. Then, as the applied voltage grows, a fully developed electrohydrodynamic flow completely occupying the electrode gap arises. In both developed and undeveloped regimes, the flow is usually of a laminar (parallel-stream) character and is directed from the "wire" (active) electrode to the plane (passive) one. Comparative analysis of developed electrohydrodynamic flows at various voltages has not revealed appreciable distinctions in their pattern. The dependence of the average velocity in the central stream on the voltage between the electrodes is either a power-type function, $v \sim U^k$ ($k > 2$), for the undeveloped flow or a linear function ($k = 1$) for the developed flow. In [1], an analytically convenient dimensionless form of the dependence $V = V(U)$ has been proposed, where average drift velocity of ions $V_0 = bU/l_0$ (where b is the mobility of ions and l_0 is the electrode gap length) is taken as a unit of measurement of the velocity. In this case, the voltage dependence of the average velocity takes the form $\mathrm{Re_{e1}} = V/V_0 = f(U)$. Quantity $\mathrm{Re_{e1}}$ is usually called the electric Reynolds number. Function $\mathrm{Re_{e1}} = f(U)$ has a threshold, which is usually higher than the voltage of transition to the supraohmic part of the *I–V* characteristic. In the undeveloped flow, the electric Reynolds number grows very rapidly; in the developed flow, it is constant.

Recently, a procedure for computer-aided processing of electrohydrodynamic flow observations has been developed [3], allowing one to reconstruct the distributions of the velocities and accelerations in the electrode gap. Such a processing visualizes the kinematic and dynamic patterns of the flows. It has been noted [4, 5] that the region occupied by an electrohydrodynamic flow can be conditionally subdivided into several zones.

(1) In the immediate vicinity of the electrodes, there is a thin adherent layer of the quiescent liquid. Within this layer, the liquid is structured.

(2) The zone near the active electrode where the liquid is intensely accelerated by the electric field. The length of this zone is 0.1–0.2 of the electrode gap. In this zone, the velocity isolines concentrate and the acceleration vectors are aligned with the flow direction or are directed at an acute angle to the flow.

(3) The zone of smoothly varying flow (the so-called central stream of electrohydrodynamic flow) or the parallel-stream part of the flow (the streamlines in this zone are essentially straight lines parallel to each other).

(4) The zone of flow stagnation, where the acceleration vectors are directed oppositely to the flow.

(5) The layer of quiescent liquid in the immediate vicinity of the surface of the passive electrode.

Studies of electrohydrodynamic flows under different conditions have concerned mainly the sizes of and flow patterns in those zones. As was already mentioned, the average electric Reynolds number in a fully developed flow usually lies within 10–20; i.e., the ions outside the quiescent near-electrode zone are "frozen" into the liquid. Therefore, there are good grounds to believe that a charge injected from the surface of the wire elec-

trode will be entrained by the flow into the electrode gap and the charged liquid region will have the form of a narrow strip between the electrodes. Certainly, the charge distributions in the liquid may vary depending on the flow conditions and the conductivity of the liquid; however, these variations are insignificant in a fully developed flow.

The objective of this paper is to establish a correspondence between the space charge distribution in the liquid and the kinematic pattern of the flow. To do this, we carry out a computer-aided simulation using the well-known software packages ANSYS and MatLab. When simulating electrohydrodynamic flows, one runs into a number of difficulties. First, the distributed electric charge has an appreciable effect on the electric field distribution, which, in this case, has as yet unstudied intriguing features and is of independent interest. Second, electrohydrodynamic flows are caused by a volume electric force. In classical fluid dynamics, such flows are poorly known.

For convenient comparison of data obtained in numerical and full-scale experiments, the dimensions of the model were taken to be comparable to the size of the cell in which the experiments were carried out; for the same reason, the liquid of choice is transformer oil. We simulated a steady-state electrohydrodynamic flow, which is the most-studied experimentally. In this case, all electric and kinematic characteristics of the flow may be taken to be time-independent. A two-step computational algorithm was applied. At the first step, a hypothetical space charge distribution is specified and the corresponding distributions of the electric field and Coulomb forces are calculated. At the second step, the electrohydrodynamic flow velocity field caused by the calculated distribution of Coulomb forces is computed. The space charge distribution was set based on semi-intuitive ideas regarding a possible flow pattern. Immediately near the surface of the cylindrical electrode, the space charge distribution is set in the form of a 0.1-mm-thick ring coaxial with the electrode. This stagnant near-electrode zone has been detected in experimental studies of electrohydrodynamic flows. In addition, we took into account the drift of the charge from the near-electrode zone into the electrode gap. In this case, the space charge distribution was set by a function of form $\rho_0 \exp\{\alpha(y)x^2 - y/\beta\}$. Parameter α specifies the width of the charged strip. Near the active electrode, the width of the strip was set to be equal to the electrode diameter. Varying parameter β, one can control the rate of decrease of the space charge density along the central stream of the flow. Parameter α was varied along the flow axis in such a way that the law of conservation of charge in the stream held true. Physically, this means the absence of charge sources and drains in the elec-

trode gap. Applying various combinations of parameters α and β, one can simulate different flow patterns.

CALCULATION OF THE ELECTRIC FORCE DENSITY DISTRIBUTION

Constructing the charge distribution is followed by calculating the corresponding distribution of Coulomb forces ρE. Throughout the simulation, the potentials on the plane electrode and cylindrical electrode were set equal to 0 V and 20 kV, respectively. At the boundaries of the cell, the normal component of the electric field was assumed to vanish. The average experimental value of the flow velocity in the central stream measured at a given voltage was taken to determine ρ_0. The electric field distribution and the electric force density were calculated using the electrostatic version of the ANSYS software package.

CALCULATION OF THE VELOCITY DISTRIBUTION IN ELECTROHYDRODYNAMIC FLOWS

The electric force density distribution obtained was substituted into the equations of hydrodynamics as the density of external forces. The boundary conditions were taken to be zero velocities at the electrodes and cell walls. The geometry of the electric and hydrodynamic problems, as well as the properties of the liquid, was assumed to remain the same and meet the experimental conditions. The hydrodynamic problem was solved by the finite-element method with the help of the ANSYS package. The calculated velocity distributions were analyzed and compared with the distributions obtained by processing the experimental data. In the numerical experiment, we studied the influence of various space charge distributions on the kinematic parameters of electrohydrodynamic flows.

Some of the authors approximate the electrodynamic flow by an immersed jet. In this case, acting forces are concentrated near the active electrode. Therefore, at the first step, we assumed that the space charge is localized near the surface of the cylindrical electrode, ignoring the entrainment of the charge by the flow (low electric Reynolds numbers). Subsequently, the charge entrainment by the flow was allowed for.

SIMULATION RESULTS

Figure 1 shows simulation results for the situation when the entire space charge concentrates in the near-electrode zone, that is, within the cylindrical ring coaxial with the electrode. Space charge isolines and the vector field of Coulomb forces (shown by the arrows) for this case are demonstrated in Fig. 1a. The electrode cross section corresponds to the inner circumference. The distribution of Coulomb forces, as well as the space charge distribution, are symmetric about the axis of the active electrode. Figure 1b shows the isolines and
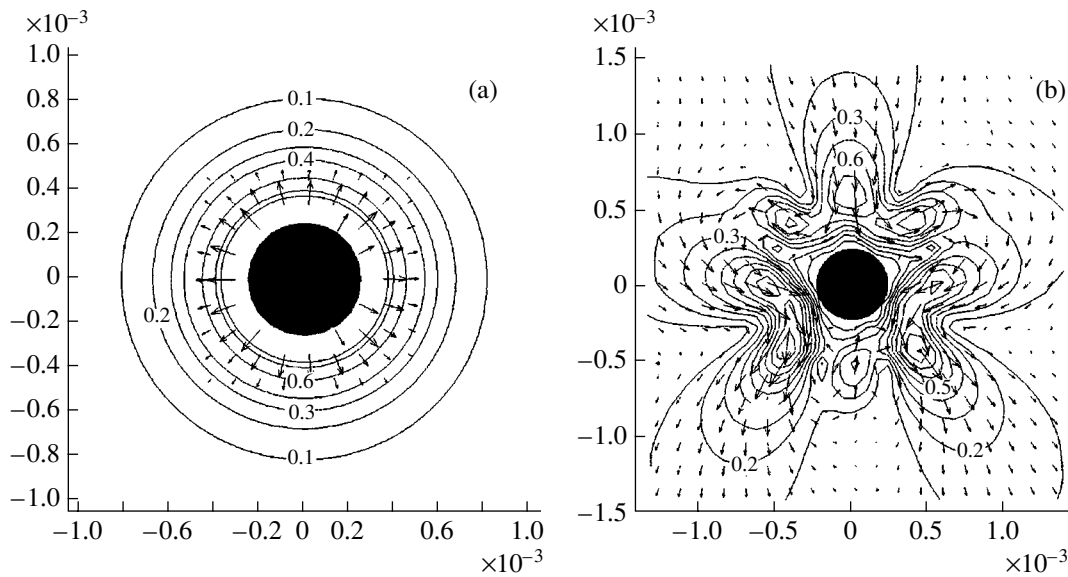
Fig. 1. (a) Electric space charge isolines and the Coulomb force vector field (shown by the arrows) for the situation where the entire space charge is in the near-electrode zone and (b) the isolines of the flow velocity and of the velocity vector field for a flow arising at such a charge distribution.
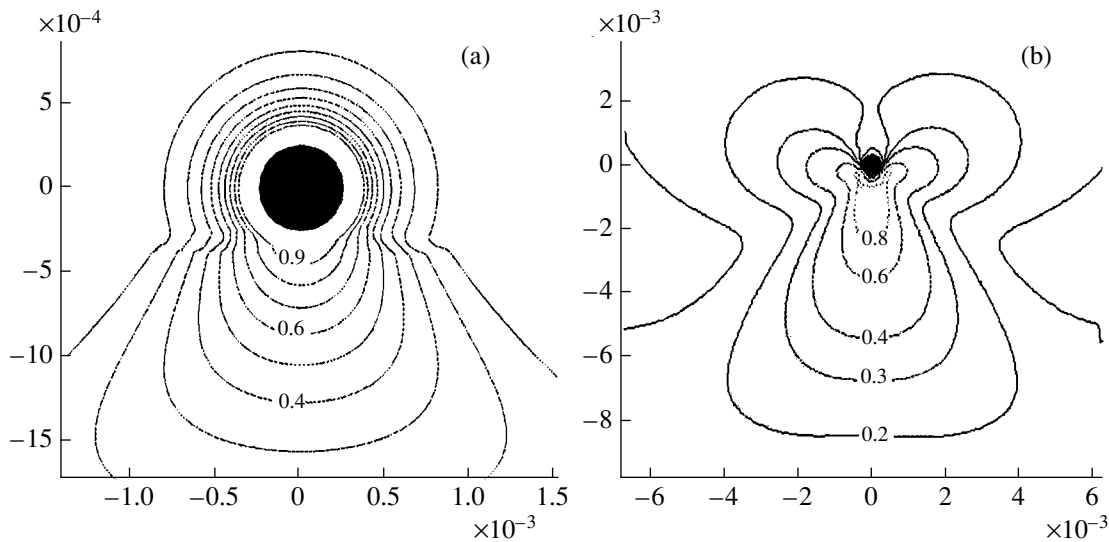
Fig. 2. (a) Space charge density distribution and (b) the flow velocity isolines at a small asymmetry ($\beta = 0.001$).

vector field of the flow velocity arising at such a charge distribution. The emerging small-scale vortices are seen to concentrate in the immediate vicinity of the electrode. They form a symmetric pattern of size much smaller than the electrode gap. As follows from the isolines, the velocity in the vortices is very low (no higher than 0.1 mm/s at a voltage of 20 kV and a charge density of 3×10^{-4} C/m³) and fades out over distances on the order of the electrode diameter. Flows of this type have not been observed in the experiments.

When a weak asymmetry appears in the charge distribution, the kinematic flow pattern changes appreciably. Figure 2 demonstrates the space charge density distributions and the flow velocity isolines in the case of a weakly asymmetric charge distribution ($\beta = 0.001$). From the charge distribution, it is seen that the charge decreases by three times toward the counterelectrode over a distance of 0.1 of the electrode gap length. Therefore, one can assume that the charge concentrates in the near-electrode zone as before is distributed asym-

metrically relative to the electrode center. However, even such an insignificant deviation from symmetry turns out to be sufficient for numerous small-scale vortices in the near-electrode zone to transform into two clear-cut ones. The flow is now directed toward the counterelectrode, and its velocity increases to 2.5 cm/s. However, the flow penetrates into the electrode gap over a distance no larger than $0.3l_0$. In the rest of the electrode gap, the flow velocity is insignificant. The zone where the liquid is accelerated is equal to the electrode diameter. It is followed by the stagnation zone. This regime has all signs of an undeveloped electrohydrodynamic flow [1] and is characterized by low values of the average electric Reynolds number. In this situation, the electrical field strength distribution is seen to be little affected by the space charge outside the small near-electrode zone. The electric forces act over a small range equal to the electrode size. Thus, in this regime, the electrohydrodynamic flow is similar to an immersed jet.

When parameter β increases, the asymmetry rapidly grows and the flow occupies an ever expanding part of the electrode gap. At $\beta > 0.003$, the electric force distribution exhibits a portion where the forces are uniformly localized and the kinematic flow pattern takes the signs of a fully developed flow. Figure 3 shows the calculation results for $\beta = 0.005$: the charge decreases tenfold over a distance of 2.5 cm, which is equal to the electrode gap length. The extent of the acceleration zone amounts to $\approx 0.2l_0$ and is virtually independent of the applied voltage. There also exists a wide zone of uniform motion, where the velocity vectors are aligned with the axis passing through the center of the active electrode normally to the plane electrode and the magnitude of the velocity varies only slightly. This zone occupies most of the electrode gap. The zone of effective stagnation starts to form at $0.6l_0$. All these features are, in general, consistent with earlier publication [4] concerning the kinematic pattern of developed electrohydrodynamic flows.

It is known that so-called smooth flow about a cylinder occurs at low Reynolds numbers. In this case, streamlines in the liquid are symmetric about the cylinder axis. The streamline pattern in the electrohydrodynamic flow near the active electrode is akin to that in the smooth flow: the streamlines describe circles centered on the electrode axis. However, the symmetry about the plane containing the axis of the active electrode and being parallel to the plane electrode is broken. This is because volume electric forces accelerate the liquid behind the cylindrical electrode toward the plane electrode. The jet of the liquid contracts, and the velocity isolines clearly demonstrate a zero-velocity domain extending in the flow direction. Behind the electrode, a low-pressure zone is observed. As the rate of decrease of the charge goes on dropping, the flow pattern as a whole remains almost unchanged but acquires new attributes near the central stream unobserved in the
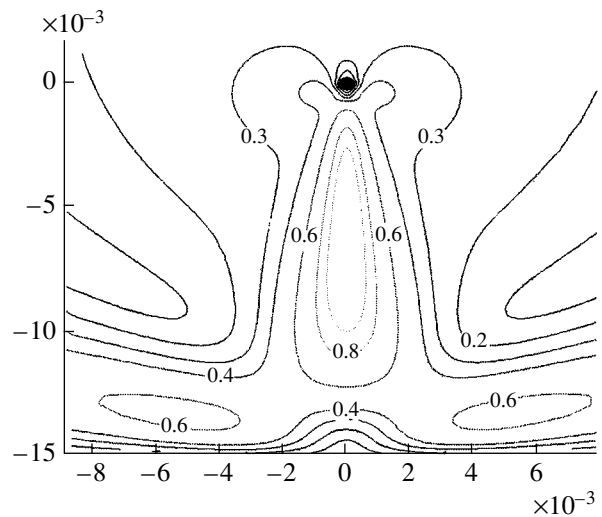
Fig. 3. Velocity isolines for the fully developed electrohydrodynamic flow ($\beta = 0.005$).

experiments. Specifically, the velocity of the central stream is no longer constant: it slowly grows, attaining a maximum closer to the plane electrode. At $\beta > 0.01$, the charged stream turns into a narrow strip of a nearly constant width and the charge density along the stream varies by no more than 30% over the electrode gap.

In the developed flow (unlike the undeveloped one), the region of localization of acting forces occupies an appreciable part of the electrode gap. Therefore, the flow pattern acquires new attributes not found in the immersed jet, namely, an extended acceleration zone and an intermediate zone of weakly varying flow. These zones are clearly seen in the acceleration distributions shown in Fig 4. Also, the calculation shows that the space charge appreciably influences the electric field distribution in the liquid under these conditions.

Figure 4 shows the acceleration distributions along a line passing through the center of the cylindrical electrode normally to the plane one for rates of decrease of the charge $\beta = 0.001, 0.003, 0.005, \ldots, 0.010$. It is seen that $\beta = 0.001$ corresponds to the undeveloped flow: the acceleration zone occupying $\approx 0.1l_0$ is followed by the extended stagnation zone, where the acceleration is negative. At β ranging between 0.005 and 0.010, the pattern of the flow is virtually invariable and consists of three zones. In the acceleration zone, a maximum of velocity is reached near $0.05l_0$. In the uniform flow zone, the velocity is directed along the flow axis and changes weakly (the accelerations are low). The location of the effective stagnation zone depends on the rate of decrease of the charge and begins at 0.4–$0.6l_0$ with a maximum of the velocity at $0.8l_0$. As β grows, the uniform flow zone expands.

As follows from the calculation, the space charge distribution typical of a fully developed electrohydrodynamic flow may appreciably distort the electric field
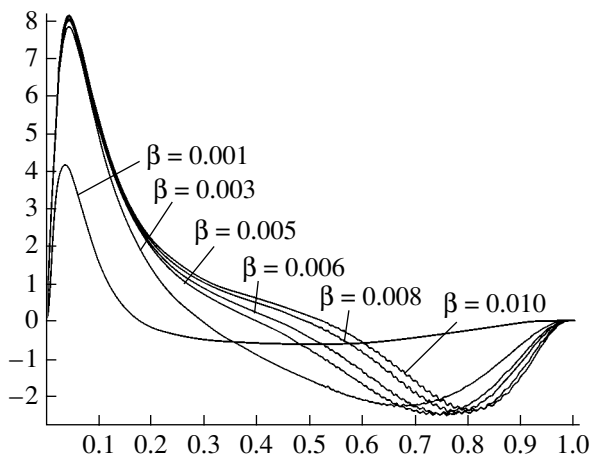
Fig. 4. Acceleration distributions along the line passing through the center of the cylindrical electrode normally to the plane electrode at different β.

distributions calculated without considering the space charge. The distributions of Coulomb forces causing electrohydrodynamic flows are of a complicated character: the forces differ both in magnitude and in direction. Unlike the distribution of accelerations, that of Coulomb forces does not show a maximum. To gain a deeper insight into the situation, we carried out a numerical experiment, setting a uniform distribution of forces acting along the central stream axis and localized within a strip with sizes roughly equal to those of the zone of space charge localization in a fully developed electrohydrodynamic flow. The pattern of the resulting flow is similar to that arising when the charge gently diffuses along the central stream. Namely, the extended acceleration zone and the intermediate uniform-flow zone are virtually absent, and the acceleration and stagnation zones alternate. The velocity reaches a maxi-

mum within the strip of forces, and the stagnation zone immediately follows the maximum.

Thus, having performed numerical analysis, we discovered a correlation between the zone structure of the electrohydrodynamic flow and the space charge distribution in the central stream of the flow.

Now that we are fully aware of the kinematic and electric characteristics of the electrohydrodynamic flow (the distributions of electric field strength E, charge density ρ, and the flow velocity), it is possible to find a distribution of local electric Reynolds numbers, $\mathrm{Re_{el}} = V/bE$, and the density of convection current, $J_c = \rho V$, in the volume of the liquid. The values of the current density (Fig. 5b) are normalized by the maximum value ($j_{\max} = 0.7 \times 10^{-4}\ \mathrm{A/m^2}$). The electric Reynolds number distribution (Fig. 5a) has a drop-shaped form with a maximum in the central stream. The convection current has the form of a narrow strip in the central stream of the flow. The regions of low $\mathrm{Re_{el}}$ and J_c near both electrodes are seen. In these regions, ion transport is accomplished chiefly by migration.

In many works (see, e.g., [6, 7]), the nonlinear part of the I–V characteristic is associated with the mechanism of conduction by convection. We share this viewpoint. The density of the total current is equal to the sum of the conduction and convection current densities, $j \cong \rho(V + bE)$. In the central stream, the density of the convection current considerably exceeds the density of the conduction current (Fig. 5b). However, in the stagnant near-electrode layers, the electric Re is small, so that the role of the conduction current in these regions becomes appreciable. The total resistance of the electrode–liquid system is the sum of the resistances of all the zones plus the resistances of the electrode–liquid interfaces. The highest resistance region is the major contributor to the total resistance, and it needs not be the region around the central stream. Studying the influ-
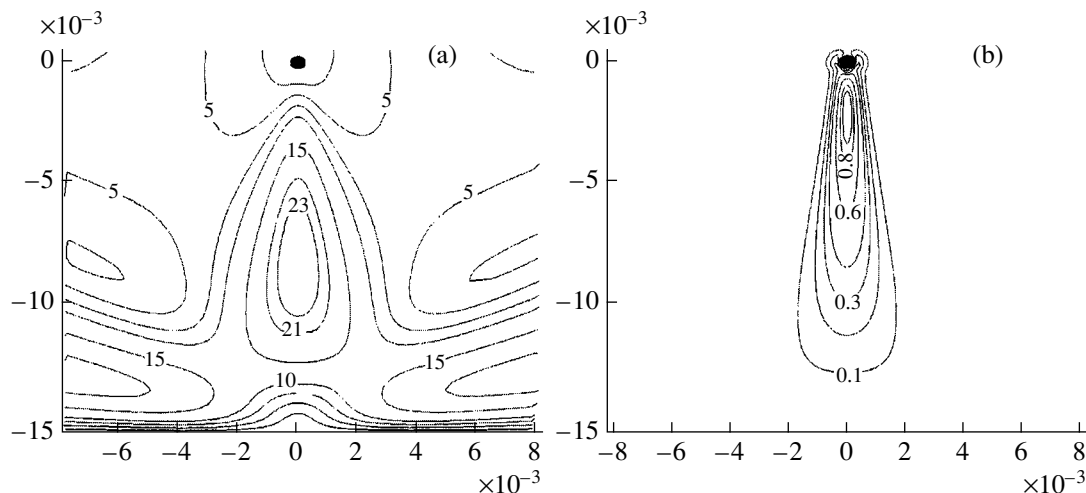

Fig. 5. (a) Electric Re isolines and (b) convection current density isolines in the fully developed electrohydrodynamic flow.

ence of the electrode–liquid interface on the form of the *I–V* characteristic, one of the authors [8] showed that the metal–liquid contact might be a controlling factor. Therefore, while diving the convection term the attention it deserves, we yet believe that the influence of all the zones on the *I–V* characteristic should be thoroughly analyzed in each particular case.

CONCLUSIONS

We carried out a computer-aided simulation of the kinematic pattern of the electrohydrodynamic flow at different space charge distributions in the liquid. The conclusions drawn are as follows.

(1) Symmetric space charge distributions in the form of a ring surrounding the electrode without considering charge drift into the liquid generate local multiple-vortex flows, which do not penetrate deep into the liquid. Such flows have not been observed experimentally.

(2) Weakly symmetric charge distributions localized near the cylindrical electrode cause two-vortex flows toward the plane electrode. The pattern of these flows is similar to that of the observable undeveloped electrohydrodynamic flow described earlier. In this flow, the influence of the space charge is insignificant and the flow pattern resembles the pattern of an immersed jet.

(3) If the charge drift from the near-electrode zone into the liquid is pronounced, the penetration depth of the flow increases and the fully developed electrohydrodynamic flow sets in. In this flow, the space charge has an appreciable effect on the electric field distribution in the liquid: the electric force density distribution and the convection current density distribution have the form of a fine filament along the flow axis. Under these conditions, the zone structure of the flow is similar to the developed electrohydrodynamic flow pattern obtained earlier from analysis of experimental data.

(4) In the immediate vicinity of the cylindrical electrode, there is a zone of very low velocities and reduced pressure, although the density of acting electric forces here is high. In this zone, the electrohydrodynamic flow is akin to the smooth flow and does not separate. For such a flow about the electrode, the charge in the form of a fine filament is transported from the near-electrode zone into the electrode gap.

(5) In the fully developed electrohydrodynamic flow, the charge transport in the near-electrode zones and in the central stream is accomplished largely by migration and convection, respectively, as follows from the simulated distributions of the electric Re and convection current. When analyzing the mechanism governing the form of the *I–V* characteristic, one should take into account all stages of the electrode processes, including those at contacts.

REFERENCES

1. Yu. K. Stishkov and A. A. Ostapenko, *Électrohydrodynamic Flows in Liquid Insulators* (Leningr. Gos. Univ, Leningrad, 1989) [in Russian].

2. Yu. K. Stishkov and M. A. Pavleyno, Elektrokhim. Obrab. Met., No. 1, 14 (2000).

3. A. V. Buyanov, M. A. Pavleyno, and Yu. K. Stishkov, in *Proceedings of the 6th International Scientific Conference on Contemporary Problems in Electrophysics and Electrohydrodynamics of Liquids, St. Petersburg, 2000,* pp. 4–9.

4. A. V. Buyanov, M. A. Pavleyno, and Yu. K. Stishkov, Vestn. St. Peterb. Univ., Ser. 4: Fiz., Khim., No. 12 (2001).

5. A. V. Buyanov and Yu. K. Stishkov, Zh. Tekh. Fiz. **73** (8), 34 (2003) [Tech. Phys. **48**, 972 (2003)].

6. P. Atten and I. Haidara, IEEE Trans. Electr. Insul. **20**, 187 (1985).

7. P. Atten and I. Haidara, IEEE Trans. Ind. Appl. **21**, 709 (1985).

8. Yu. K. Stishkov, Dokl. Akad. Nauk SSSR **288**, 861 (1986) [Sov. Phys. Dokl. **31**, 500 (1986)].

Translated by N. Mende

GASES
AND LIQUIDS

Effect of External Power Supply to a Shock Layer on the Shock Wave Structure

V. Yu. Golyatin*, V. V. Kuchinskiy**, and V. S. Sukhomlinov*

** Physics Research Institute, St. Petersburg State University,
Ul'yanovskaya ul. 1, Petrodvorets, St. Petersburg, 198903 Russia*

e-mail: SUKHOMLINOV@POBOX.ru

*** Research Institute of Hypersonic Systems, St. Petersburg, 196066 Russia
e-mail: AJAX@COMSET.NET*

Received June 28, 2004

Abstract—A physical model that describes the structure of a 1D shock wave in a gas containing a moving heat source is put forward. A stationary equation for the profile of a shock wave in a gas with an arbitrary-shape heat source that is at rest relative to this wave is derived. Analytical solutions to this equation make it possible to analyze the flow pattern in the case of external power supply. © 2005 Pleiades Publishing, Inc.

INTRODUCTION

Much interest has recently arisen in the effect of external actions on supersonic flows. In particular, the propagation of shock waves (SWs) in a weakly ionized gas-discharge plasma was studied in [1, 2]. Anomalous changes in the SW structure (SW front broadening, SW attenuation, and the nonmonotonicity of the SW structure) were observed in experiments with traveling SWs and with bodies flying through a plasma [3–5]. It was supposed [6] that, when SWs propagate along an electric field, some of these effects can be accounted for by the appearance of an electric field behind the shock wave, where the density of the gas rises. Such conditions are equivalent to the presence of an external power supply.

Basic formulas that extend the Hugoniot–Rankine relationships to the case when a heat source that is at rest relative to a shock wave is present near the SW front were derived in [7]. The results obtained in [7] relate the asymptotic values of gas flow parameters in the undisturbed (supersonic) flow with the asymptotic values of the parameters in the disturbed flow and cannot describe the shock layer structure.

In this work, we suggest simple analytical techniques for calculating the SW shape at any point of the space in the 1D approximation. Analytical techniques make it possible to clearly illustrate how local energy supply influences the SW structure. In addition, such methods will allow researchers to reveal general trends in a phenomenon under consideration. Specifically, we consider here the supersonic motion of aerodynamic objects in a plasma, that is, ballistic-range experiments

[3]. Experimental approaches to determining the SW parameters can be found elsewhere [1–5].

1. BASIC EQUATIONS

Consider a plane SW moving in a gas flow that contains a heat source near the SW front. The set of 1D nonstationary equations describing the SW parameters in the presence of a heat source moving in the x direction with velocity w has the form

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0,$$

$$\frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}\left(\rho u^2 + p - \frac{4}{3}\mu\frac{\partial u}{\partial x}\right) = 0,$$

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\rho u^2 + \rho e\right) + \frac{\partial}{\partial x}\left\{\left(\frac{1}{2}u^2 + e\right)\rho u\right.$$

$$\left. + \left(p - \frac{4}{3}\mu\frac{\partial u}{\partial x}\right)u - \lambda\frac{\partial T}{\partial x}\right\} - Q(x - wt) = 0.$$

$$(1)$$

Here, ρ, u, e, μ, λ, and T are, respectively, the density, mass velocity, internal energy, viscosity, thermal conductivity, and absolute temperature of the gas and p is the static pressure.

A stationary solution to set (1) will be sought in the coordinate system moving with velocity w. To this end, we pass to variable $\xi = x - wt$. As a desired function, we take $v = w - u$, which corresponds to the coordinate system moving with the heat source and also makes it possible to consider the case of a stationary (unmoving) heat source in a gas flowing along the ξ axis. Using

well-known relationships, one can easily derive from set (1)

$$\frac{4}{3}\mu\frac{dv}{d\xi} = P_0 - Q_0\left(v + \frac{\gamma-1}{\gamma}\frac{H}{v}\right),$$

$$\frac{4}{3}\mu\frac{d}{d\xi}\left(H + \frac{1}{2}v^2\right) = E_0 - Q_0\left(H + \frac{1}{2}v^2\right) + \int_\xi^\infty Q(\xi')d\xi',$$

(2)

where H is the gas enthalpy per unit mass; Q_0, P_0, and E_0 are positive constants; $\gamma = c_p/c_v$; and c_p and c_v are specific heats at constant pressure and volume, respectively.

Integration of the second equation in (2) subject to

$$\varphi(\xi) \xrightarrow[\xi\to\infty]{} \frac{E_0}{Q_0}; \quad \varphi(\xi) \xrightarrow[\xi\to-\infty]{} \frac{E_0}{Q_0} + \frac{1}{Q_0}\int_{-\infty}^\infty Q(\xi')d\xi'$$

at infinity gives for function $\varphi(\xi) = H + (1/2)v^2$ the expression

$$\varphi(\xi) = \frac{E_0}{Q_0} + \frac{1}{Q_0}\left\{\int_\xi^\infty Q(\xi')d\xi'\right.$$

$$\left. + \exp\left(-\frac{3Q_0}{4\mu}\xi\right)\int_{-\infty}^\xi \exp\left(\frac{3Q_0\xi'}{4\mu}\right)Q(\xi')d\xi'\right\}.$$

(3)

Let $\lim_{\xi\to\infty} v = v_1$, where v_1 is the mass velocity of the gas relative to the source moving in the undisturbed flow. Suppose that there exists a point $\xi = \xi_0$ where

$$\left.\frac{dv}{d\xi}\right|_{\xi=\xi_0} = 0,$$

$v(\xi)$ being a continuous function at $\xi > \xi_0$, and let $v|_{\xi=\xi_0} = v_2$. For enthalpy $H(\xi)$, it follows from (3) that

$$H(\xi) = \frac{E_0 + \Delta E(\xi)}{Q_0} - \frac{1}{2}v^2(\xi).$$

Putting $\xi \longrightarrow \infty$ and $\xi = \xi_0$, we obtain two relationships

$$\frac{P_0}{Q_0} = \frac{\gamma+1}{2\gamma}(v_2 + v_1) + \frac{\gamma-1}{\gamma}\frac{\Delta E_0}{v_2 - v_1}\frac{1}{Q_0},$$

$$\frac{P_0}{Q_0} = \left\{\frac{\gamma+1}{2\gamma}v_1 + \frac{\gamma-1}{\gamma}\frac{E_0}{v_1}\frac{1}{Q_0}\right\},$$

(4)

where

$$\Delta E(\xi) = \int_\xi^\infty Q(\xi')d\xi'$$

$$+ \exp\left(-\frac{3Q_0\xi}{4\mu}\right)\int_{-\infty}^\xi \exp\left(\frac{3Q_0\xi'}{4\mu}\right)Q(\xi')d\xi'$$

is the delivered power flux per unit area with $\Delta E_0 = \Delta E(\xi_0)$.

Substituting equalities (4) into the first equation in (2) yields

$$\frac{4}{3}\mu\frac{dv}{d\xi} = \frac{\gamma+1}{2\gamma}\frac{Q_0}{v}\left\{(v_1 - v)(v - v_2)\right.$$

$$\left. + \frac{2(\gamma-1)}{(\gamma+1)}\frac{\Delta E_0(v - v_1)}{(v_2 - v_1)Q_0} - \frac{2(\gamma-1)}{(\gamma+1)}\frac{\Delta E(\xi)}{Q_0}\right\}.$$

(5)

In the dimensionless coordinate system

$$y = \frac{3}{8}\frac{\rho_1 a_1}{\mu}\frac{(\gamma+1)}{\gamma}(x - wt),$$

(6)

Eq. (5) turns into a stationary (in the coordinate system selected) inhomogeneous equation for velocity $v = w - u$,

$$v\frac{dv}{dy} = (v_1 - v)(v - v_1\omega^2)M - v_1^2\frac{\beta(y)}{\gamma^2 M^2},$$

(7)

where

$$\beta(y) = \frac{16(\gamma-1)}{3(\gamma+1)^2}\frac{\gamma^3\mu}{\rho_1^2 a_1^4}\left\{\int_y^\infty Q(y')dy'\right.$$

$$\left. + \exp\left(-\frac{3\rho_1 a_1 y}{4\mu}\right)\int_{-\infty}^y \exp\left(\frac{3\rho_1 a_1 y}{4\mu}\right)Q(y')dy'\right\} > 0,$$

$$v_2 = v_1\frac{\gamma M^2 + 1 - (M^2 - 1)\sqrt{1 - \dfrac{\beta(y_0)}{\beta_{\lim}}}}{(\gamma+1)M^2},$$

(8)

$$\beta_{\lim} = \frac{\gamma^2(M^2 - 1)^2}{(\gamma+1)^2 M}, \quad \beta(y_0) = \lim_{y\to y_0}(\beta(y)),$$

$$\omega = \sqrt{\frac{2 + (\gamma-1)M^2}{(\gamma+1)M^2}}.$$

Here, v_1 and ρ_1 are, respectively, the velocity and density of the gas, $M \equiv M_1 = v_1/a_1$ is the Mach number, and a_1 is the sound velocity in the undisturbed flow (at $y \longrightarrow +\infty$).

It is evident that

$$v \xrightarrow[y \to \infty]{} v_1, \quad \frac{dv}{dy}\bigg|_{y = y_0} = 0,$$

$$v\big|_{y = y_0} = v_2, \quad y_0 = \frac{3\rho_1 a_1(\gamma + 1)}{8M}\frac{\xi_0}{\gamma}.$$

From the expression for v_2, it follows that, if the inequality

$$\beta(y_0) \le \beta_{\lim} = \frac{\gamma^2(M^2 - 1)^2}{(\gamma + 1)^2 M} \tag{9}$$

is met, there exists a point $y = y_0$ where $v\big|_{y = y_0} = v_2$ and

$$\frac{dv}{dy}\bigg|_{y = y_0} = 0.$$

Detailed analysis of Eqs. (7)–(9) shows that the above conditions are met iff $y_0 \longrightarrow -\infty$.

If the heat source does not generate power, Eq. (7) turns into the equation derived in [8]. In this case ($\beta = 0$), a solution to Eq. (7) can be found in implicit form,

$$1 - \frac{v}{v_1} = \left[\frac{(v - v_2)(v_1 - \sqrt{v_1 v_2})}{v_1(\sqrt{v_1 v_2} - v_2)}\right]^{\frac{v_2}{v_1}}$$

$$\times \exp\left(-\frac{v_1 - v_2}{a_1}y\right). \tag{10}$$

The constant of integration here is taken such that point $y = 0$ corresponds to the maximum of the first derivative of the velocity $(d^2 v(y)/dx^2|_{y = 0} = 0)$. Velocity $v = v(y)$ and its first derivative at point $y = 0$ are given by the simple expressions

$$v(0) = v_1\omega, \quad \frac{dv}{dy}\bigg|_{y = 0} = Mv_1(1 + \omega)^2 \tag{11}$$

$$(\beta(y) = 0).$$

When solving Eq. (7) for $\beta(y) \ne 0$ in order to find the absolute value of the SW front displacement due to heating, one must keep the origin at the point where the first derivative of the velocity with respect to distance is maximal (i.e., at $y = 0$). A constant of integration should be taken such that a solution obtained will continuously turn into solution (10) when $\beta(y) \ne 0$ continuously passes to $\beta(y) = 0$. To this end, it will suffice to find the first derivative of the velocity from the condition of vanishing of the second derivative and substitute $\frac{dv_1}{dy}\bigg|_{y = 0}$ into basic equation (7). Such a procedure yields a fourth-degree polynomial in correction Δ to velocity $v(0)$ at $\beta(y) = 0$ (see formula (7)),

$$v(0) = v_1\omega + v_1\Delta, \quad \Delta^4 + a_3\Delta^3 + a_2\Delta^2 + a_1\Delta + a_0 = 0,$$

$$a_0 = c_0\omega(1 - \omega)^2 - c_0^2 - c_1\omega^2,$$

$$a_1 = c_0(1 + \omega^2) - 2\omega^2(1 - \omega)^2 - 2c_1\omega, \tag{12}$$

$$a_2 = -c_1 - 3\omega(1 - \omega)^2, \quad a_3 = -\omega^2 + 4\omega - 1,$$

$$c_0 = \frac{1}{\gamma^2 M^3}\beta(0), \quad c_1 = \frac{1}{\gamma^2 M^4}\frac{d\beta(y)}{dy}\bigg|_{y = 0}.$$

The second (in magnitude) root of polynomial (12) gives Δ needed to calculate $v(0) = v_1(\omega + \Delta)$ (at $\beta \longrightarrow 0$, this root $\Delta \longrightarrow 0$, since $a_0 \longrightarrow 0$). In most cases, $v(0)$ can be fairly accurately found from the approximate formula

$$v(0) \approx v_1\left(\omega - \frac{a_1 + \sqrt{a_1^2 - 4a_0 a_2}}{2a_2}\right). \tag{13}$$

2. COMPUTATIONAL FORMULAS FOR SHOCK WAVE PARAMETERS

Knowing a solution at $y = 0$, one can uniquely determine the velocity profile. This may be done with routine programs (using, e.g., any version of the Runge–Kutta method) or with approximants given below (these approximants make it possible to determine the velocity profile with a fairly high accuracy). Then, using the well-known relationships, one can find the pressure, temperature, and density of the gas, as well as the local Mach number,

$$p(y) = p_1\left\{1 + \gamma M^2\left[1 - \frac{v(y)}{v_1}\right]^2\right.$$

$$+ \frac{v_1 M^2(\gamma + 1)}{2v(y)}\left[\left(1 - \frac{v(y)}{v_1}\right)\left(\omega^2 - \frac{v(y)}{v_1}\right) + \frac{\beta(y)}{\gamma^2 M^3}\right]\right\}, \tag{14}$$

$$T(y) = 3.4837 \times 10^{-3}\frac{p(y)}{v_1\rho_1}v(y),$$

$$\rho(y) = \frac{\rho_1 v_1}{v(y)}, \quad \chi(y) = \frac{p(y)}{p_1} - 1,$$

$$M(y) = \frac{v(y)}{v_1}\left[\frac{\gamma - 1}{2}\left(1 - \left(\frac{v(y)}{v_1}\right)^2\right)\right.$$

$$\left. + \frac{1}{M^2} + \frac{(\gamma + 1)}{2\gamma^2 M^3}\beta(y)\right]^{-1/2}. \tag{15}$$

From (15), one can find the velocity at which the local Mach number (designated as $M(y)$, whereas M is the Mach number in the undisturbed flow) equals unity at point y_s,

$$v(y_s) = v_1\sqrt{\omega^2 + \frac{\beta(y_s)}{\gamma^2 M^3}}. \tag{16}$$

Point y_s, where equality (16) is valid, specifies the SW front displacement due to heating; thus, expression (16) is an equation from which the SW front displacement can be found.

As estimators, it is convenient to take analytical formulas that describe the SW parameter variation with an accuracy sufficient for our purposes. For the SW velocity in the absence of a heat source ($\beta(y) = 0$), asymptotic estimates yield two approximate relationships,

$$v(y) \approx v_u(y) = v_1 \omega \left[1 + \frac{(1 - \omega)\left(e^{My\left(\frac{1}{\omega} - \omega\right)} - 1 \right)}{\omega e^{My\left(\frac{1}{\omega} - \omega\right)} + 1} \right], \quad (17)$$

$$v(y) \approx v_d(y) = v_1 \begin{cases} \omega^2 + \dfrac{1 - \omega^2}{1 + \dfrac{1}{\omega} e^{-My(1 - \omega^2)}} & \text{at } y \geq 0 \\[4mm] 1 - \dfrac{1 - \omega^2}{1 + \omega e^{My\left(\frac{1}{\omega^2} - 1\right)}} & \text{at } y \leq 0. \end{cases} \quad (18)$$

Comparing the results obtained with (17) and (18) shows that formula (17) provides an upper estimate of the exact solution (implicitly, the exact solution is given by (10)), while formula (18), a lower estimate At $\beta(y) \neq 0$, the SW velocity is approximately given by formula (26) in the Appendix. Also, formula (28) in the Appendix estimates SW front width Δ_Q, and approximate formula (29), SW front displacement Δ_s.

3. WEAK-WAVE APPROXIMATION

In the weak-wave approximation ($M - 1 \ll 1$), it is convenient to take advantage of function

$$\psi(z) = \frac{1}{\alpha}\left(1 - \frac{v(z)}{v_1}\right), \quad (19)$$

where

$$\alpha = 1 - \omega^2 = \frac{2(M^2 - 1)}{(\gamma + 1)M}, \quad z = \alpha M y. \quad (20)$$

In this approximation, function $\psi(z)$ satisfies the equation

$$\frac{d\psi}{dz} = \psi^2 - \psi + \frac{\bar{\beta}}{4}f, \quad (21)$$

where

$$\bar{\beta} = \frac{\beta(-\infty)}{\beta_{lim}}, \quad \beta(z) = \beta(-\infty)f(z). \quad (22)$$

As in Section 2, the origin should be placed at point z where $d^2\psi/dz^2|_{z=0} = 0$. Note that

$$\psi_0 = \psi(0) = \frac{1}{2}(1 + \delta),$$

where δ is the second (in magnitude) root of the cubic equation

$$\delta^3 + (\bar{\beta}f(0) - 1)\delta + \bar{\beta}\frac{df}{dz}\bigg|_{z=0} = 0. \quad (23)$$

Equation (15) can be solved exactly, e.g., if the integral heat supply is described by the exponential

$$f(z) = \frac{1}{2} \begin{cases} e^{-\frac{z}{L}} & \text{at } z \geq 0 \\[3mm] 2 - e^{\frac{z}{L}} & \text{at } z < 0 \end{cases} \quad (24)$$

or is approximated by a linear function in the interval $-L \leq z \leq L$,

$$f(z) = \begin{cases} 1 & \text{at } z < -L, \\[2mm] \dfrac{1}{2}\left(1 - \dfrac{z}{L}\right) & \text{at } -L \leq z < L \ (0 \leq f(z) \leq 1) \\[2mm] 0 & \text{at } z \geq L. \end{cases} \quad (25)$$

For the SW velocity subject to exponential integral power supply (24), we obtain formula (30); for linear power supply (25), we get formula (32) (formulas (30) and (32) are given in the Appendix).

Figure 1 compares the numerical solution of exact equation (7) (continuous line) with the analytical solution obtained by the formulas in the weak-wave approximation (dots). The weak-wave approximation works worst near the minimum of the SW velocity distribution along the z coordinate. For $M \leq 1.4$ and $\beta < \beta_{lim}$, this approximation yields almost exact values. Different variants of the formulas for the SW velocity make it possible to estimate the behavior of other SW parameters (see formulas (14) and (15)).

Thus, we have derived estimators for SW shape variation in the presence of a heat source. They relate the amount of supplied heat with temperature, density, and pressure variation.

4. RESULTS AND DISCUSSION

The formulas derived and basic equation (7) lead us to the following conclusions: (i) in the case of external power supply, the SW profile may be nonmonotonic; (ii) the local Mach number may either be a monotonic function of coordinate or have a single minimum downstream of the SW center; and (iii) at any Mach number, the SW intensity drops and the SW front broadens as the integral amount of the heat supplied increases (Figs. 2 and 3).
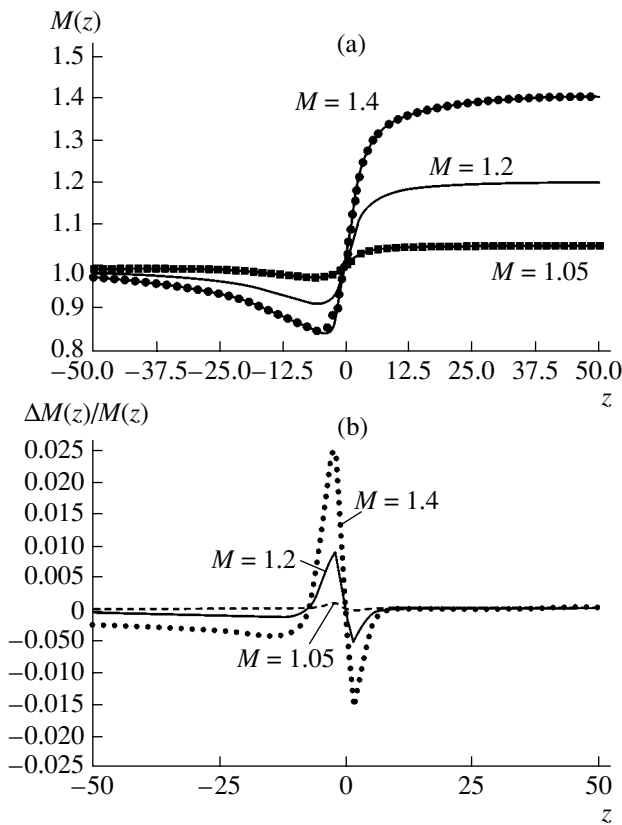
Fig. 1. (a) Coordinate dependence of the local Mach number. Continuous curves, the exact solution to Eq. (3); dotted curves, the weak-wave approximation. (b) Relative error associated with the weak-wave approximation at different initial Mach numbers (figures by the curves).

Fig. 2. (a) Velocity and (b) Mach number distributions at different positions of the local heat source. The source has a Gaussian shape with a half-width of 2.5 mm and generates the maximal heat (β_{lim}). (*1*) Distributions in the absence of heat, (*2*) heat source is shifted downstream by 1 mm, (*3*) heat source is at the shock front (unshifted), and (*4*) heat source is shifted upstream by 1 mm.

When analyzing the variation of the flow structure subject to external heat supply, one has to determine the boundaries of the super- and subsonic flows. In the absence of external heat ($\beta = 0$), it can be shown, using formulas (15), (17), and (18), that $M_0(y) > 1$ for $y > 0$ and $M_0(y) < 1$ for $y < 0$. Thus, in our coordinate system, point $y = y_s$ separating the regions of super- and subsonic flows in the SW coincides with the point of maximum of derivative $dv_1(y)/dy$; so, $y_s = 0$. If heat is delivered to the shock layer, the boundary between the super- and subsonic flows in the SW shifts upstream. However, in the subsonic region at $y \longrightarrow -\infty$, the Mach number increases compared with the case when external heat is not supplied.

From the formulas derived above, it follows that, when the amount of heat equals a critical value (i.e., at $\beta(y_0) \equiv \beta(-\infty) = \beta_{lim}$; formula (9)), the flow behind the wave asymptotically tends to become sonic (Fig. 2). Hence, an SW, which is treated as a continuous deceleration through the sonic speed (the supersonic-to-subsonic transition), disappears in this case. Let us consider reasons for such behavior of the flow subject to external heat supply and also why the Navier–Stokes equations that are stationary in the coordinate system

related to a moving source cannot be solved at $\beta(-\infty) > \beta_{lim}$.

We found that, at any $\beta(-\infty) < \beta_{lim}$, the flow splits into the supersonic ($y > y_s$) and subsonic ($y < y_s$) regions. As the shock layer is approached, the flow deceleration in the absence of external heat is due to gas compression. It is known that, when heat is delivered to the supersonic region, the higher the amount of heat, the more effective the flow deceleration, and the gas temperature rises [9]. This additional deceleration reduces the mass velocity in front of the shock layer; that is, $v(y)$ and the local Mach number decrease (Figs. 2 and 3). As the Mach number decreases, so does the compression in the SW and, hence, the deceleration of the flow. In going to the subsonic region ($y < y_s$), external heat supply to the flow is known [9] to raise the flow velocity, the mass velocity increasing with the rate of heat delivery. Note that the increase in the mass velocity in the subsonic region is not compensated for

Fig. 3. Distributions of the (a) velocity, (b) Mach number, (c) intensity, (d) density, (e) pressure, ad (f) temperature for different amounts of supplied heat β. $\beta(-\infty)/\beta_{\lim}$ (*1*) 0, (*2*) 0.25, (*3*) 0.50, and (*4*) 1.00.
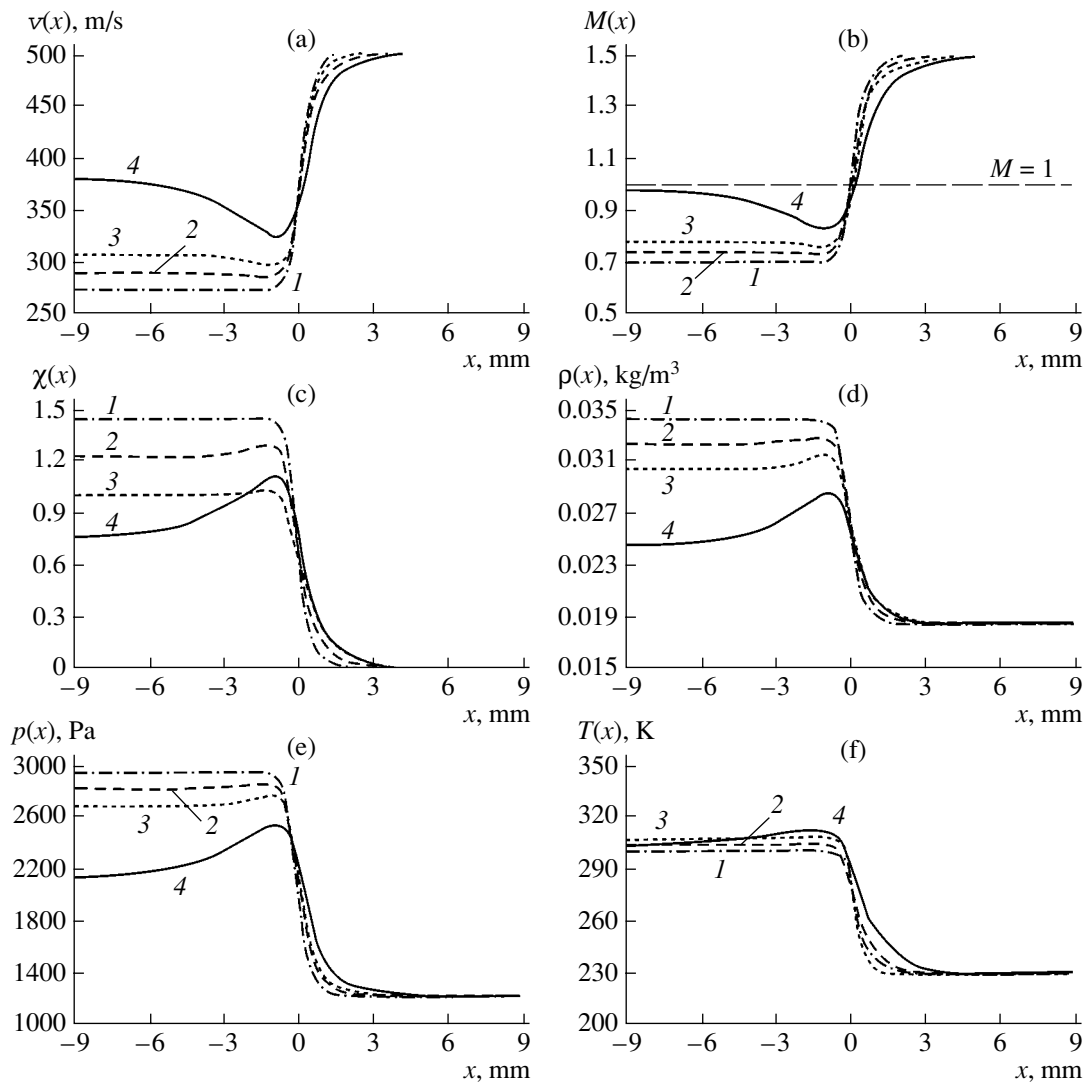
by an increase in the gas temperature. Therefore, while generally decreasing, the local Mach number becomes the higher (at $y \longrightarrow -\infty$) or the lower (at $y \longrightarrow +\infty$), the higher the rate of heat delivery (Fig. 3). As follows from Fig. 3c, when the amount of heat grows, the intensity of the SW decreases and the wave itself diffuses (the gradients of $v(y)$, $\rho(y)$, etc. smooth out). Thus, it can be argued that, at $\beta(-\infty) = \beta_{\lim}$, the shock wave somewhat transforms into a domain where the gas parameters vary smoothly.

A significant remark is in order. Let the heat source be in the subsonic region. Then, the heat will readily propagate up- and downstream via heat conduction and downstream via transfer by the flow.

Now consider the case when the heat source is in the supersonic region. The heat transfer downstream will be as efficient as in the former case. Since the heat is transferred with a thermal velocity, it seemingly cannot

propagate upstream, i.e., against the direction of the mass velocity. Actually, however, this is not the case, because gas molecules obey the Maxwell velocity distribution. For any M, there exist particles with velocities directed against the flow and exceeding the mass velocity. Obviously, the number of such particles decreases with increasing the mass velocity of the gas. Thus, irrespective of whether the flow is supersonic or subsonic, heat delivered to a limited space will propagate both downstream and (while with a lower efficiency) upstream. Let us substantiate this statement. If we represent volume power density $Q(\xi)$ of a spatially limited heat source, e.g., in the form $Q(\xi) = Q_0(\xi)\theta(\xi - \xi_1)\theta(\xi_2 - \xi)$ ($\theta(\xi)$ is the Heaviside function), function $\Delta E(\xi)$ defined by formula (4) will exponentially drop down to zero at $\xi > \xi_2$ and remain constant at $\xi < \xi_1$. Such reasoning explains why a spatially limited source supplying heat to the supersonic flow changes the flow

pattern upstream of the heat release area in the stationary solution.

Consider now the gas flow when two heat sources, one ($Q_1(y)$) in the supersonic region and the other ($Q_2(y)$) in the subsonic one, generate the same power. On the right of point $y = y_s$, Q_2 has a weaker effect than Q_1 because of the minor heat flux upstream, as we have just concluded. In other words, the gas deceleration in the supersonic region is stronger: $v_2(y) > v_1(y)$ at $y \longrightarrow \infty$. As for the local Mach number, $M_1(y) < M_2(y)$. Because of the more significant decrease in the Mach number in the former case (in the supersonic region), the SW is attenuated more significantly; hence, the compression of the flow in the shock layer exerts a smaller effect on the gas deceleration. Therefore, a point must exist (y_{21}) where $v_1(y_{21}) = v_2(y_{21})$. In going into the subsonic region, Q_1 will make a higher contribution to the flow acceleration until $y \geq y_{21}$. This is because a major fraction of heat Q_1 is transferred downstream via mass velocity and heat conduction. At the same time, a part of heat Q_2 propagates upstream; i.e., Q_2 is less efficient. At $y \leq y_{21}$, the effect of Q_1 and Q_2 on the flow parameters is nearly the same and the mass velocities are asymptotically equalized. Therefore, $v_1(y) > v_2(y)$ and $v_1(-\infty) = v_2(-\infty)$ at $y \longrightarrow -\infty$. The effect of the shape, intensity, and location of the heat sources on other flow parameters may be analyzed in a similar way. From the aforesaid, it follows, in particular, that, when heat is supplied to the supersonic region (in front of the SW), the wave broadens to a greater extent than when the same amount of heat is supplied to the subsonic region (behind the wave).

The formulas derived allow one to study at length changes in the SW structure that are induced by heat delivered to the SW front.

APPENDIX

The approximate formula for the velocity (the approximate solution to Eq. (7)) is given by

$$
v_h(y) \approx v_1 \left\{ \frac{1+\gamma M^2}{(\gamma+1)M^2} + \frac{v_2 \dfrac{d\beta(y)}{dy}}{4(\beta_{\lim}-\beta(y))Mv_1} \right.
$$

$$
\left. + \tanh\left[v_1 \frac{\displaystyle\int_0^y \sqrt{M(\beta_{\lim}-\beta(z))}\,dz}{\gamma M v_2} + \theta \right] \sqrt{\frac{\beta_{\lim}-\beta(y)}{\gamma^2 M^3}} \right\},
$$
(26)

$$
\theta = -\tanh^{-1}\left[\gamma\sqrt{M^3} \frac{\dfrac{1+\gamma M^2}{(\gamma+1)M^2} - \dfrac{v(0)}{v_1} + \dfrac{d\beta(0)}{dy}\dfrac{v_2}{4(\beta_{\lim}-\beta(0))Mv_1}}{\sqrt{\beta_{\lim}-\beta(0)}} \right].
$$
(27)

The shock front width is

$$
\Delta_Q = \frac{8\mu\gamma}{3\rho_1 a_1(\gamma+1)} \frac{v_1 - v_2}{\left(\dfrac{dv(y)}{dy}\right)_{y=0}}.
$$
(28)

The approximate formula for the SW front displacement has the form

$$
\Delta_s \approx \frac{8\mu\gamma}{3\rho_1 a_1(\gamma+1)}
$$

$$
\times \frac{1}{c_3}\left[\frac{c_4}{2} + \sqrt{\left(\frac{c_4}{2}\right)^2 + \omega^2 + c_0 - \frac{v(0)}{v_1}c_4 - \frac{v(0)}{v_1}} \right],
$$

$$
c_3 \equiv \frac{1}{v_1}\frac{dv(y)}{dy}\bigg|_{y=0}
$$
(29)

$$
= \frac{M}{v(0)}\left\{ [v_1 - v(0)]\left[\frac{v(0)}{v_1} - \omega^2\right] - c_0 v_1 \right\},
$$

$$
c_4 = \frac{1}{\gamma^2 M^3 c_3}\frac{d\beta(y)}{dy}\bigg|_{y=0},
$$

where $\left(\dfrac{dv(y)}{dy}\right)_{y=0}$ is calculated by formula (7).

The exact solution to Eq. (21) (weak-wave approximation) for the exponential integral heat supply (see (24)) is given by

$$
v(z) = v_1\left(1 - \alpha\begin{cases} \Psi_-(z) & \text{at } z < 0 \\ \Psi_+(z) & \text{at } z \geq 0 \end{cases}\right),
$$
(30)

where

$$\Psi_-(z) = \frac{1}{2}\left\{1 - \left(\sqrt{1-\bar{\beta}} + \frac{\zeta}{L}\left[\frac{K_{\nu+1}(\zeta) - C_-I_{\nu+1}(\zeta)}{K_\nu(\zeta) + C_-I_\nu(\zeta)}\right]\right)\right\},$$

$$\zeta = L\sqrt{\frac{\bar{\beta}e^{\frac{z}{L}}}{2}}, \quad \nu = L\sqrt{1-\bar{\beta}},$$

$$C_- = \frac{\zeta_0 K_{\nu+1}(\zeta_0) + L(1 - \sqrt{1-\bar{\beta}} - 2\psi_0)K_\nu(\zeta_0)}{\zeta_0 I_{\nu+1}(\zeta_0) + L(-1 + \sqrt{1-\bar{\beta}} + 2\psi_0)I_\nu(\zeta_0)}, \quad (31)$$

$$\zeta_0 = L\sqrt{\frac{\bar{\beta}}{2}},$$

$$\Psi_+(z) = \sqrt{\frac{\bar{\beta}e^{-\frac{z}{L}}}{8}}\frac{J_{L-1}(\xi) + C_+Y_{L-1}(\xi)}{J_L(\xi) + C_+Y_L(\xi)}, \quad \xi = -L\sqrt{\frac{\bar{\beta}e^{-\frac{z}{L}}}{2}},$$

$$C_+ = -\frac{\psi_0 J_L(\xi_0) + 0.35355\sqrt{\bar{\beta}}J_{L-1}(\xi_0)}{\psi_0 Y_L(\xi_0) + 0.35355\sqrt{\bar{\beta}}Y_{L-1}(\xi_0)},$$

$$\xi_0 = -L\sqrt{\frac{\bar{\beta}}{2}}.$$

For the case of the linear heat supply (see (25)), the SW velocity calculated through the exact solution for $\psi(z)$ is expressed through Airy functions Ai(ζ) and Bi(ζ),

$$v(z)$$

$$= v_1\left[1 - \alpha\left\{\begin{array}{l}\frac{1}{2}\left[1 - \sqrt{1-\bar{\beta}}\tanh\left[\frac{\sqrt{1-\bar{\beta}}}{2}\left(z + L - \frac{2}{\sqrt{1-\bar{\beta}}}\right)\right.\right. \\ \left.\left.\times\tanh^{-1}\left(\frac{2\psi_{-L} - 1}{\sqrt{1-\bar{\beta}}}\right)\right)\right]\right] \text{ at } z < -L, \quad (32) \\ \Psi(z) \text{ at } -L \le z \le 1, \\ \frac{1}{1 + \left(\frac{1}{\psi_L} - 1\right)e^{z-L}} \text{ at } z > L, \end{array}\right.\right.$$

where

$$\Psi(z) = \frac{1}{2}\left[1 - \sqrt[3]{\frac{\bar{\beta}}{L}}\frac{Ai'(\zeta) + C_aBi'(\zeta)}{Ai(\zeta) + C_aBi(\zeta)}\right],$$

$$\zeta = \sqrt[3]{\frac{\bar{\beta}}{L}}\left(\frac{L}{\bar{\beta}} + \frac{z-L}{2}\right),$$

$$C_a = -\frac{Ai'(\zeta_0)\sqrt[3]{\frac{\bar{\beta}}{L}} - Ai(\zeta_0)(2\psi_0 - 1)}{Bi'(\zeta_0)\sqrt[3]{\frac{\bar{\beta}}{L}} - Bi(\zeta_0)(2\psi_0 - 1)}, \quad (33)$$

$$\zeta_0 = \sqrt[3]{\frac{\bar{\beta}}{L}}\left(\frac{L}{\bar{\beta}} - \frac{L}{2}\right),$$

$$\psi_{-L} = \Psi(-L), \quad \psi_L = \Psi(L).$$

If the integral heat supply is describable by formula (22) for $f(z) = \left(1 + e^{\frac{2z}{L}}\right)^{-1}$, velocity $v(z)$ is found as the half-sum of the velocities calculated by (30) and (32) with a reasonable accuracy.

REFERENCES

1. D. M. van Wie, A. L. Wesner, and L. R. Gauthier, in *Proceedings of the 3rd Workshop on Weakly Ionized Gasses AIAA-99-4824, Norfolk, 1999*.

2. B. N. Ganguly and P. Bletzinger, in *Proceedings of the Workshop on Weakly Ionized Gases, Colorado, 1997*, Vol. 2, pp. HH1–HH13.

3. A. P. Bedin and G. I. Mishin, Pis'ma Zh. Tekh. Fiz. **21** (1), 5 (1995) [Tech. Phys. Lett. **21**, 5 (1995)].

4. A. I. Klimov and G. I. Mishin, Pis'ma Zh. Tekh. Fiz. **16** (12), 89 (1990) [Sov. Tech. Phys. Lett. **16**, 960 (1990)].

5. B. N. Ganguly, P. Bletzinger, and A. Garscadden, Phys. Lett. A **230**, 218 (1997).

6. V. V. Kuchinskiy, V. S. Sukhomlinov, V. A. Sheverev, and M. V. Oguten, in *Proceedings of the Conference "Prospects of MHD and Plasma Technologies in Aerospace Applications," Moscow 1999* (IVT RAN, Moscow, 2000), pp. 307–312.

7. V. Yu. Golyatin, V. V. Kuchinskiy, and V. S. Sukhomlinov, Zh. Tekh. Fiz. **75** (4), 25 (2005) [Tech. Phys. **50**, 408 (2005)].

8. G. B. Whitham, J. Fluid Mech., No. 4, 337 (1958).

9. L. A. Vulis, *Thermodynamics of Gas Flows* (Gosénergoizdat, Moscow, 1950) [in Russian].

Translated by V. Isaakyan

<div style="text-align:center">

GASES
AND LIQUIDS

</div>

On the Solution of the Kinetic Equation for a Heat Flux between Concentric Spheres

P. S. Aleshin and S. A. Savkov

Orel State University, Orel, 302015 Russia
Received September 15, 2003; in final form, July 26, 2004

Abstract—The problem of calculating a heat flux in a spherical layer is considered. The results are obtained in terms of the Bhatnagar–Gross–Krook model and the Boltzmann collision integral. A general (independent of the form of the kinetic equation and the solution method) expression for the heat flux as a function of the energy accommodation coefficient is derived. The results are compared with the experiment and the analytical results obtained previously. © 2005 Pleiades Publishing, Inc.

Investigation of heat transfer processes remains a challenging problem in the kinetic theory of gases. In particular, measurements of the heat flux from a physical body are used for gaining a better insight into gas–surface interaction [1–3]. In the majority of the relevant studies, the emphasis is on a heat flux from a single sphere or on heat exchange between coaxial cylinders. At the same time, nobody, except Lees [4], has centered due attention on the analysis of a heat flux between two concentric spheres.

As in [5], here we employ a version of the method of half-spatial moments. The results are obtained in terms of the Bhatnagar–Gross–Krook (BGK) model [6] and Boltzmann collision integral for molecules interacting as elastic hard spheres. The procedure of calculating all the moments of this integral is described elsewhere [7]. General (independent of both the form of the kinetic equation and the solution method) expressions relating the heat flux to the energy accommodation coefficient are presented. The results are compared with experimental data and the analytical data obtained previously.

As the unit of length, we take

$$l = \frac{3\lambda}{\sqrt{\pi}}, \qquad (1)$$

where

$$\lambda = \frac{\kappa}{15 n_0} \sqrt{\frac{8\pi m}{k^3 T_0}}$$

is the mean free path of gas molecules [8] and κ is the thermal conductivity of the gas.

Let us consider a gas layer between two concentric spheres with radii $R_1 < R_2$ and constant surface temperatures $T_s^1 > T_s^2$. Assume that the temperature drop

$\Delta T_s = T_s^1 - T_s^2$ is sufficiently small that our consideration may be restricted to a linear approximation.

The gas law is stated as [8]

$$\mathbf{V} \cdot \nabla f = J_{col}[f]. \qquad (2)$$

Here, $\mathbf{V}$ is the intrinsic thermal velocity of gas molecules, f is the distribution function, and J_{col} is the collision integral.

By virtue of the linearity condition assumed above, Eq. (2) can be represented in the form

$$f = f_0(1 + \varphi),$$

where

$$f_0 = n_0 \left(\frac{m}{2\pi k T_0} \right)^{3/2} \exp(-C^2)$$

is the equilibrium (Maxwellian) distribution function; m is the molecular mass; k is the Boltzmann constant; $\mathbf{C} = \mathbf{V}\sqrt{m/2kT_0}$; and T_0 and n_0 are, respectively, some temperature and concentration of gas molecules that are taken to be equilibrium.

Passing to the spherical system of coordinates with the origin at the center of the spheres and assuming that function φ depends on only r, C, and C_r, we have from (2)

$$C_r \frac{\partial \varphi}{\partial r} + \frac{C^2 - C_r^2}{r} \frac{\partial \varphi}{\partial C_r} = I_{col}[\varphi], \qquad (3)$$

where I_{col} is the linearized collision operator corresponding to J_{col}. Derivative $\dfrac{\partial}{\partial C_r}$ is taken with respect to the projection alone and does not affect the functions of the velocity magnitude.

As boundary conditions, we use diffuse reflections of gas molecules from the surfaces of either sphere,

$$\varphi|_{r=R_k} = \Phi_r^k = \frac{\Delta n_r^k}{n_0} + \left(C^2 - \frac{3}{2}\right)\frac{\Delta T_r^k}{T_0}, \qquad (4)$$

where

$$\Delta T_r^k = T_r^k - T_0, \quad \Delta n_r^k = n_r^k - n_0;$$

and T_r^k and n_r^k are, respectively, the temperature and concentration of the molecules reflected from the surface of a kth sphere.

Quantities T_r^k and n_r^k are defined by the requirement that the bulk motion of the gas be absent,

$$\int C_r \varphi \exp(-C^2) d^3 C = 0, \qquad (5)$$

and by the way of energy accommodation,

$$\alpha_K = \frac{E_i^k + E_r^k}{E_i^k + E_s^k}. \qquad (6)$$

Here,

$$E_i^k = \pi^{-3/2} \int_{(-1)^k C_r < 0} C_r C^2 \varphi(R_k) \exp(-C^2) d^3 C \qquad (7)$$

and

$$E_r^k = \pi^{-3/2} \int_{(-1)^k C_r > 0} C_r C^2 \Phi_r^k \exp(-C^2) d^3 C \qquad (8)$$

are, respectively, the incoming dimensionless energy flux due to incident molecules and the outgoing dimensionless energy flux due to the molecules reflected from the surface of a kth sphere and

$$E_s^k = \pi^{-3/2} \int_{(-1)^k C_r > 0} C_r C^2 \Phi_s^k \exp(-C^2) d^3 C$$

is the energy that would be carried away by the molecules that reflect with temperature T_s^k, i.e., by those obeying the distribution function

$$\Phi_s^k = \frac{\Delta n_s^k}{n_0} + \left(C^2 - \frac{3}{2}\right)\frac{T_s^k - T_0}{T_0}.$$

In view of linearity, any of the gas characteristics can be represented in the form

$$F = F^* \frac{\Delta T_r}{T_0}, \quad \Delta T_r = T_r^1 - T_r^2.$$

Hereafter, we mark by the asterisk the values related to $\Delta T_r/T_0$, or, equivalently, to $\Delta T_s/T_0$ if the energy accommodation is complete.

Let us first consider the internal sphere. From (5), (7), and (8), we find that

$$\frac{\Delta n_r^1}{n_0} = -\frac{1 + 4I_0^1}{2}\frac{\Delta T_r}{T_0},$$

$$E_i^1 = \frac{I_1^1}{\sqrt{\pi}}\frac{\Delta T_r}{T_0}, \quad E_r^1 = \frac{1 - 2I_0^1}{\sqrt{\pi}}\frac{\Delta T_r}{T_0}, \qquad (9)$$

$$I_i^1 = \pi^{-1} \int_{C_r < 0} C_r C^{2i} \varphi^*(R_1) \exp(-C^2) d^3 C.$$

Similarly,

$$\frac{\Delta n_s^1}{n_0} = -\frac{1}{2}\frac{T_s^1 - T_0}{T_0} - 2I_0^1\frac{\Delta T_r}{T_0},$$

$$E_s^1 = \frac{T_s^1 - T_0}{T_0} - \frac{2I_0^1}{\sqrt{\pi}}\frac{\Delta T_r}{T_0}.$$

Substituting these expressions into Eq. (6) yields

$$\alpha_1(T_s^1 - T_0) = \Delta T_r(1 + (1 - \alpha_1)(I_1^1 - 2I_0^1)). \qquad (10)$$

The desired energy flux is defined by the relationship

$$q = n_0 \sqrt{\frac{2k^3 T_0^3}{m}} Q,$$

$$Q = \pi^{-3/2} \int C_r C^2 \varphi \exp(-C^2) d^3 C. \qquad (11)$$

On the strength of the energy conservation law and the linearity of the problem, we may write

$$Q = Q^* \frac{R^2}{r^2}\frac{\Delta T_r}{T_0}, \qquad (12)$$

where Q^* is a dimensionless constant. Without loss of generality, it can be set equal to the density of the energy flux from the surface of the internal sphere per unit relative temperature difference under the condition of complete energy accommodation.

On the other hand,

$$Q|_{r=R_1} = E_r^1 + E_i^1 = \frac{1 - 2I_0^1 + I_1^1}{\sqrt{\pi}}\frac{\Delta T_r}{T_0}. \qquad (13)$$

Comparing Eqs. (12) and (13), we get

$$Q^* = \frac{1 - 2I_0^1 + I_1^1}{\sqrt{\pi}}.$$

Substitution of this expression into Eq. (10) gives

$$\alpha_1(T_s^1 - T_0) = (\sqrt{\pi}(1 - \alpha_1)Q^* + \alpha_1)\Delta T_r. \qquad (14)$$

Consider now the external sphere. Without loss of generality, the temperature and concentration of the molecules reflected from it can be viewed as equilib-

rium; i.e., $T_0 = T_r^2$ and $n_0 = n_r^2$. Under this assumption, we obtain

$$\Phi_r^2 = 0, \quad E_r^2 = 0, \quad E_i^2 = \frac{I_1^2}{\sqrt{\pi}} \frac{\Delta T_r}{T_0},$$

$$E_s^2 = -\frac{1}{\sqrt{\pi}} \frac{T_s^2 - T_0}{T_0},$$

$$\alpha_2(T_s^2 - T_0) = -I_1^2(1 - \alpha_2)\Delta T_r, \tag{15}$$

$$I_1^2 = \pi^{-1} \int\limits_{C_r > 0} C_r C^{2i} \varphi^*(R_2) \exp(-C^2) d^3 C.$$

Taking into account that

$$E_i^2 = Q|_{r = R_2} = Q^* \frac{R_1^2}{R_2^2} \frac{\Delta T_r}{T_0},$$

we arrive at

$$I_1^2 = \sqrt{\pi} Q^* \frac{R_1^2}{R_2^2}.$$

Eventually,

$$\alpha_2(T_s^1 - T_0) = -\sqrt{\pi}(1 - \alpha_2)Q^* \frac{R_1^2}{R_2^2} \Delta T_r. \tag{16}$$

Comparing Eqs. (14) and (16) yields

$$\Delta T_r = \left(\sqrt{\pi} Q^* \left(\frac{1}{\alpha_1} - 1 + \frac{R_1^2}{R_2^2} \left(\frac{1}{\alpha_2} - 1 \right) \right) + 1 \right)^{-1} \Delta T_s.$$

Accordingly,

$$Q = \left(\sqrt{\pi} \left(\frac{1}{\alpha_1} - 1 + \frac{R_1^2}{R_2^2} \left(\frac{1}{\alpha_2} - 1 \right) \right) + \frac{1}{Q^*} \right)^{-1} \frac{\Delta T_s}{T_0}. \tag{17}$$

In the collisionless (ballistic) mode, which sets in when the distance between the spheres or the radius of the internal sphere is much less than the molecular free path, one can evidently neglect a change in the distribution function in the volume and assume that function φ^* is zero in the domain of integration (9) and integrals I_i^k vanish. Therefore, in this approximation,

$$Q^* = Q_{\text{bal}}^* = \pi^{-1/2}.$$

Accordingly,

$$Q_{\text{bal}} = \pi^{-1/2} \left(\frac{1}{\alpha_1} + \frac{R_1^2}{R_2^2} \left(\frac{1}{\alpha_2} - 1 \right) \right)^{-1} \frac{\Delta T_s}{T_0},$$

$$\frac{Q}{Q_{\text{bal}}} = \left(\frac{1}{\alpha_1} + \frac{R_1^2}{R_2^2} \left(\frac{1}{\alpha_2} - 1 \right) \right) \tag{18}$$

$$\times \left(\frac{1}{\alpha_1} + \frac{R_1^2}{R_2^2} \left(\frac{1}{\alpha_2} - 1 \right) + \frac{Q_{\text{bal}}^*}{Q^*} - 1 \right)^{-1}.$$

The above relationship has a general character and depends neither on the form of the kinetic equation nor on the solution method.

To calculate Q^*, we will take advantage of the general idea of the method of half-spatial moments, specifically, representing a solution to Eq. (3) as a sum of given velocity polynomials.

Note that the distribution function exhibits a discontinuity on the surface of either sphere and also the fact that any point inside the gas can be assigned three invariant cones in the velocity space, the boundaries of these cones being crossed by molecules only when they collide with each other. Bearing this in mind, we write

$$\varphi = \sum_{i=1}^{4} \varphi_i H_i,$$

where

$$\varphi_i = a_1^i + a_2^i C^2 + a_3^i C_r + a_4^i C_r C^2, \tag{19}$$

$$H_1 = H(C_r - \gamma C), \quad H_2 = H(C_r) - H_1,$$

$$H_4 = H(-C_r - \gamma C),$$

$$H_3 = H(-C_r) - H_4, \quad \gamma = \sqrt{1 - \frac{R_1^2}{r^2}},$$

and $H(x) = \dfrac{|x| + x}{2x}$ is the Heaviside function.

Coefficients a_j^i, which depend only on the distance to the centers of the spheres, are determined from a system of the moment equations that can be obtained by successively multiplying Eq. (3) with φ in form (19) by $H_i \exp(-C^2)$, $C^2 H_i \exp(-C^2)$, $C_r H_i \exp(-C^2)$, and $C_r C^2 H_i \exp(-C^2)$ and integrating the result over the velocity space. Omitting here calculations that are similar to, but much more tedious than, those presented in [5, 7], we note that, by virtue of (4), the desired solution must satisfy the conditions

$$a_1^1 = \frac{\Delta n_r}{n_0} - \frac{3}{2} \frac{\Delta T_r}{T_0}, \quad a_2^1 = \frac{\Delta T_r}{T_0},$$

$$a_3^1 = a_4^1 = 0, \quad \text{at} \quad r = R_1, \tag{20}$$

$$a_i^3 = a_i^4 = 0 \ (i = 1, 2, 3, 4), \text{ at } r = R_2. \tag{21}$$

The value of $\Delta n_r/n_0$ is defined by condition (5) (the absence of mass transfer). Therefore, the first expression in (20) can be replaced by the equality

$$4(a_1^1 - a_1^4) + 8(a_2^1 - a_2^4)$$
$$+ \sqrt{\pi}(2a_3^1 + 5a_4^1 + 2a_3^4 + 5a_2^4) = 0, \text{ at } r = R_1, \tag{22}$$

which follows from (5).

Next, it is necessary to take into consideration that the system of differential equations presented above has a singularity on the surface of the internal sphere, which stems from the collapse of the central (i.e., corresponding to functions φ_2 and φ_3) cones in the velocity space. Expanding the desired solution in powers of $\xi = r - R_1$ and including the requirement of the finiteness of the distribution function, we get four additional conditions,

$$a_i^2\gamma^2 = a_{i+2}\gamma^3 = 0 \ (i = 1, 2), \text{ at } r = R_1. \tag{23}$$

Thus, the boundary conditions are specified by (i) the distribution function for the molecules reflected from the external sphere (i.e., by the values of functions φ_3 and φ_4 at $r = R_2$), (ii) the distribution function for the molecules reflected from the internal sphere (i.e., by the value of φ_1 at $r = R_1$), and (iii) the requirement that function φ_2 be finite at $r \longrightarrow R_1$.

The heat flux from a unit surface of the internal sphere can be determined as

$$Q = \pi^{-3/2}\int C_r C^2 \varphi(R_1) \exp(-C^2) d^3 C$$
$$= \frac{a_1^1 - a_1^4}{\sqrt{\pi}} + 3\frac{a_2^1 - a_2^4}{\sqrt{\pi}} + \frac{5}{8}(a_3^1 + a_3^4) + \frac{35}{16}(a_4^1 + a_2^4).$$

Figure 1 presents numerical solutions to the system of equations considered subject to the boundary conditions listed above for $R_1/R_2 = 0.1, 0.5, 0.9,$ and 0.99 (the bundles of curves 1, 2, 3, and 4, respectively). At $R_1/R_2 \lesssim 0.01$, the heat flux becomes independent of the ratio between the radii and nearly coincides with the results obtained in the limit $R_1/R_2 \longrightarrow 0$ (Fig. 2), which is equivalent to the case of a solitary sphere. Specific values of the heat flux are listed in the table.

It should be noted that most authors treat experimental data using the Lees method. In this case, the distribution function is taken in the form

$$\varphi = (a_1 + a_2 C^2)H_1 + (a_3 + a_4)(1 - H_1),$$

and the system of moment equations is constructed by using factors 1, C^2, C_r, and $C_r C^2$. This yields (see, e.g., [2])

$$Q^* = \left(\frac{4}{5}R_1\left(1 - \frac{R_1}{R_2}\right) + \sqrt{\pi}\right)^{-1}. \tag{24}$$

The disadvantages of such an approach were discussed, in particular, in [5, 7]. The values of Q^* calcu-
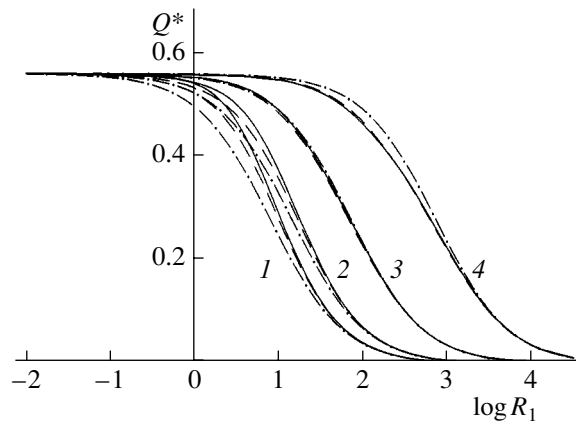
Fig. 1. Heat flux vs. the radius of the internal sphere for $R_1/R_2 = $ (1) 0.1, (2) 0.5, (3) 0.9, and (4) 0.99. The continuous curves correspond to the model of elastic hard spheres; the dashed curves, to the BGK model of the collision integral; and the dot-and-dash curves, to the standard Lees method.

Fig. 2. Heat flux from a solitary sphere. The symbols are the results obtained by direct numerical integration [9] in terms of ($\bigcirc$) the BGK collision integral and ($\bullet$) the model of hard spheres and ($\times$) by the variational method [11].

Fig. 3. Ratio Q/Q_{th} calculated for $R_2 = 7R_1$ and $\alpha_1 = \alpha_2 = 0.3$. ($\square$) Data points [10].

Values of Q^* for a solitary sphere

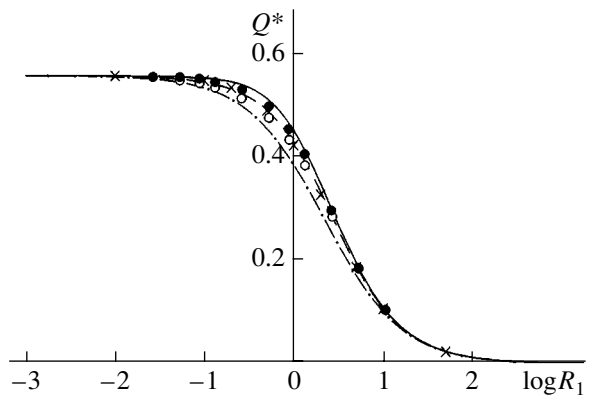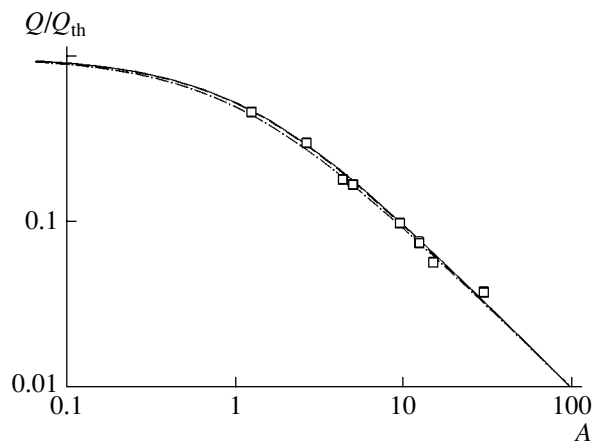| $k = \dfrac{15}{8R_1}$	BGK model		Molecules as hard spheres	
	1	2	3	4
0.05	0.10342	0.10334	0.10494	0.1064
0.1	0.18247	0.18125	0.18689	0.1885
0.2	0.28891	0.28416	0.30055	0.2996
0.4	0.39594	0.38751	0.41684	0.4095
0.6	0.44607	0.43712	0.46986	0.4595
1	0.49175	0.48395	0.51494	0.5032
2	0.52810	0.52331	0.54561	0.5361
4	0.54624	0.54384	0.55720	0.5511
6	0.55220	0.55065	0.56005	0.5559
10	0.55694	0.55593	0.56186	0.5593
20	0.56051	0.55954	0.56298	0.5616

Note: The first and the third columns list the results calculated by the method of moments that is described in this paper, the second column shows the results obtained by solving the system of integral equations that follows from (3) [12], and the fourth column presents the results of direct numerical integration of the kinetic equation [9].

lated by formula (24) are shown by the dash-and-dot lines in Fig. 1.

As is seen from the table and plots, the results of solving the kinetic equation by the method proposed agree with the values calculated in [9, 12] within 2.5% for both the BGK and Boltzmann collision integrals. The variational method provides nearly the same order of accuracy. However, its application to the class of problems similar to that considered in this paper is restricted to model equations only. The error involved in the standard Lees method is 10% or higher.

Figure 3 plots ratio Q/Q_{th} calculated for $R_2 = 7R_1$ and $\alpha_1 = \alpha_2 = 0.3$ (i.e., under the same conditions as those used to measure the heat flux through a helium layer bounded by glass spheres [11]). Here,

$$A = \frac{5\sqrt{\pi}}{4R_1}\left(1 - \frac{R_1}{R_2}\right)\left(\frac{1}{\alpha_1} + \frac{R_1^2}{R_2^2}\left(\frac{1}{\alpha_2} - 1\right)\right),$$

and

$$Q_{th} = \frac{5}{4R_1}\left(1 - \frac{R_1}{R_2}\right)^{-1}$$

is the heat flux from a unit surface of the internal sphere at a unit temperature drop between the surfaces that is calculated in the thermodynamic limit.

REFERENCES

1. O. A. Kolenchits, *Thermal Accommodation of Gas–Solid System* (Nauka i Tekhnika, Minsk, 1977) [in Russian].
2. Yu. A. Koshmarov and Yu. A. Ryzhov, *Applied Dynamics of Rarefied Gas* (Mashinostroenie, Moscow, 1977) [in Russian].
3. S. F. Borisov, N. F. Balakhonov, and V. A. Gubanov, *Interaction of Gases with Solid Surfaces* (Nauka, Moscow, 1988) [in Russian].
4. L. Lees, J. Soc. Ind. Appl. Math. **13**, 278 (1965).
5. S. A. Savkov and A. A. Yushkanov, Zh. Tekh. Fiz. **70** (11), 9 (2000) [Tech. Phys. **45**, 1383 (2000)].
6. P. L. Bhatnagar, E. P. Gross, and M. A. Krook, Phys. Rev. **94**, 511 (1954).
7. S. A. Savkov, A. A. Yushkanov, and Yu. I. Yalamov, Teplofiz. Vys. Temp. **39**, 657 (2001).
8. C. Cercignani, *Theory and Application of the Boltzmann Equation* (Elsevier, New York, 1975; Mir, Moscow, 1978).
9. S. Takata, Y. Sone, D. Lhuillier, and M. Wakabayshi, Comput. Math. Appl. **35**, 193 (1998).
10. C. Cercignani and C. D. Pagani, Phys. Fluids **9**, 1167 (1966).
11. G. S. Springer and S. F. Wan, AIAA J., No. 8, 800 (1966).
12. S. A. Savkov, Inzh.-Fiz. Zh. **75**, 111 (2002).

Translated by A. Sidorova

GAS DISCHARGES,
PLASMA

Electromagnetic Instability of a Homogeneous Plasma in Interstellar Space

A. S. Baranov

Pulkovo Astronomical Observatory, Russian Academy of Sciences,
Pulkovskoe sh. 65/1, St. Petersburg, 196140 Russia
Received February 26, 2004; in final form, September 20, 2004

Abstract—A study is made of an electromagnetic instability of a homogeneous plasma with an arbitrary velocity distribution in interstellar space on long time scales typical of gas-dust clouds without a significant magnetic field. It is shown that such instabilities develop on time scales of a few months or years and that it is only a narrow class of nearly spherical solutions that conforms to the stability requirements. © 2005 Pleiades Publishing, Inc.

INTRODUCTION

The problem of stability of an inhomogeneous plasma is important in the physics of interstellar and interplanetary media, especially because not all particle velocity distributions can occur in a natural (even collisionless) plasma. In fact, unstable distributions should rapidly rearrange themselves into stable ones; this circumstance should be taken into account in constructing theoretical models of the evolution of a medium [1, 2]. An example is the isotropization of an interstellar medium in which two particle flows meet [3]. The authors of [3] emphasized, however, the particular character of this situation—the model of two overlapping (interpenetrating) particle flows is not the only one possible and many other anisotropic models may be constructed as well. In addition, because of the huge spatiotemporal scales of plasma objects in space, the basic accepted views about the theory of plasma stability, which is now better suited for laboratory applications, are already in need of revision. As will be seen below, the conclusions derived here for space plasmas are somewhat unusual; in particular, a particle velocity distribution centered in a conventional manner remains stable only when the second harmonic in the expansion of the phase density in spherical functions vanishes (rather than satisfies a certain inequality). Physically, this indicates that the velocity diagram becomes isotropic on time scales far shorter than the general evolutionary scales.

Existing theories are usually restricted to the so-called potential plasma instabilities [4, 5], i.e., the perturbations are treated only in terms of their electrostatic field. It is known, however, that there is the Weibel electromagnetic instability [6], which is rarely mentioned in the literature and which involves a perturbed magnetic field. However, the growth rate of this instability is substantially slower than that of the potential instabilities. This is why the electromagnetic instability is almost never taken into account in describing ground-based experiments: under terrestrial conditions, this instability does not have time to develop in the course of experiment. In astrophysics, the situation is different: the relevant time scales are relatively long (a matter of years or more). At the same time, observational data and general considerations are insufficient to choose between different realistic models [7, 8]; therefore, stability criteria have to be invoked.

It should be stressed that electromagnetic instability should also manifest itself at moderate (nonrelativistic) particle velocities; in this case, however, the wavenumbers k of the instability are small and its growth rate is slow. This can be readily demonstrated, e.g., by the elementary example presented in [5]. The Weibel instability of two discrete particle flows with the same density n_0 and with the velocities $(0, 0, -V_0)$ and $(0, 0, V_0)$ (the wave vector being oriented along the x axis) is described by the dispersion relation

$$\frac{c^2 k^2}{\omega^2} - 1 + \frac{2\omega_p^2}{\omega^2}\left(1 + \frac{k^2 V_0^2}{\omega^2}\right) = 0, \tag{1}$$

where

$$\omega_p = \sqrt{\frac{4\pi e^2 n_0}{m}} \tag{2}$$

is the Langmuir frequency. Here, e and m are the charge and mass of a particle of a given species, c is the speed of light, and ω is the oscillation frequency. We write the solution to dispersion relation (1) in the form

$$\omega^2 = -\frac{4\omega_p^2 k^2 V_0^2}{2\omega_p^2 + k^2 c^2 + \sqrt{(2\omega_p^2 + k^2 c^2)^2 + 8\omega_p^2 k^2 V_0}},$$

and, in the limit of small wavenumbers k, obtain

$$\omega \approx ikV_0, \tag{3}$$

which agrees with the conclusion obtained earlier in [5] that the system is unstable.

Note that asymptotic formula (3) does not contain the speed of light. It will be seen later that this is a common feature of a wide class of nonrelativistic velocity distributions: although the electromagnetic effects are formally relativistic, the parameter c drops out in the limit $k \longrightarrow 0$. It is another matter, however, that, in the example at hand, the Weibel instability occurs over the entire range of k values and that, in the case of smoother velocity distributions, there is a critical k value, which approaches zero as the ratio of the mean velocity to the speed of light decreases. The critical value of k will be estimated at the end of the paper.

In order for the analysis to be more convenient from the formal point of view and to comply better with the existing literature on the subject [9], we begin with a model of N discrete beams, which can be generalized to a model with a continuous velocity distribution in a trivial way. First of all, keeping in mind further possible generalization, we nevertheless consider arbitrary particle velocities and then restrict ourselves to the limit $V \ll c$ in order to single out and investigate the electromagnetic instability.

BASIC EQUATIONS

We adopt a model of a homogeneous plasma consisting of the particle flows designated as $i = 1, 2, \ldots, N$, with the corresponding flow velocity vectors $V_i(u_i, v_i, w_i)$, spatial densities n_i particle masses m_i, and particle charges e_i. The model also assumes a background with a space charge density of opposite sign:

$$d = -\sum_i n_i e_i.$$

We consider the propagation of a wave in the linear approximation, assuming that the z axis points along the wave vector $\mathbf{k}$. In this case, all of the linearized perturbed quantities contain the factor $\exp(\lambda t + ikz)$, where t is the time and λ is the growth rate.

If there were no perturbation, each particle would move by inertia,

$$\mathbf{r} = \mathbf{r}_{0i} + \mathbf{V}_i t, \tag{4}$$

where the vector $\mathbf{r}_{0i}(x_{0i}, y_{0i}, z_{0i})$ refers to a certain initial position of a particle.

In fact, we must take into account small perturbations of both the electric ($\mathbf{E}$) and magnetic ($\mathbf{H}$) fields. In order to obtain an evolutionary equation for the velocity vector of an individual particle, we first write the famil-

iar equation for the particle momentum

$$\frac{d\mathbf{p}_i}{dt} = e_i\left(\mathbf{E} + \frac{\mathbf{V}_i \times \mathbf{H}}{c}\right) \tag{5}$$

and then take into account the algebraic relationships between the velocity and the momentum,

$$\mathbf{p} = \frac{m\mathbf{V}}{\sqrt{1 - \dfrac{V^2}{c^2}}}, \quad \mathbf{V} = \frac{\mathbf{p}}{\sqrt{m^2 + \dfrac{p^2}{c^2}}},$$

$$\sqrt{m^2 + \frac{p^2}{c^2}} = \frac{m}{\sqrt{1 - \dfrac{V^2}{c^2}}}. \tag{6}$$

Following what was said above, we can single out the dependence on time and coordinates in the expressions for $\mathbf{E}$ and $\mathbf{H}$:

$$\mathbf{E} = \boldsymbol{\varepsilon} e^{\lambda t + ikz}, \quad \mathbf{H} = \mathbf{h} e^{\lambda t + ikz}.$$

Linearizing Eq. (5) yields

$$\frac{d\delta\mathbf{p}_i}{dt} = e_i\left(\boldsymbol{\varepsilon} + \frac{\mathbf{V}_i \times \mathbf{h}}{c}\right)e^{\lambda t + ikz}. \tag{7}$$

Here and below, the symbol δ denotes the perturbation of a given quantity.

Note, however, that the z coordinate on the right-hand side of Eqs. (7) refers to the instantaneous position of a particle and thus should be determined from Eq. (4); in this case, the perturbation gives rise to only a small (second-order) correction to Eq. (4). With these remarks in mind, we integrate Eq. (7) from the infinite past to find

$$\delta\mathbf{p}_i = e_i\left(\boldsymbol{\varepsilon} + \frac{\mathbf{V}_i \times \mathbf{h}}{c}\right)\int_{-\infty}^{t} e^{ik(z_{0i} + w_i t) + \lambda t}dt$$

$$= e_i\left(\boldsymbol{\varepsilon} + \frac{\mathbf{V}_i \times \mathbf{h}}{c}\right)\frac{e^{\lambda t + ikz}}{\lambda + ikw_i}. \tag{8}$$

On the right-hand side of this equation, the perturbed momentum is expressed in terms of the z coordinate the particle would have in the absence of perturbation. Since we are interested exclusively in unstable perturbations, we set $\operatorname{Re}\lambda > 0$. According to relationships (6), we obtain the following equation for the velocity perturbation:

$$\delta\mathbf{V}_i = \frac{e_i\sqrt{1 - \dfrac{V_i^2}{c^2}}}{m_i(\lambda + ikw_i)}e^{\lambda t + ikz}$$

$$\times \left[\boldsymbol{\varepsilon} + \frac{\mathbf{V}_i \times \mathbf{h}}{c} - \frac{(\mathbf{V}_i \cdot \boldsymbol{\varepsilon})\mathbf{V}_i}{c^2}\right]. \tag{9}$$

We again integrate over t and use for the moment the expression for the instantaneous z coordinate. As a result, according to Eq. (4), we get

$$\delta \mathbf{r}_i = \frac{e_i \sqrt{1 - \dfrac{V_i^2}{c^2}}}{m_i(\lambda + ikw_i)^2} e^{\lambda t + ikz}$$

$$\times \left[\boldsymbol{\varepsilon} + \frac{\mathbf{V}_i \times \mathbf{h}}{c} - \frac{(\mathbf{V}_i \cdot \boldsymbol{\varepsilon})\mathbf{V}_i}{c^2} \right]. \tag{10}$$

The total current at each point should be found from the relationship

$$\mathbf{J} = \mathbf{j} e^{\lambda t + ikz}.$$

The corresponding inverse relationship has the form

$$\mathbf{j} = \frac{ke^{-\lambda t}}{2\pi} \int_0^{2\pi/k} \mathbf{J} e^{-ikz} dz.$$

The current J is the sum of all the contributions $e_i n_i \mathbf{V}_i$ from each particle in each plasma flow; of course, the contributions should also be summed over the subscript i. In the expansion of the quantity $e_i n_i \mathbf{V}_i \exp(-ikz)$, the zero-order terms entirely cancel each other, reflecting the fact that, in an unperturbed state, there are no currents. Accordingly, we have

$$\delta \mathbf{j} = \mathbf{j} = \frac{ke^{-\lambda t}}{2\pi} \sum_{i=1}^N e_i \int_0^{2\pi/k} (\delta \mathbf{V}_i - ik\mathbf{V}_i \delta z) e^{-ikz} dz$$

or

$$\delta \mathbf{j} = \sum_{i=1}^N n_i \frac{e_i^2}{m_i} \left\{ \left[\boldsymbol{\varepsilon} + \frac{\mathbf{V}_i \times \mathbf{h}}{c} - \frac{(\mathbf{V}_i \cdot \boldsymbol{\varepsilon})\mathbf{V}_i}{c^2} \right] L \right.$$

$$\left. - ik\mathbf{V}_i \left[\varepsilon_z + \frac{(\mathbf{V}_i \times \mathbf{h})_z}{c} - \frac{w_i(\mathbf{V}_i \cdot \boldsymbol{\varepsilon})}{c^2} \right] L^2 \right\} \tag{11}$$

$$\times \sqrt{1 - \frac{V_i^2}{c^2}} \quad (\mathrm{Re}\,\lambda > 0).$$

This way of deriving the perturbed current is advantageous in that it clearly illustrates how the factor $L = 1/(\lambda + ikw)$ arises in this expression. The sought-for current can also be obtained by linearizing the hydrodynamic equations for each of the flows; in this way, we again arrive as expected at the above expression (11) for $\delta\mathbf{j}$. Note that, if the quantities k and λ are assumed to be of the same order of magnitude, then the generated current is inversely proportional to k.

Formula (11) should be used in combination with Maxwell's equations. However, in the case at hand, we have $H_z = 0$. Therefore, in terms of the wave vector

$\mathbf{k}(0, 0, k)$, Maxwell's equations become

$$\frac{\lambda}{c}\mathbf{h} = i(\boldsymbol{\varepsilon} \times \mathbf{k}), \quad \frac{\lambda}{c}\boldsymbol{\varepsilon} + \frac{4\pi}{c}\delta\mathbf{j} + i(\mathbf{h} \times \mathbf{k}) = 0. \tag{12}$$

Following the scheme of analysis adopted here, we consider the long-wavelength limit: $k \longrightarrow 0$. Let the quantities k and λ be of the same order of smallness. Note that the generated current is generally inversely proportional to k. Consequently, in order to satisfy Eqs. (12), the relationship between k and λ should be chosen in such a way that, when the vector $\mathbf{h}$ from the first of Eqs. (12) is substituted into relationship (11), the right-hand side of the latter, i.e.,

$$\sum_{i=1}^N \frac{n_i e_i^2}{m_i} \left[\frac{\mathbf{R}_i}{\lambda + ik\mathbf{V}_i} - i\mathbf{V}_i \frac{k\mathbf{R}_i}{(\lambda + ik\mathbf{V}_i)^2} \right], \tag{13}$$

where

$$\mathbf{R}_i = \boldsymbol{\varepsilon} + \frac{i}{\lambda}(\mathbf{V}_i \times (\boldsymbol{\varepsilon} \times \mathbf{k})) - \frac{(\mathbf{V}_i \cdot \boldsymbol{\varepsilon})\mathbf{V}_i}{c^2},$$

becomes a null vector (to within relatively small corrections).

Recall that, at this step of the analysis, we are ignoring a specific relativistic correction—the last terms in the expression for $\mathbf{R}_i$. We also restrict ourselves to a certain special class of velocity distributions. Namely, we assume that all the particles are of the same species and that the velocity distribution (i) possesses a rotational symmetry, $f(u, v, w) = \tilde{f}(\rho, w)$ ($\rho = \sqrt{u^2 + v^2}$), and (ii) is symmetric with respect to the equatorial plane, $f(u, v, -w) = f(u, v, w)$.

In the above specific formulation of the problem, it is preferable to identify the z axis with the symmetry axis of the velocity diagram; in this case, the wave vector $\mathbf{k}$ will be in an inclined position. Without loss of generality, the plane that passes through the symmetry axis and contains the wave vector can be assumed to be the xz plane. Accordingly, the wave vector has the components $(k\sin\sigma, 0, k\cos\sigma)$, where σ is the angle between the wave vector and the symmetry axis. In formula (13), we switch from summation to integration (for one particle species) and consider the components of the expression obtained. All terms in the y component of this expression contain the quantity ε_y, so we divide by this quantity to obtain

$$\iiint \left\{ 1 - \frac{k^2 v^2}{[\lambda + ik(u\sin\sigma + w\cos\sigma)]^2} \right\} f\, du\, dv\, dw = 0 \tag{14}$$

$$(\varepsilon_y \neq 0).$$

For the quantities ε_x and ε_z, we arrive at the following set of two coupled equations:

$$A\varepsilon_x + B\varepsilon_z = 0, \quad B\varepsilon_x + C\varepsilon_z = 0, \tag{15}$$

where

$$A = \iiint \left\{ \frac{\lambda + ik(w\cos\sigma - u\sin\sigma)}{\lambda + ik(u\sin\sigma + w\cos\sigma)} \right.$$

$$\left. - \frac{k^2 u^2}{[\lambda + ik(u\sin\sigma + w\cos\sigma)]^2} \right\} f du dv dw,$$

$$B = -\iiint \left\{ \frac{ik(w\sin\sigma + u\cos\sigma)}{\lambda + ik(u\sin\sigma + w\cos\sigma)} \right.$$

$$\left. + \frac{k^2 uw}{[\lambda + ik(u\sin\sigma + w\cos\sigma)]^2} \right\} f du dv dw,$$

$$C = \iiint \left\{ \frac{\lambda + ik(u\sin\sigma - w\cos\sigma)}{\lambda + ik(u\sin\sigma + w\cos\sigma)} \right.$$

$$\left. - \frac{k^2 w^2}{[\lambda + ik(w\cos\sigma + u\sin\sigma)]^2} \right\} f du dv dw.$$

ANALYSIS OF THE BASIC EQUATIONS

Note that, according to the above conditions (i) and (ii), changing the sign of the coordinates u, v, and w leaves the function f unchanged but changes the quantities A, B, and C into their complex conjugates; i.e., these quantities are real for real values of λ. As $\lambda \longrightarrow \infty$, the left-hand side of Eq. (14) approaches its limiting value

$$\iiint f du dv dw = \nu > 0,$$

where ν is the spatial density of the particles of a given species. Analogously, in the same limit, we have $A = C = \nu$ and $AC - B^2 = \nu^2 > 0$. By continuity, Eq. (14) and the equation $AC - B^2 = 0$ necessarily have the roots $\lambda > 0$, which imply instability, provided that the left-hand side of Eq. (14) and the expression $AC - B^2$ are negative for small λ values. These are sufficient conditions for the plasma instability. In order to analyze them in more detail, we first turn to Eq. (14). To make the analysis more illustrative, we begin by considering the class of velocity distributions on a unit sphere. By virtue of the above symmetry conditions, the expansion of the surface density F on this sphere in spherical functions has the form

$$F = \sum_{n=0}^{\infty} c_n P_n(\cos\theta), \qquad (16)$$

where the sum is taken only over even values of n. Here and below, we use the ordinary spherical coordinates: θ is the angle with respect to the w axis, and μ is the azimuthal angle in the equatorial plane (u, v). It is conve-

nient to rotate the coordinate system so as to pass over to the new velocity components,

$$u_1 = u\cos\sigma - w\sin\sigma, \quad v_1 = v,$$
$$w_1 = u\sin\sigma + w\cos\sigma. \qquad (17)$$

Applying transformation of rotation (17) to expansion (16) reduces to using the familiar theorem of summation of the spherical functions. As a result, the left-hand side of Eq. (14) becomes

$$M = \iint \left[1 - \frac{k^2 \sin^2\theta_1 \sin^2\mu_1}{(\lambda + ik\cos\theta_1)^2} \right]$$

$$\times \sum_{n=0}^{\infty} c_n [P_n(\cos\theta_1) P_n(\cos\sigma)$$

$$+ 2\sum_{v=0}^{\infty} \frac{(n-v)!}{(n+v)!} P_n^{(v)}(\cos\theta_1) \qquad (18)$$

$$\times P_n^{(v)}(\cos\sigma)\cos v\mu_1 \bigg] \sin\theta_1 d\theta_1 d\mu_1,$$

where θ_1 and μ_1 are again the coordinates on the sphere, but with respect to the new polar axis w_1, and the Legendre associated polynomials are abbreviated in the conventional notation $P_n^{(v)}$.

Integration over μ_1 on the right-hand side of expression (18) is a simple matter; the result is that there remain only the terms with $v = 0$ and 2:

$$M = \pi \sum_{n=0}^{\infty} c_n \left\{ P_n(\cos\sigma) \right.$$

$$\times \int_0^{\pi} \left[2 - \frac{k^2 \sin^2\theta}{(\lambda + ik\cos\theta)^2} \right] P_n(\cos\theta)\sin\theta d\theta$$

$$+ \frac{\sin^2\sigma P_n''(\cos\sigma)}{(n+2)(n+1)n(n-1)} \qquad (19)$$

$$\times \int_0^{\pi} \frac{k^2 \sin^5\theta}{(\lambda + ik\cos\theta)^2} P_n''(\cos\theta) d\theta \bigg\}.$$

We are interested in the form of expression (19) in the limit $\lambda \longrightarrow 0$. Integrating by parts yields

$$I_n = \int_0^{\pi} \frac{\sin\theta P_n(\cos\theta)}{(\lambda + ik\cos\theta)^2} d\theta$$

$$= \frac{i}{k}\left(\frac{1}{\lambda + ik} - \frac{1}{\lambda - ik} \right) - \frac{i}{k}\int_{-1}^{1} \frac{P_n'(t)}{\lambda + ikt} dt.$$

In the limit $\lambda = 0$, we obtain

$$I_n = \frac{2}{k^2} - \frac{1}{k^2}\int_{-1}^{1}\frac{P'_n(t)}{t}dt.$$

Using the recurrence formula presented in [10], we find

$$\int_{0}^{1}\frac{P'_n(t)}{t} = 1 - (n+1)\int_{0}^{1}\frac{P_{n+1}(t)}{t}dt = 1 - \frac{1}{P_n(0)}$$

(the number n is always assumed to be even). Consequently, we arrive at the integral

$$I_n = \frac{2}{k^2 P_n(0)} \quad (\lambda = 0).$$

Other similar integrals on the right-hand side of expression (19) reduce to this one and, hence, in the limit $\lambda \longrightarrow 0$, are equal to

$$\int_{0}^{\pi}\frac{\sin^3\theta P_n(\cos\theta)}{(\lambda + ik\cos\theta)^2}d\theta = \frac{4}{k^2} \ (n=0); \quad \frac{2}{k^2 P_n(0)} \ (n\geq 2).$$

We then get

$$K_n = \int_{0}^{\pi}\frac{\sin^3\theta}{(\lambda + ik\cos\theta)^2}P''_n(\cos\theta)d\theta$$

$$= \int_{-1}^{1}\frac{(1-t^2)[2tP'_n(t) - n(n+1)P_n(t)]}{(\lambda + ikt)^2}dt,$$

and thus are left with the problem of calculating the auxiliary integral. This is done by integrating by parts,

$$\int_{-1}^{1}\frac{t(1-t^2)P'_n(t)}{(\lambda + ikt)^2}dt = \frac{2}{k^2 P_n(0)},$$

so we obtain

$$K_n = -\frac{2(n+2)(n-1)}{k^2 P_n(0)} \quad (n\geq 2).$$

As a result, we have

$$M = -2\pi\sum_{n=2}^{\infty}\frac{c_n}{P_n(0)}\left[P_n(\cos\theta) + \frac{\sin^2\sigma P''_n(\cos\sigma)}{n(n+1)}\right]$$

$$= -4\pi\sum_{n=2}^{\infty}\frac{c_n}{P_n(0)}(\cos\sigma)P'_n(\cos\sigma). \tag{20}$$

A sufficient condition for the function M to be negative is derived by integrating the right-hand side of

expression (20) with a weighting factor of $\sin^2\sigma$:

$$\int_{0}^{\pi}\sin^3\sigma\cos\sigma P'_n(\cos\sigma)d\sigma = \frac{4}{5} \ (n=2)0; \ (n\geq 4)$$

so that

$$\int_{0}^{\pi}\sin^3\sigma\cos\sigma M(\sigma)d\sigma = \frac{32\pi}{5}c_2. \tag{21}$$

By virtue of formula (21), the function $M(\sigma)$ for $c_2 < 0$ cannot be positive everywhere. There always exists a σ value such that $M(\sigma) < 0$; this implies instability. Hence, the sufficient instability condition is the negativeness of the corresponding coefficient in the expansion of the phase density in spherical functions:

$$\int_{0}^{\pi}FP_2(\cos\theta)\sin\theta d\theta < 0. \tag{22}$$

Let us now turn to the expression $AC - B^2$. It is readily verified that the following equality (which is a result of the above transformation of rotation of the coordinate system) is satisfied:

$$AC - B^2 = A_1C_1 - B_1^2, \tag{23}$$

where the expressions

$$A_1 = A\cos^2\sigma - 2B\sin\sigma\cos\sigma + C\sin^2\sigma,$$

$$B_1 = (A-C)\sin\sigma\cos\sigma + B(\cos^2\sigma - \sin^2\sigma), \tag{24}$$

$$C_1 = A\sin^2\sigma + 2B\sin\sigma\cos\sigma + C\cos^2\sigma;$$

derive from the explicit expressions for A, B, and C for the above class of velocity distributions on a unit sphere. In particular, we have

$$A_1 = \iint\left[1 - \frac{k^2u_1^2}{(\lambda + ikw_1)^2}\right]F\sin\theta d\theta d\mu.$$

Substituting the expressions for u_1, w_1, and F into this relationship, we arrive at an expression that coincides with expression (18) to within the replacement of $\sin^2\mu_1$ by $\cos^2\mu_1$. We then simply have to change the sign of the last term on the right-hand side of expression (19) to obtain

$$A_1 = 2\pi\sum_{n=2}^{\infty}\frac{c_n}{P_n(0)}\left[\frac{\sin^2\sigma P''_n(\cos\sigma)}{n(n+1)} - P_n(\cos\sigma)\right]$$

$$= 4\pi\sum_{n=2}^{\infty}\frac{c_n}{(2n+1)P_n(0)}\left(\frac{n+1}{n}P'_{n-1}(\cos\sigma)\right) \tag{25}$$

$$-\frac{n}{n+1}P'_{n+1}(\cos\sigma)\Bigg)\ (\lambda = 0).$$

The quantity C_1 is described by the expression

$$C_1 = \lambda^2 \iint \frac{F\sin\theta}{(\lambda + ikw_1)^2}d\theta d\mu$$

$$= \frac{4\pi\lambda^2}{k^2}\sum_{n=0}^{\infty}\frac{c_n P_n(\cos\sigma)}{P_n(0)} \tag{26}$$

which is valid to within small terms on the order of λ^3 and higher. Finally, we get

$$B_1 = -ik\lambda\iint\frac{u_1 F\sin\theta}{(\lambda + ikw_1)^2}d\theta d\mu$$

$$= -2\pi ik\lambda\sum_{n=2}^{\infty}\frac{c_n}{n(n+1)}\sin\sigma P'_n(\cos\sigma)$$

$$\times\int\frac{\sin^3\theta P'_n(\cos\theta)}{(\lambda + ik\cos\theta)^2}d\theta.$$

The auxiliary integral can be taken by parts at $\tau = i\lambda/k$:

$$\int_{-1}^{1}\frac{(1 - t^2)P'_n(t)}{(t - \tau)^2}dt = -n(n+1)\pi i P_n(0)\ (\lambda = 0).$$

Hence, we have

$$B_1 = 2\pi^2\frac{\lambda}{k}\sum_{n=2}^{\infty}c_n P_n(0)\sin\sigma P'_n(\cos\sigma). \tag{27}$$

Expressions (25)–(27) are sufficient to determine the leading-order term of $A_1 C_1 - B_1^2$, which is proportional to λ^2. Note that the above expression for B_1 is not needed for further analysis and is given only to furnish a more complete picture.

Let us estimate the mean value of the product $A_1 C_1$ with a weighting function of $\sin^3\sigma$. Based on formulas (25) and (26) and on the recurrence formula $(2n + 1)P_n(x) = P'_{n+1}(x) - P'_{n-1}(x)$, we find

$$\int_{0}^{\pi}\sin^3\sigma A_1 C_1 d\sigma$$

$$= \int_{-1}^{1}(1 - x^2)\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\alpha_n\alpha_m(P'_{m+1}(x) - P'_{m-1}(x))$$

$$\times\left(\frac{n+1}{n}P'_{n-1}(x) - \frac{n}{n-1}P'_{n+1}(x)\right)dx,$$

where

$$\alpha_n = \frac{c_n}{(2n+1)P_n(0)}.$$

The orthogonality property of the Legendre associated polynomials yields

$$\int_{0}^{\pi}\sin^3\sigma A_1 C_1 d\sigma = -\sum_{n=2}^{\infty}\frac{2(2n+1)(2n^2 + 2n - 3)}{(2n+3)(2n-1)}\alpha_n^2$$
$$\tag{28}$$
$$+ \sum_{n=2}^{\infty}\frac{2n(n+2)}{2n+5}\alpha_n\alpha_{n+2} + \sum_{n=0}^{\infty}\frac{2(n+1)(n+3)}{2n+3}\alpha_n\alpha_{n+2}.$$

The sought-for estimate is obtained by estimating the mixed products $\alpha_n\alpha_{n+2}$ with $n \geq 2$ by means of the obvious inequality $2\alpha_n\alpha_{n+2} \leq \alpha_n^2 + \alpha_{n+2}^2$. This gives

$$\int_{0}^{\pi}\sin^3\sigma A_1 C_1 d\sigma$$

$$\leq -4\sum_{n=4}^{\infty}\frac{n^2\left(4n^2 + 8n - 9 - \frac{16}{n}\right)}{(2n-1)(2n+1)(2n+3)(2n+5)}\alpha_n^2 \tag{29}$$

$$-\frac{79}{63}\alpha_2^2 + 2\alpha_0\alpha_2.$$

We always have $c_0 > 0$ and $\alpha_0 > 0$. If $\alpha_2 < 0$ and $c_2 > 0$, then the right-hand side of inequality (29) is negative and the above examination shows that, at a certain σ value, the inequality $A_1 C_1 < 0$, and especially the inequality $A_1 C_1 - B_1^2 < 0$, should hold. This indicates that the sufficient instability condition is given by the inequality opposite to inequality (22). Accordingly, the necessary stability condition is, on the whole, given by the equality

$$\int_{0}^{\pi}FP_2(\cos\theta)\sin\theta d\theta = 0. \tag{30}$$

GENERAL REMARKS

Up to this point, we have used an example of a velocity distribution on a unit sphere. Now, the task is to establish the extent to which the results obtained apply to general velocity distributions.

For $\lambda = 0$, the contribution of a particle with the velocity vector $(\xi u, \xi v, \xi w)$ to the left-hand side of Eq. (14) is the same as that of a particle with the velocity vector (u, v, w) because the limiting transitions $\lambda \longrightarrow 0$ and $\lambda/\xi \longrightarrow 0$ are equivalent. We thus can

immediately write the generalization of instability condition (22):

$$\int\limits_0^\infty \int\limits_0^\pi f(V, \theta) P_2(\cos\theta) V^2 \sin\theta \, d\theta \, dV < 0. \tag{31}$$

The situation with the expansions of A, B, and C in powers of λ is more complicated. The replacement of (u, v, w) by $(\xi u, \xi v, \xi w)$ generates the replacement $A(\lambda) \longrightarrow A(\lambda/\xi)$ (and the same replacements for B and C); hence, the series expansion

$$A = A^{(0)} + \lambda A^{(1)} + \lambda^2 A^{(2)} + \dots$$

becomes

$$A^{(0)} + \lambda \frac{A^{(1)}}{\xi} + \lambda^2 \frac{A^{(2)}}{\xi^2} + \dots.$$

Consequently, in formulas (25)–(27), the expressions for the coefficients c_n of the corresponding powers of λ will be different. In accordance with what was said above, the coefficients in formulas (20) and (25) are given by the expressions

$$c_n = \frac{2n+1}{4\pi} \int\limits_0^\infty \int\limits_0^\pi f(V, \theta) P_n(\cos\theta) V^2 \sin\theta \, d\theta \, dV. \tag{32}$$

As for the formula for C_1, it contains the factor λ^2; therefore, the coefficients c_n have some other form:

$$c_n = \frac{2n+1}{4\pi} \int\limits_0^\infty \int\limits_0^\pi f(V, \theta) P_n(\cos\theta) \sin\theta \, d\theta \, dV. \tag{33}$$

Finally, formula (27) for B_1 contains the factor λ, so the coefficients on the right-hand side of this formula read

$$\hat{c}_n = \frac{2n+1}{4\pi} \int\limits_0^\infty \int\limits_0^\pi f(V, \theta) P_n(\cos\theta) V \sin\theta \, d\theta \, dV. \tag{34}$$

For some types of velocity distributions, the instability condition supplementary to condition (31) can be constructed in a relatively simple way.

(1) Distributions in which the angular variables are separated from V,

$$f(V, \theta) = f_1(V) f_2(\theta). \tag{35}$$

Substituting distribution (35) into expressions (32)–(34) yields the integrals

$$\int\limits_0^\pi f_2(\theta) P_n(\cos\theta) \sin\theta \, d\theta; \quad n = 0, 1, 2, \dots \tag{36}$$

with different coefficients, which, for all three expressions, are independent of n. Consequently, the expressions for A_1 and C_1 will differ from those obtained for a

velocity distribution on a unit sphere only in these constant coefficients; it is obvious that they will not change the final results. In complete analogy with condition (30), a necessary condition for stability is obtained by equating integrals (36) with $n = 2$ to zero. In terms of the initial distribution function f, this indicates that

$$\int\limits_0^\infty \int\limits_0^\pi f(V, \theta) P_2(\cos\theta) \sin\theta \, d\theta = 0. \tag{37}$$

(2) Distributions without higher harmonics,

$$f(V, \theta) = \varphi_0(V) + P_2(\cos\theta) \varphi_2(V).$$

For such distributions, we naturally have $c_n = \bar{c}_n = \hat{c}_n$ for $n \geq 4$ and there remain only the first terms in expressions (25) and (26),

$$A_1 = \frac{4\pi c_2}{5} \cos 2\sigma, \quad C_1 = \frac{4\pi\lambda^2}{k^2} [\bar{c}_0 - 2\bar{c}_2 P_2(\cos\sigma)].$$

For certain values of σ, the product $A_1 C_1$ can take on negative values, except when (a) $c_2 = 0$ and (b) $\bar{c}_2 = 2\bar{c}_0$, $c_2 < 0$.

In case (b), instability condition (31) is certainly satisfied because its left-hand side coincides with c_2 to within a positive factor. Stability is possible only in case (a), namely, for $c_2 = 0$, which coincides with condition (37). Moreover, even when $c_2 = 0$, the quantity $A_1 C_1 - B_1^2$ will be negative if $\hat{c}_2 \neq 0$. Hence, for distributions of type 2, the necessary stability conditions have the form $c_2 = \hat{c}_2 = 0$ or, equivalently,

$$\int\limits_0^\infty V^2 \varphi_2(V) dV = \int\limits_0^\infty V \varphi_2(V) dV = 0. \tag{38}$$

(3) Nearly spherical distributions

$$f(V, \theta) = \varphi_0(V) + \kappa \psi(V, \theta)$$

with $\kappa \longrightarrow 0$. All the coefficients c_n, $\bar{c}_n$, and $\hat{c}_n$ with $n \geq 2$ are proportional to κ. For $\kappa \longrightarrow 0$, the leading-order term in the expression for $A_1 C_1 - B_1^2$ is

$$\frac{4\pi c_2}{5} \cos 2\sigma \frac{4\pi\lambda^2}{k^2} c_0.$$

This term, which is first-order in κ, is of variable sign for $c_2 \neq 0$ and is identically zero for $c_2 = 0$. By analogy with distributions of the previous type, we conclude that equalities (38) constitute necessary conditions for stability in a certain range $0 < \kappa < \kappa_0$.

Thus, the requirement that there be no electromagnetic instability severely restricts the class of admissible velocity distributions. It has been shown that, under certain additional assumptions, the stability condition

should have the form of an equality (a conclusion somewhat unusual for the relevant problems); specifically, the second zonal harmonic defined in one way or another should vanish. In this respect, the electromagnetic instability is stronger than the electrostatic one, because the latter manifests itself when the velocity distribution deviates substantially from being spherically symmetric: it may cause a separation of two beams, and so on.

Notably, from energy considerations, it follows in advance that a spherically symmetric distribution (such that $\partial f/\partial V < 0$) is completely stable: according to the Liouville theorem, any perturbation in this case increases the kinetic energy of the system [11].

Similar results were obtained in a number of earlier papers. The result that most closely corresponds to those presented above was obtained in [12] for an ellipsoidal velocity distribution. Although the dispersion relations, as well as the conclusions about the instability, are the same as in the present paper, the model of [12] is more specific in character because it refers to a particular spheroidal distribution. In [13], a somewhat different model was considered, namely, a superposition of two Gaussian distributions. That model, too, showed the onset of instability for a sufficiently strong total anisotropy of the particle velocities. Those two papers concentrated on technical applications and paid little attention to astrophysical aspects of the plasma behavior. This is why, in my recent work [14], a study was made of the ultrarelativistic limit (a case opposite to that considered here) on the basis of a model of a triaxial ellipsoidal velocity distribution. Such a distribution was also found to be unstable, except, of course, for the particular case of a spherical distribution.

The linear approximation used above to describe the electromagnetic instability is valid only for sufficiently long waves. The critical wavelength can be estimated from dimensionality considerations, as is quite clear from simple case (1), in which it was in fact assumed that

$$k \ll \frac{\omega_P}{c}, \tag{39}$$

where the Langmuir frequency ω_p is defined by formula (2). Thus, for electron densities of about $n_e = 0.1$ cm^{-3}, which are more or less typical of interstellar media [15], formula (39) gives a critical wavelength of $\lambda^* \sim 2\pi/k \sim 100$ km. So, any of the characteristic spatial scales in interstellar space is certainly far longer than λ^* and the approximation in question is valid. As for the instability growth rate, it is on the order of the time required for an average particle to traverse the distance equal to the wavelength. As a result, the instability develops in certain astrophysical situations, e.g., at distances of 10^{-4} pc or less and on fairly short time scales (about half a month).

Note that, here, the particle velocities were always assumed to be nonrelativistic. In multicomponent systems, in which the particle velocities can differ by orders of magnitude, there also may be some additional phenomena, one reason being that charge transfer by cosmic rays may sometimes be significant [16].

Finally, it should again be emphasized that taking into account the electromagnetic instability severely restricts the choice of models for describing anisotropic velocity distributions in interstellar space.

REFERENCES

1. S. A. Kaplan and S. B. Pikelner, *Interstellar Medium* (Fizmatgiz, Moscow, 1963; Harvard Univ. Press, Cambridge, 1982).

2. L. Spitzer, Jr., Annu. Rev. Astron. Astrophys. **28**, 71 (1990).

3. S. Heinz and R. Sunyaev, Astron. Astrophys. **390**, 751 (2002).

4. *Plasma Electrodynamics*, Ed. by A. I. Akhiezer (Nauka, Moscow, 1974; Pergamon, Oxford, 1975).

5. A. B. Mikhailovskii, *Theory of Plasma Instabilities* (Atomizdat, Moscow, 1971; Consultants Bureau, New York, 1974), Vol. 1.

6. E. S. Weibel, Phys. Rev. Lett. **2**, 83 (1959).

7. V. M. Lyutyi, in *Astrophysics and Space Physics Reviews (Soviet Scientific Reviews, Section E: Astrophysics and Space Physics Reviews)*, Ed. by R. A. Syunyaev (Nauka, Moscow, 1982; Harwood Academic, 1983), pp. 66–87.

8. É. A. Dibaĭ, in *Active Nuclei and Stellar Cosmogony*, Ed. by D. A. Martynov (MGU, Moscow, 1987), pp. 6–33 [in Russian].

9. A. A. Ivanov, *Physics of Highly Nonequilibrium Plasmas* (Atomizdat, Moscow, 1977) [in Russian].

10. H. Bateman and A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1968; Nauka, Moscow, 1974).

11. R. Devidson, in *Basic Plasma Physics: A Handbook of Plasma Physics*, Ed. by A. A. Galeev and R. N. Sudan (North-Holland, Amsterdam, 1983; Énergoatomizdat, Moscow, 1983), Vol. 1, pp. 443–502.

12. R. N. Sudan, Phys. Fluids **8**, 153 (1965).

13. A. B. Mikhailovskii, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Atomizdat, Moscow, 1972; Consultants Bureau, New York, 1978), Vol. 6, pp. 70–138.

14. A. S. Baranov, Fiz. Plazmy **29**, 956 (2003) [Plasma Phys. Rep. **29**, 890 (2003)].

15. D. Ya. Martynov, *Course of General Astrophysics* (Nauka, Moscow, 1979) [in Russian].

16. V. A. Antonov and A. S. Baranov, Astron. Zh. **79**, 387 (2002) [Astron. Rep. **46**, 347 (2002)].

Translated by O. Khadin

GAS DISCHARGES, PLASMA

Pollution Control and Depollution Processes by Plasma Techniques[1]

J. Amouroux*, D. Morvan*, S. Cavadias*, Ph. Adam**, M. F. Gonnord*, K. Coulibaly*, A. Vincent*, S. Morel*, F. Daou*, S. Ognier*, P. Rousseau*, and L. Martin*

* Laboratoire de Génie des Procédés Plasmas et Traitement de Surface, ENSCP, Paris, 11 rue P&M. Curie, 75231 Paris Cedex 05, France

** DGA-CEB BP3, 91710 Vert le petit

Received July 21, 2004

Abstract—The European Union is creating strict standards for air and water pollution and waste treatment and implementing aggressive regulations. Compliance with these regulations is impossible without the development of new depollution processes involving plasma or laser technology. Time is one of the major problems in monitoring pollutants with the use of the time-resolution laser-induced breakdown spectroscopy technique, which can perform online analysis without sampling with a high level sensitivity for all the species of the Periodic Table. Plasma-enhanced desorption from fly ashes or polluted soils associated with a mass spectrometer or an optical emission spectrometer allows the monitoring of volatile organic compounds (VOCs) and organochlorine species in a few seconds. In the last DBD point, we present the treatment of VOCs in air by the plasma technique in order to destroy or trap PAH molecules. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

Pollution is the contamination of the environment as a result of human activities (combustion, mineral extraction, chemical plants, energy production, etc.).

Each year in the European Union, we throw away 1.3 billion tons of waste, some 40 million tons of it hazardous. Add to this total a further 700 million tons of agricultural waste, and it is clear that treating and disposing of all this material, without harming the environment becomes a major headache. Between 1990 and 1995, the amount of waste generated in Europe increased by 10%, according to the Organization for Economic Cooperation and Development (OECD). Most of what is thrown away is either burned in incinerators or dumped into landfill sites (67%). By 2020, the OECD estimates that it could be generating 45% more waste than we did in 1995.

The first step for the reduction of the pollutant emission and waste is their identification and characterization by fast analytical techniques. European Union is acting on air, water pollution and waste treatment, implementing aggressive regulations. The compliance with these regulations is impossible without the development of new depollution techniques.

Therefore the second step must be precisely the development of such techniques based on plasma science.

For the first step in our laboratory we have developed two techniques for the online analysis of gases:

(1) (TRELIBS) Time-resolution laser-induced breakdown spectroscopy allows identification by atomic emission of the elements resulting from high power laser pulse on gases, liquids, powders, or surfaces of materials.

(2) (MS) Mass spectrometry on gases desorbed from surfaces or powders, like fly ashes produced from municipal solid wastes.

For a second step we have developed thermal plasma techniques for the vitrification of fly ashes produced from the municipal solid wastes and DBD discharges for the treatment of (VOCs).

In Section 2 is present the European legislation for waste treatment. Section 3 is devoted to different types of online pollutant analysis (TRELIBS and nonequilibrium plasma desorption coupled with mass spectrometry). The industrial treatment of waste by plasma techniques and DBD processes for VOC elimination from gases are presented in Section 4.

1. EUROPEAN LEGISLATION FOR WASTE TREATMENT

The European Union's Sixth Environment Action Program identifies waste prevention and management as one of its top four priorities. It wants to reduce the quantity of waste going to "final disposal" by 20% from 2000 to 2010 and by 50% by 2050, with special emphasis on cutting hazardous waste.

---

[1] This article was submitted by the authors in English.

Table 1. Air emission limit values

Pollutants	Air emission limit values
Total dust	10 mg/m^3 (#)
Gaseous and vaporous organic substances, expressed as total organic carbon	10 mg/m^3 (#)
Hydrogen chloride (HCl)	10 mg/m^3 (#)
Hydrogen fluoride (HF)	1 mg/m^3 (#)
Sulphur dioxide (SO$_2$)	50 mg/m^3 (#)
Nitrogen monoxide (NO) and nitrogen dioxide (NO$_2$) for existing incineration plants with a nominal capacity exceeding 6 t/h or new incineration plants	200 mg/m^3 (*) (#)
Nitrogen monoxide (NO) and nitrogen dioxide (NO$_2$) for existing incineration plants with a nominal capacity of 6 t/h or less	400 mg/m^3 (*) (#)
Cd + Tl	Total 0.05 mg/m^3 (+)
Hg	0.05 mg/m^3 (+)
Sb + As + Pb + Cr + Co + Cu + Mn + Ni + V	Total 0.5 mg/m^3 (+)
Dioxins and furans	0.1 ng/m^3 ($)

Note: (#) Daily average values; (+) average over a period of 30 min–8 h; ($) average over a period of 6–8 h; and (*) until January 1, 2007, the emission limit value for NO$_x$ (does not apply to plants, only incinerating hazardous waste).

The European Union's approach to waste management is based on three principles:

1.1. Waste prevention: This is a key factor in any waste management strategy. If we can reduce the amount of waste generated in the first place and reduce its hazardousness by reducing the presence of dangerous substances in products, then disposing of it will automatically become simpler. Waste prevention is closely linked with improving manufacturing methods and influencing consumers to demand greener products and less packaging.

1.2. Recycling and reuse: If waste cannot be prevented, as many of the materials as possible should be recovered, preferably by recycling. The European Commission has defined several specific "waste streams" for priority attention, the aim being to reduce their overall environmental impact. This includes packaging waste, used-up vehicles, batteries, and electrical and electronic waste. EU directives now require member states to introduce legislation on waste collection, reuse, recycling, and disposal of these waste streams. Several EU countries are already managing to recycle over 50% of packaging waste.

1.3. Improving final disposal and monitoring: Where possible, waste that cannot be recycled or reused should be safely incinerated, with landfills only used as a last resort. Both these methods need close monitoring because of their potential for causing severe environmental damage. The European Union has recently approved a directive that sets strict guidelines for landfill management. It bans certain types of waste, such as used tires, and sets targets for reducing quantities of biodegradable rubbish. Another recent directive lays down tough limits on emission levels from incinerators. The Union also wants to reduce emissions of dioxins and acid gases, such as nitrogen oxides (NO$_x$), sulfur dioxides (SO$_2$), and hydrogen chlorides (HCl), which can be harmful to human health.
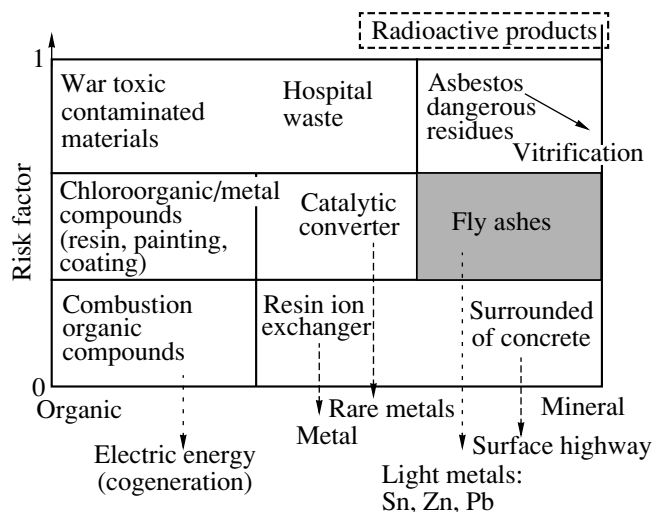
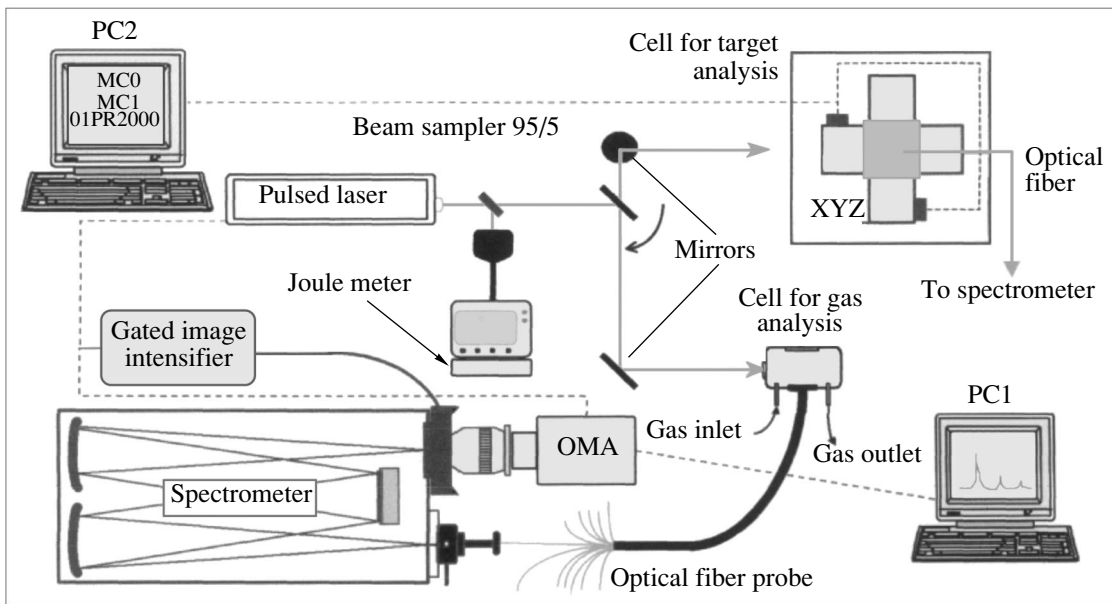**Fig. 1.** Waste products of human activities.
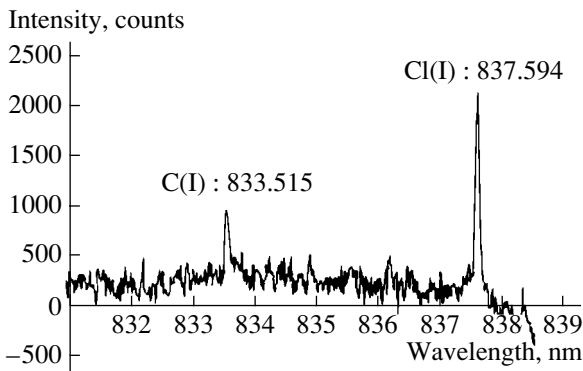
Fig. 2. Experimental setup of TRELIBS analysis.

Fig. 3. Spectral resolution and choice of atomic emission lines: chlorine and carbon examples.
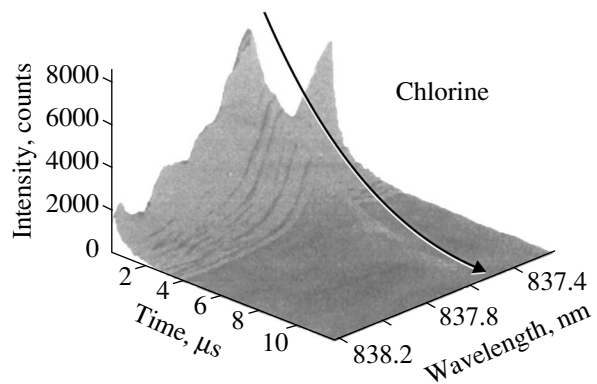

Fig. 4. Time evolution of emission intensities.

A recent directive (Directive 2000/76/EC) on the incineration of waste that was published on December 28, 2000, lays down tough limits on emission levels from incinerators. Although the volume of waste incineration is predicted to increase across the European Union in the near future, this directive will lead to significant reductions in emissions of several key pollutants. Considerable reductions will be achieved for acidic gases such as nitrogen oxides, sulfur dioxide, and hydrogen chloride as well as for heavy metals: Cd from 16 t to 1 t in 2005 and dioxins and furans from an annual 2400 g in 1995 to only 10 g.

This directive states the conditions of control and monitoring of emissions and defines the periodicity of the measurements.

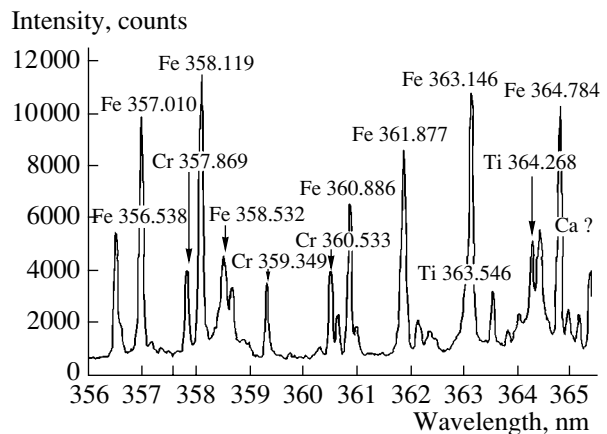The Table 1 defines the air-emission limit values.

Fig. 5. Identification of atomic elements from fly ash. The Ni atomic line at 361.939 nm is overlapped by the Fe line at 361.876 nm.
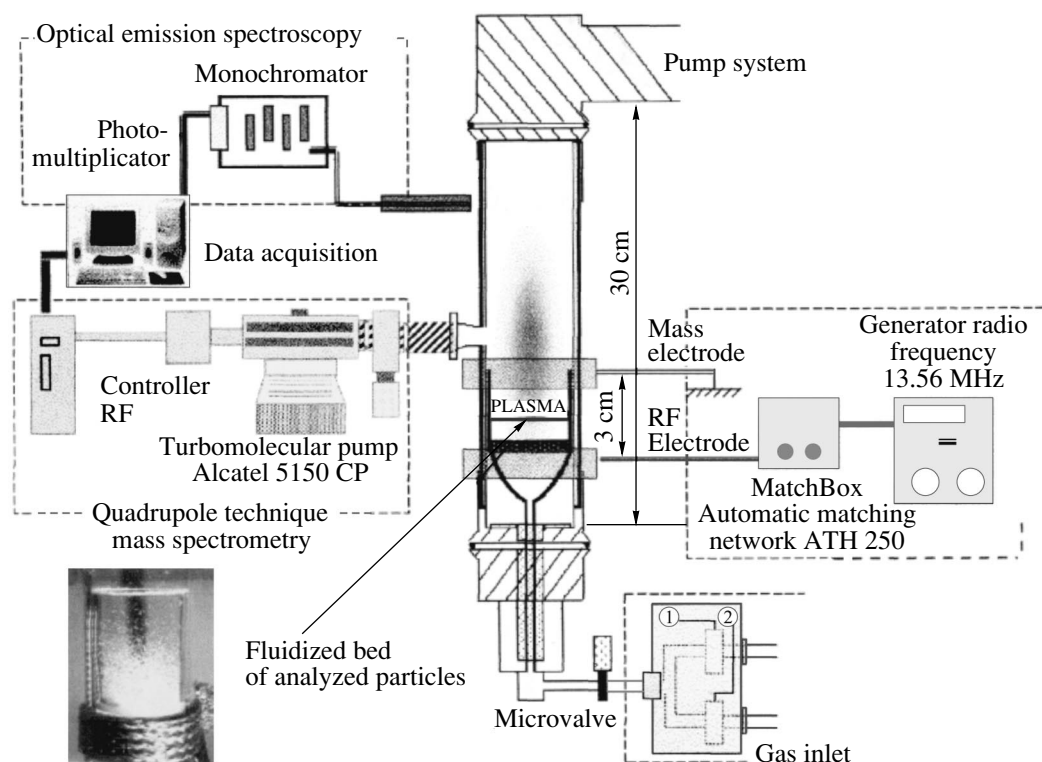
Fig. 6. Experimental device for qualification of desorbed species from ash. Radio frequency low pressure plasma process (pressure 5–10 Pa).

Fig. 7. Mass spectrum of low pressure argon plasma treatment of fly ash. (Mass of fly ash 0.7 g; power 100 W; Ar flow rate: 284 sccm, pressure 5 Pa.)

2. ANALYSIS OF POLLUTANT EMISSION

Our purpose is to apply online diagnostics for monitoring of emission gases from incineration of waste materials. The requirements are rapid responses and high levels of sensitivity. TRELIBS analysis and plasma-enhanced desorption of gases coupled with MS analysis are new methods that meet these requirements.

2.1. TRELIBS Online Analysis [1–4]

The goal of this technique is to identify, using atomic emission, the elements in any kind of material, gas, powder, or material surface.

For this purpose, we use a YAG laser ($\lambda = 1064$ nm) with an energy beam between 25 to 160 mJ per pulse with a frequency of 50 Hz (Fig. 2). At the focus of the lens, the laser beam pulse produces plasma with a duration of 10 ms and a size of 6 mm in length and 3 mm in

Low pressure argon (284 sccm)–H_2 (24 sccm) plasma

Identification:

Peak (m/e)	Species		Peak (m/e)	Species	
2	H_2^+	↗	40	Ar^+	
16	O^+, CH_4^+	↘	50, 51, 52		
18	H_2O^+	↘			
20	Ar^+		70	Cl_2^+	↘
26	$C_2H_2^+$	→	78	$C_6H_6^+$	→
28	N_2^+, CO^+		91, 92	$C_7H_8^+$	→
30	$C_2H_6^+$	↗			
32	O_2^+	↗	128	$C_6H_4ClOH^+$	↘
35, 37	Cl^+	↗	146	$C_6H_4Cl_2^+$	↘
36	HCl^+	↗	180	$C_6H_3Cl_3^+$	↘

Intensity, ×10⁻⁸

Cl^+(35)
HCl^+(36)
Cl^+(37)   Heavy molecules
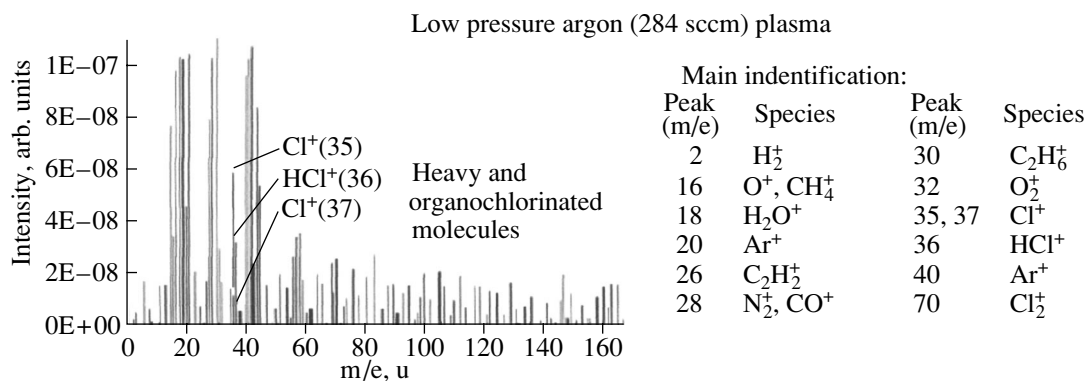78: $C_6H_6^+$; 92: $C_7H_8^+$
H/H_2
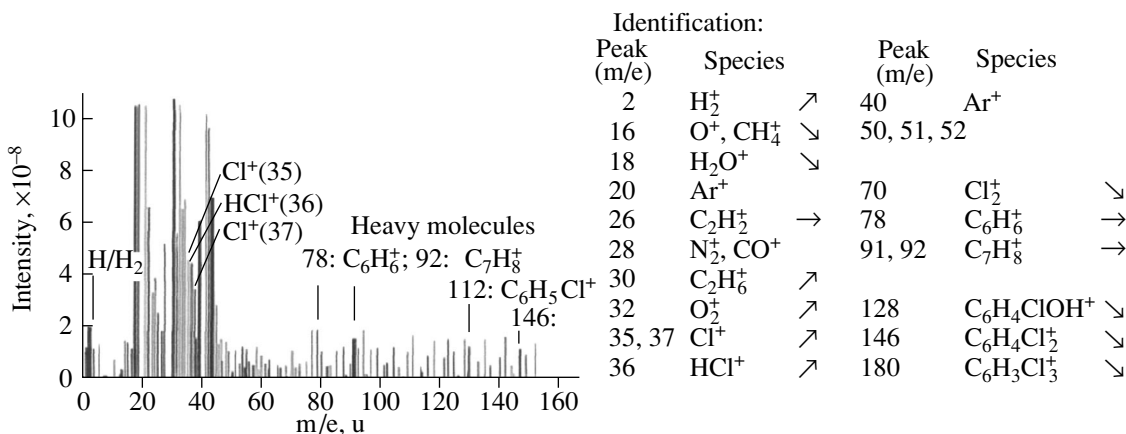112: $C_6H_5Cl^+$
146:

m/e, u

Fig. 8. Mass spectrum of low pressure argon hydrogen plasma treatment of fly ash. (Mass of fly ash 0.7 g; power 100 W; flow rates: Ar 284 sccm, H_2 = 24 sccm, pressure 5 Pa.)

height. In this plasma ellipsoid, a temperature increase of 20 000 K and, after 500 ns of bremsstrahlung, the atomic lines for a period of a few microseconds give us the possibility to, using an optical fiber, a spectrometer, and an OMA, measure the light intensity for each wavelength while a software program gives us the nature of the elements and their concentration.

The detection limit, the time evolution, and the stoichiometric composition of the molecules have been found for fluorine, chlorine, phosphorus, and sulfur (Figs. 3–5).

The measurement lasts for about 10 μs in a volume of a few mm³ of gas, and the technique permits a sensitivity of 10^{-6} g/g. With this technique, we can also determine the chemical composition of the melted fly ashes by identifying elements in the bulk (Si, Ca, Mg, or Al) or heavy metals trapped in the oxides of the matrix (Ti, Cr, Ni, Mn, Zn, Fe, or Pb).

The main advantage of this method is that, without any long and complex chemical preparation of the samples, it can provide direct analysis of solids. The volume needed for the analysis is a few mm³ of gas or a surface of 100 μm diameter for liquids or solids. Only optical access to the sample is required, and the risk of sample contamination is negligible.

This method allows detection and identification of the main heteroatoms (fluorine, chlorine, sulfur, etc.) and atomic elements such as Si, Ca, Al, Cr, Mg, Pb, or Fe in fly ash at very low concentrations (10^{-6} g/g). Another advantage of this method is the use of an optical fiber to carry and treat the signal far from the location of the measurements.

2.2. Nonequilibrium Plasma Desorption Coupled with Mass Spectrometry for Fly Ash Analysis of a Trapped VOC [5–8]

The organic compounds analysis of powders trapped in fly ash generally need desorption methods such as liquid/solider extraction by Sohxlet and GC/MS analysis. This technique requires a long analysis time, which is why we have developed a low pressure plasma desorption process for polluted powders with online mass spectrometry analysis of the desorbed organic compounds (VOCs and organochlorine species). The time from sampling of powders to MS analysis of the organocompounds trapped in the powders is only few seconds.

HCl + Cl content (×10⁻⁹)

×18   ×6   69.5   ×2.5
3.9   12.1   27

Under argon fluidization   Argon plasma   Ar + H_2 plasma   Ar + H_2 plasma + CaO
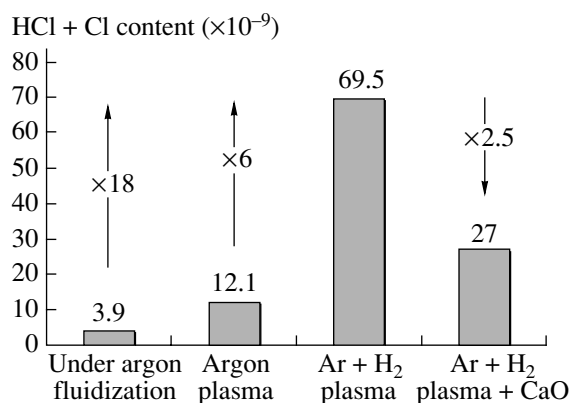
Fig. 9. Chlorinated mass balance of the desorbed species (Cl + HCl) for different treatments. Experimental conditions: mass of fly ash 2 g; Ar 300 sccm, H_2 = 24 sccm, O_2 = 10 sccm; P_w = 100 W; P = 1 mbar.

Table 2. Inventory of chemical species qualified by Soxhlet extraction and GC/MS analysis

Alkanes:	**Alkyl PAHs**
C_{10} to C_{31}	
Polychlorobenzenes (2, 3, 4, 5 Cl)	Methylnaphtalene
Dichlorotetracycloheptane $C_7H_6Cl_2$	Ethylnaphtalene
Ethylmethylbenzene	Di/tri-methylnaphtalene
Benzaldehyde (C_7H_6O)	Methylphenanthrene
Biphenyl ($C_{12}H_{10}$)	**Alkyl phenols**
Dimethylbiphenyl ($C_{14}H_{14}$)	**Phosphoric compounds**
Sulfur S_8 (256)	Triethylene phosphoric acid $C_6H_{15}O_4P$
Phenol $C_{15}H_{24}O$, $C_{16}H_{26}O$	**Bromic compounds**
PAHs	Terbutylbromomethylphenol ($C_{15}H_{23}OBr$)
Naphtalene	**Sulphur compounds**
Acenaphtylene	Ethylphenylthiopentanone ($C_3H_{18}OS$)
Acenaphtene	Acethylthienylbutanone ($C_{10}H_{12}O_2S$)
Anthracene/phenanthrene	Sulfonylbenzene ($C_{12}H_{10}O_2S$)
Fluoranthene	
Pyrene Chrysene/Benzo(a)anthracene	**Aromatic compounds**
Benzo(b)fluoranthene	Methylphenol
Benzo(k)fluoranthene	Ethanolphenoxy ($C_8H_{10}O_2$)
Benzo(a)pyrene	Methanonebiphenyl ($C_{13}H_{10}O$)
Benzo(ghi)perylene/Indeno(1,2,3-cd) pyrene	

This experimental technique is constituted by a fluidized bed in a quartz tube (0.04 m diameter and 0.3 m length) with a gas injection (argon/H_2/O_2) at the bottom and a connection with the vacuum system at the top.

An RF generator at 13.56 MHz produces a nonequilibrium plasma in the fluidized bed (Fig. 6) of the ash powder. The OES emission allows the measurement of the vibrational and rotational temperatures of the
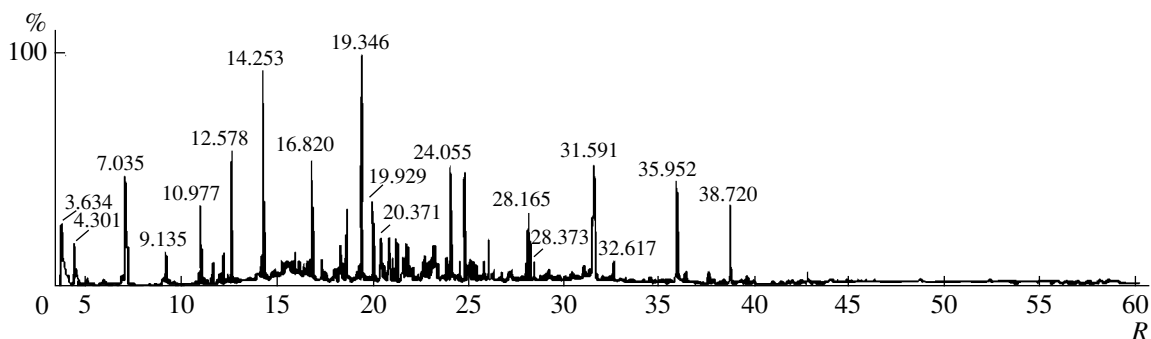
Fig. 10. Chromatogram of soluble organic fraction of industrial fly ash.
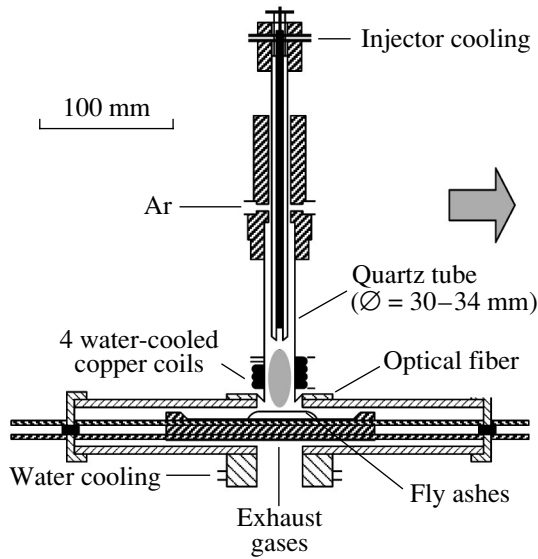
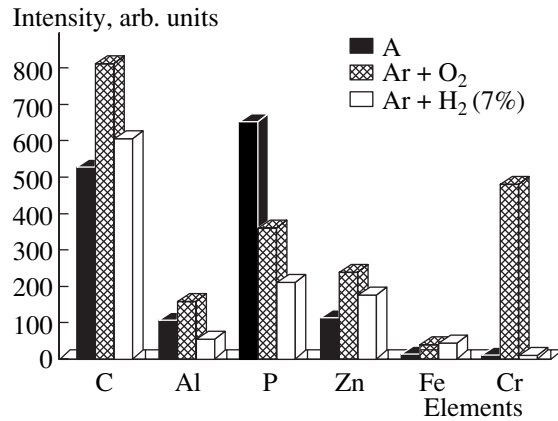**Fig. 11.** Experimental setup for fly ash vitrification by thermal plasma.

Fig. 12. Evaporation of each element obtain by integration of the surface of each atomic emission line.
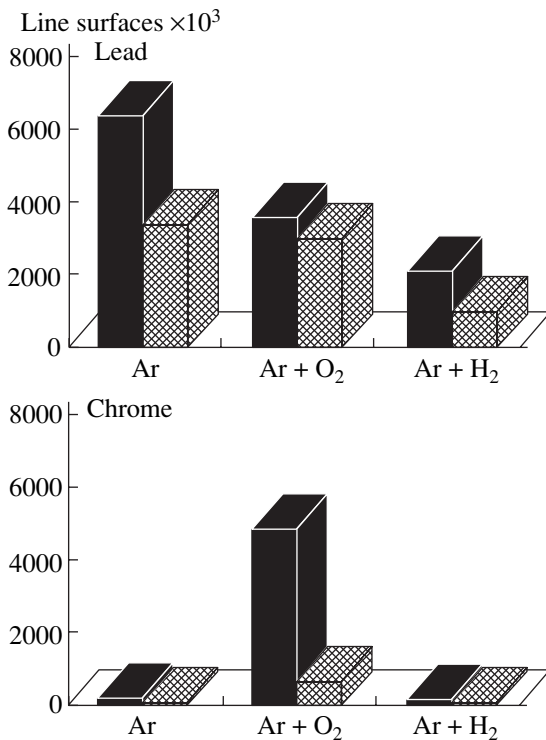
Fig. 13. Competition between the desorption of metallic oxides and metallic chlorides (example of Pb and Cr).
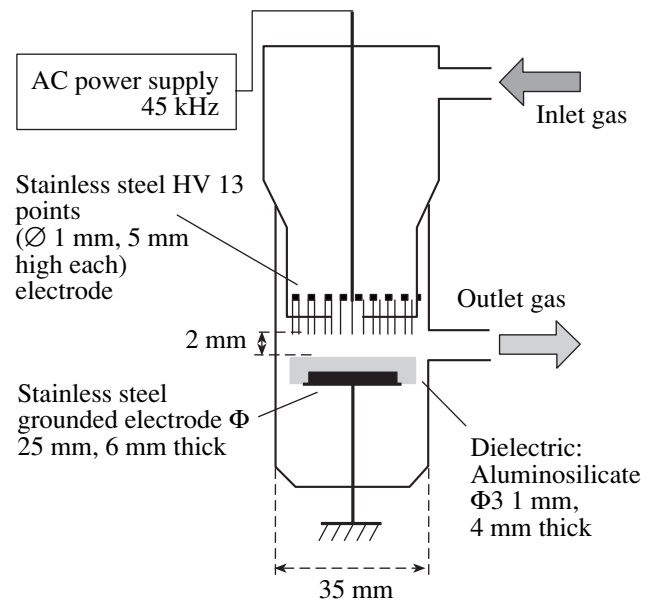
Fig. 14. The electric barrier discharge system: discharge plasma reactor.

plasma. Analysis of the desorbed organic species is performed directly by a mass spectroscopy quadripole connected by a capillary tube to the reactor.

The use of a nonequilibrium plasma for fly ash treatment provides a specific reactivity where electron

impact is the initiation step of the overall reaction without any thermal energy requirements.

The main advantages of this technique are the detection and the monitoring of the chlorine, the organochlorine molecules, and the alkane compounds trapped in

Fig. 15. Hydrocarbon analysis from CH_3CHO air mixture treated by DBD. 1 min at 80°C and 80°C $\longrightarrow$ 240°C at a rate of 10°C/min; 1 min at 240°C and 240°C $\longrightarrow$ 320°C at a rate of 10°C/min; N_2, 20% O_2, 1.5% acetaldehyde, $D = 1$ l/min, $d = 6$ mm.

Fig. 16. Mass balance of hydrocarbons from CH_3CHO air treated in a DBD. N_2, 20% O_2, 1.5% acetaldehyde, $D = 1$ l/min, $d = 6$ mm, Upk = 10 kV.

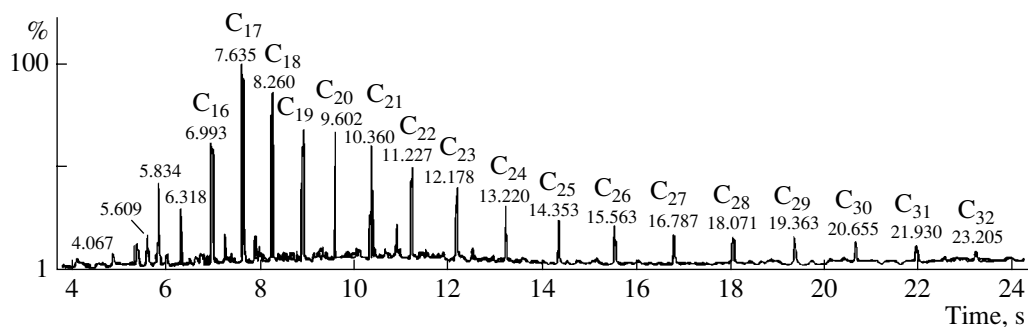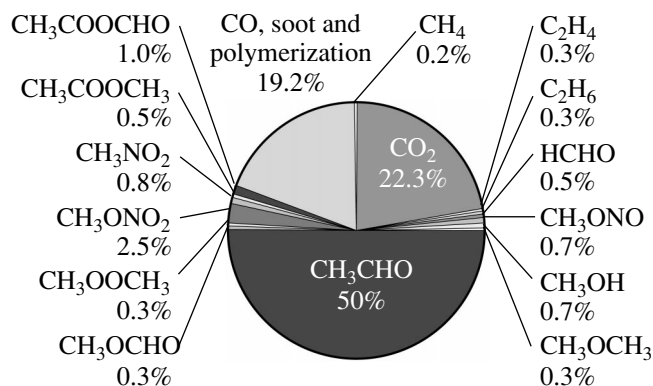the porous structure of the matrix. This process lasts for a few seconds. Depending on gas composition (Figs. 7–9), desorption of chlorine or chlorine acid is observed. Indeed, when pure argon is used, chlorine is desorbed preferentially. In the case of Ar–H_2 plasma, chlorine acid is the desorbed species, chlorine does not react with benzene molecules, and so the chlorobenzene decreases. Moreover, this technique indicates the production of the $COCl_2$ phosgene, which is easily synthesized by the reaction between CO and chlorine atoms from chlorinated metals.

2.3. Soxhlet Extraction and GC/MS Analysis

This technique uses dichloromethane for liquid extraction of organic molecules trapped in the porous structure of ash powders. The extraction requires 24 h, while concentration of the solvent and its analysis need a total of 2 h (Fig. 10). Table 2 gives a list of the main species analyzed by GC/MS. As this technique is very complex and time consuming, it is impossible to apply it in online control of an industrial process. The GC/MS analyses of the 75 dioxins (PCDD isomers) and the 135 furans (PCDF isomers) done by the team M.F. Gonnord showed the concentration in ng/g and the nature of the

isomers of PCDD and PCDF that have been identified. The European regulations of these compounds (0.1 ng/N m³) require a very long analysis procedure and, consequently, are very expensive.

3. PLASMA PROCESSES FOR INDUSTRIAL TREATMENT OF WASTE MATERIALS AND POLLUTANTS

The two methods that have been developed in the laboratory and used in industry are plasma vitrification of fly ashes and VOC elimination by DBD plasma process.

3.1. Plasma Vitrification for Ash Powders [9–17]

Arc plasma or RF plasma (Fig. 11) are used for ash powder treatment in order to reduce the volume and to produce a melted material that resembles slag.

Thermal plasma induces heating of the fly ashes, and melting of the matrix creating a slag. The result is the evaporation of volatile oxides or chlorines and heavy metals (Figs. 12 and 13) trapped in the melted slag (aluminosilicate matrix). The efficiency of the process depends on the fusion temperature. In order to reduce the fusion temperature of the slag, calcosodic materials are added. The evaporation depends on the plasma gas composition, and the competition between oxygen and hydrogen modifies the complex equilibrium composition as predicted by complex chemical equilibrium calculation.

3.2. Plasma Depollution by DBD Process for VOCs in Gas Flow [18–23]

The DBD discharge is a high speed electrical breakdown between a high voltage electrode and a ceramic grounded material. The gas is blown through the gap of the two electrodes, and an electrical discharge of a few nanoseconds (50–100 ns) is produced. The energy content of a few nanocoulombs leads to dissociation of the main molecules of the gas by electron impact and the formation of radicals from the VOC.
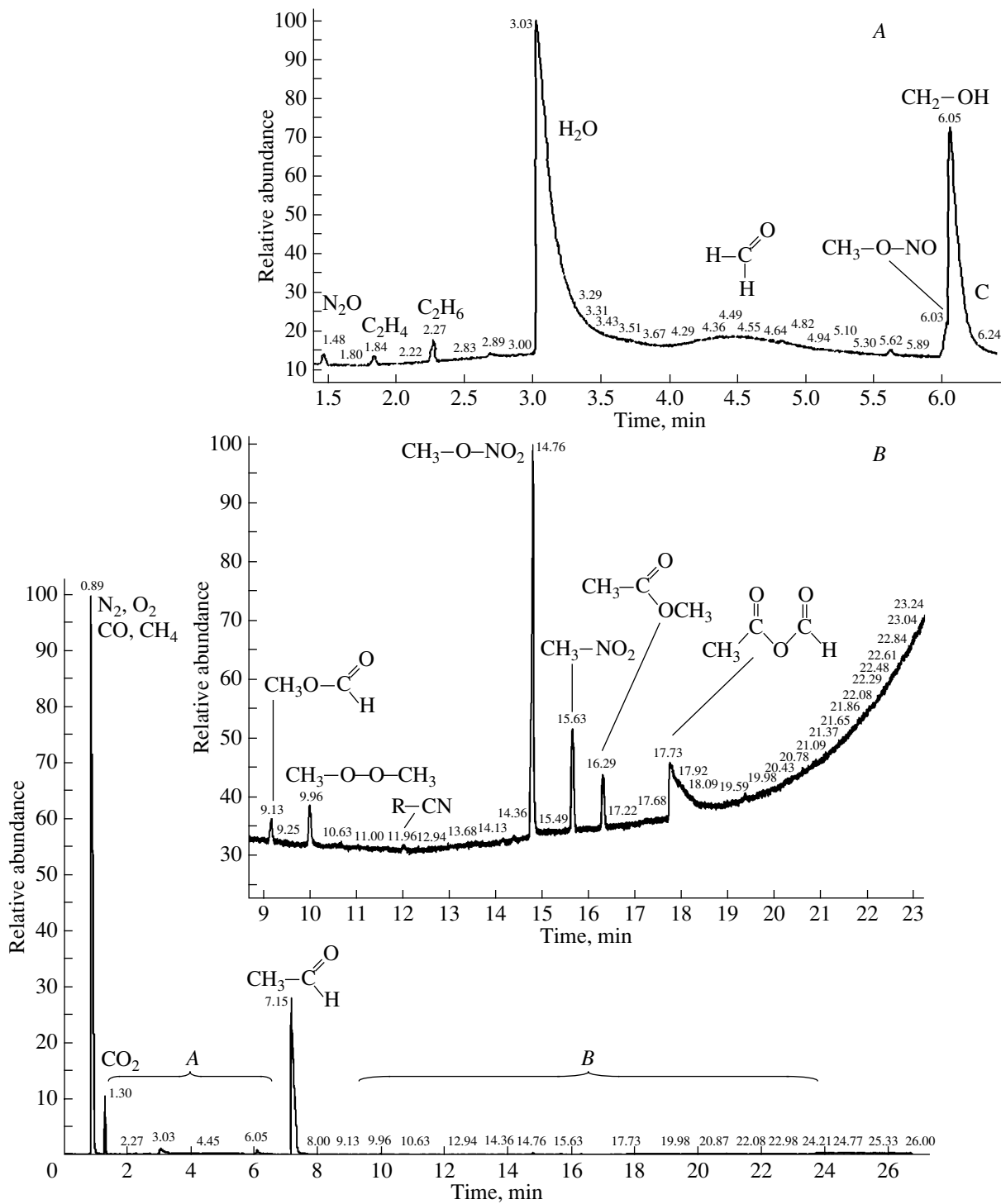
Fig. 17. Chromatogram exhaust gas composition from air CH_3CHO (1000 ppm) mixture treated in a DBD reactor N_2, 20% O_2, 1.5% acetaldehyde, $D = 1$ l/min, $d = 6$ mm.

The high kinetic rate of these excited species are responsible of molecules destruction, oxidation, or polymerization. Thus, a large part of the VOC is destroyed. The experimental devices have a multipoint-to-plane geometry (Fig. 14) or cylindrical geometry

with a high tension screw electrode in the center. These two main techniques are used for gas treatment. The mass balance from Fig. 16 gives some details of the excited phenomena responsible for the oxidation of CH_3CHO as an example of an organic pollutant.

The process is suitable for the elimination of hydrocarbon vapors or solvents from industry such as combustion, painting, solvents, printing, etc.

CONCLUSIONS

Plasma processes for depollution open new routes for gas treatment, toxic powders such as fly ashes, or chemical waste destruction. The main plasma processes use an arc jet for solid treatment, fluidized bed for halogen organic molecules, and DBD reactors for VOCs in very low concentrations in industrial emission gases. However, online gas analysis is a key for optimization of the processes through the working parameters. TRELIBS and non-plasma-enhanced desorption and mass spectrometry analysis, due to their high sensitivity and short time of analysis (less than 1 s), can meet the requirement of EU regulations for "continuous monitoring of emission pollutants." Also, the short residence time of toxic compounds in the plasma reactor allows decreasing the toxic risk of the treated material or gas, and these dry processes using arc discharge or streamers are very flexible techniques that agree with large modifications of the chemical composition of waste compounds.

ACKNOWLEDGMENTS

REFERENCES

1. J. Amouroux, H. Lancelin, K. Coulibaly, *et al.*, in *Proceedings of the 1st World Congress on Microwave Processing, Orlando, 1997*, Vol. 80, pp. 565–575.

2. V. Dudragne, Ph. Adam, J. Amouroux, *et al.*, Appl. Spectrosc. **52**, 1321 (1998).

3. H. Lancelin, L. Dudragne, P. Adam, *et al.*, High Temp. Mater. Processes **10**, 121 (1999).

4. S. Morel, P. Adam, and J. Amouroux, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 2001), pp. 11–22.

5. F. Genet, P. Rousseau, S. Cavadias, *et al.*, in *Proceedings of the 15th International Symposium on Plasma Chemistry, Orléans (France), 2001*, Vol. 7, pp. 3059–3065.

6. P. Rousseau, S. Cavadias, and J. Amouroux, High Temp. Mater. Processes **7**, 391 (2003).

7. P. Rousseau, F. Genet, S. Cavadias, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 2001), pp. 785–792.

8. S. Cavadias, P. Rousseau, F. Genet, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 2001), pp. 785–791.

9. K. Coulibaly, F. Genet, D. Morvan, *et al.*, in *Proceedings of the 5th European Conference on Thermal Plasma Processes, St. Petersburg, 1998*.

10. Ph. G. Rutberg, A. A. Safronov, A. N. Bratsev, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by E. Popov *et al.* (Begell House, New York, 2001), pp. 745–760.

11. N. Cerqueira, I. Ghiloufi, B. Barthelemy, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 2001), pp. 761–768.

12. S. Bernard, M. El. Ganaoui, P. Fauchais, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 2001), pp. 769–776.

13. V. L. Goriachev, A. I. Kulishhevich, A. A. Ufimtsev, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 2001), pp. 827–832.

14. Goldfard and G. Shneerson, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 1999), pp. 701–707.

15. K. Coulibaly, F. Genet, D. Morvan, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 1999), pp. 753–758.

16. F. Genet, K. Coulibaly, S. Cavadias, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 1999), pp. 837–845.

17. Ph. G. Rutberg, A. A. Ufimtsev, A. N. Bratsev, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 1999), pp. 821–827.

18. A. Vincent, F. Daou, and J. Amouroux, High Temp. Mater. Processes **6**, 167 (2002).

19. E. Francke, S. Robert, and J. Amouroux, High Temp. Mater. Processes **4**, 138 (2000).

20. J. Amouroux, H. Lancelin, K. Coulibaly, *et al.*, in *Proceedings of the 1st World Congress on Microwave Processing, Orlando, 1997*, Vol. 80, pp. 565–575.

21. L. Parissi, E. Odic, S. Dupre, *et al.*, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 2001), pp. 777–784.

22. E. Francke, S. Robert, and J. Amouroux, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 1999), pp. 759–765.

23. M. F. Renou-Gonnord, K. Coulibaly, and F. Genet, in *Progress in Plasma Processing of Materials*, Ed. by P. Fauchais and J. Amouroux (Begell House, New York, 1999), pp. 765–771.

====================== **SOLIDS** ======================

Adiabatic Invariance of Diffraction by Regularly Strained Crystals

M. B. Shevchenko and O. V. Pobidaĭlo

Institute for Metal Physics, National Academy of Sciences of Ukraine,
pr. Vernadskogo 36, Kiev, 03680 Ukraine
e-mail: sh_m@list.ru
Received February 4, 2004

Abstract—Adiabatic invariants occurring in the dynamic Laue diffraction of X rays, high-energy electrons, and thermal neutrons are calculated for asymmetric scattering by regularly strained crystals. A criterion for adiabatically smooth variations of strain is found using a generalized pendulum analogy. The conceptual equivalence of the adiabatic and ray modes of Bloch wave propagation in strained crystals is established. © 2005 Pleiades Publishing, Inc.

INTRODUCTION

The concept of adiabatic invariance, which originates from classical mechanics, has been extended to various fields of contemporary physics [1–4]. This concept was used to greatest advantage in quantum-mechanical applications, as well as in nuclear physics in designing magnetic traps. However, it has not been applied to the problems of dynamic diffraction of X rays, electrons, and neutrons by regularly strained crystals. At the same time, keen interest in such problems stimulates elaboration of theoretical methods taking into account multiple scattering by the strained crystal lattice. Thus, analysis of adiabatic invariance of the above processes seems to be a promising trend toward developing new analytic approaches to problems of such a sort.

In this study, the adiabatic invariants of the dynamic theory are determined for asymmetric Laue (transmission) diffraction. The invariants are valid for the diffraction of X rays, high-energy electrons, and thermal neutrons (provided that spin-dependent interaction is negligibly weak) owing to the equivalence of the corresponding equations of scattering. It should be noted that our results are correct when the displacement field slowly varies with distance. Therefore, bearing in mind asymmetric diffraction conditions, we worked out a criterion of adiabatically smooth lattice distortions. We believe that the invariants calculated here may be of interest in solving nonlinear equations of the dynamic theory. These equations would be much more readily integrable if the corresponding adiabatic constants were known. In this respect, it is worthy to note a nonlinear Riccati-type equation derived for the reflection coefficient [5], which can be solved in terms of the concept of adiabatic invariance.

ADIABATIC INVARIANTS FOR ASYMMETRIC LAUE DIFFRACTION

Assuming that lattice strains are smooth, we will solve the problem of adiabatic invariance under the most general asymmetric conditions of transmission diffraction. In this case, for amplitudes $\Phi_{0,h}$ of the fields of the transmitted and diffracted waves in a regularly strained crystal, we can write the Takagi–Taupin equations [6], which are valid for X rays, high-energy electrons, and thermal neutrons if nuclear scattering is strong,

$$
\begin{cases}
\gamma_0 \dfrac{d\Phi_0}{dz} = i\pi\sigma_h \exp\{i\mathbf{h} \cdot \mathbf{u} + isz\}\Phi_h + i\pi\sigma_0\Phi_0, & (1) \\[2mm]
\gamma_h \dfrac{d\Phi_h}{dz} = i\pi\sigma_h \exp\{-i\mathbf{h} \cdot \mathbf{u} - isz\}\Phi_0 + i\pi\sigma_0\Phi_h. & (2)
\end{cases}
$$

Here, quantities $\Phi_{0,h}$ may describe the corresponding amplitudes of the electric displacement field for X rays or the coordinate part of the wave function for electrons or neutrons. In these equations, $\sigma_{0,h} = \sqrt{\gamma_0\gamma_h}/\Lambda_{0,h}$, where $\Lambda_{0,h}$ are the extinction lengths in the directions of the transmitted and diffracted beams; γ_0 and γ_h are the cosines of the angles between the normal to the crystal surface and the propagation directions of these beams; s is the deviation from the position of Bragg reflection; and $\mathbf{u}$ is the strain field, which is regarded as one-dimensional and varying along the z axis, which is aligned with the normal to the crystal surface. Field $\mathbf{u}$ corresponds to regular strains in the crystal lattice, which may be associated, for example, with bending, a single dislocation, an ultrasonic wave, etc. It should be noted that, in the case of electron diffraction, the system of Eqs. (1) and (2) is also referred to as the Takagi–Howie–Whelan equations, which are derived in the column approximation of the scattering theory [7].

The adiabatic invariants corresponding to the diffraction problem formulated can be determined without resorting to the Takagi–Taupin equations. One may use an approach [8] that is based on determining the adiabatic invariants for the pendulum model. Generalizing the above to asymmetric diffraction conditions, we make the following change of variables in Eqs. (1) and (2):

$$\Phi_{0,h} = \frac{\tilde{\Phi}_{0,h}}{\sqrt{\gamma_{0,h}}} \exp\left\{ i\frac{\gamma_0}{2\Lambda_0}\left(\frac{1}{\gamma_0} + \frac{1}{\gamma_h}\right)\right\} \tag{3}$$

$$\times \exp\{\pm i\mathbf{h}\cdot\mathbf{u}/2 \pm isz/2\}.$$

Substituting (3) into Eqs. (1) and (2), we obtain

$$\frac{d\tilde{\Phi}_{0,h}}{dz} = \frac{2i}{\gamma_0}\sqrt{\xi_0^{(1)}\xi_0^{(2)}}\tilde{\Phi}_{h,0} \mp \frac{i}{\gamma_0}(\xi_0^{(1)} - \xi_0^{(2)})\tilde{\Phi}_{0,h}, \tag{4}$$

where coefficients $\xi_0^{(1)}$ and $\xi_0^{(2)}$ have the form

$$\xi_0^{(1,2)} = \left[\pm\frac{k\gamma_0}{4\gamma_h}\beta + \sqrt{\left(\frac{k\gamma_0}{4\gamma_h}\beta\right)^2 + \frac{\gamma_0^2}{4\Lambda_h^2}}\right]. \tag{5}$$

In expression (5), we use the standard designations for the deviation, $\beta = (\mathbf{h}\cdot d\mathbf{u}/dz + s)/k + \gamma_0(1 - \gamma_h/\gamma_0)/(k\Lambda_0)$, and the magnitude of the incident wave wavevector, $\mathbf{k}$. It is also worth noting that coefficients $\xi_0^{(1,2)}$ correspond to those variables in the dynamic theory defining, in the reciprocal space, the distance from points of excitation on the dispersion surface to the asymptote of the incident beam, which passes through the Lorentz point. Variables $\xi_0^{(1)}$ and $\xi_0^{(2)}$ then correspond to the points of excitation on the lower and upper branches, respectively.

The Takagi–Taupin equations represented in form (4) make it possible to pass from the diffraction problem to an equivalent mechanical problem with known adiabatic invariants. In our case, such an analogue is a mechanical system of two weakly coupled pendulums with different masses for which the following equations of motion are valid within sufficiently small time interval Δt:

$$\dot{\varphi}_{1,2} = \frac{i\omega_1\omega_2}{2\omega_0}\varphi_{2,1} \mp \frac{i(\omega_1^2 - \omega_2^2)}{4\omega_0}\varphi_{1,2}. \tag{6}$$

Here, $\omega_{1,2} = (\kappa/m_{1,2})^{1/2}$ and $\omega_0 = (g/l)^{1/2}$ are the frequencies, where $m_{1,2}$, κ, g, and l are the masses of the pendulums, coefficient of elasticity, free fall acceleration, and length of the pendulum suspension, respectively. Equations (6) written in "slow" angular variables $\varphi_{1,2}$ differ from the conventional Newton equations in that the former do not contain second-order derivatives. These derivatives will be negligibly small, because the oscillations are subdivided into a "slow" component and "fast" component with high frequency $\omega_H = \omega_0 +$

$\omega_b/2$, where ω_b is the beat frequency. Over time interval Δt considered, masses $m_{1,2}$ of the pendulums (hence, frequencies $\omega_{1,2}$) are assumed to be constant; consequently, we can establish local equivalence relations between the parameters of the diffraction and mechanical problems. It follows from equivalent equations (4) and (6) that these relations have the form

$$\varphi_{1,2} \longrightarrow \tilde{\Phi}_{0,h} \text{ and } \omega_{1,2} \longrightarrow 2\sqrt{\xi_0^{(1,2)}}, \tag{7}$$
$$\omega_0 \longrightarrow \gamma_0.$$

Relations (7) show that an angular deviation from the exact Bragg condition due to a regular strain or scattering asymmetry can be simulated by varying the masses of the pendulums in the given mechanical model. Then, as for a perfect crystal under symmetric diffraction conditions ($s = 0$), we must set $m_1 = m_2$. It should also be noted that the pendulum analogy proposed here remains correct in the case of extremely high strains, when the points of excitation are beyond the two-wave region of a local dispersion surface. It is known that, under such conditions, energy exchange between the transmitted and diffracted waves ceases and the refraction mode sets in. Obviously, a similar process will also be "observed" in terms of the pendulum model. Indeed, setting $\beta > 0$ for definiteness and assuming that $\xi_0^{(1)} \gg \xi_0^{(2)}$, we can easily derive from relations (7) the equivalent condition $m_2 \gg m_1$; under this condition, energy transfer between the pendulums ceases [9]. Naturally, the generalized pendulum analogy of asymmetric diffraction suggested here applies to X rays, high-energy electrons, and thermal neutrons. Provided that the masses slowly vary with time, the adiabatic invariants in the mechanical system will be variable actions

$$I_{1,2} = \frac{\langle E\rangle_{1,2}}{\Omega_{1,2}}. \tag{8}$$

Here, $\langle E\rangle_{1,2}$ are the pendulum energies averaged over period $T_{1,2} = 2\pi/\Omega_{1,2}$, where $\Omega_{1,2}$ are the frequencies of normal pendular oscillations. It is clear that to calculate $\langle E\rangle_{1,2}$ requires that the Lagrange function corresponding to the vibrations of coupled pendulums with variable masses be diagonalized. Passing to variable $\varphi_{1,2}$ and assuming that frequencies $\omega_{1,2}$ are functions of time, we derive from Eq. (8) the expressions for $I_{1,2}$,

$$I_{1,2} = (1 + \omega_{1,2}^2/\omega_{2,1}^2)|\varphi_1^\pm|^2. \tag{9}$$

Here, $\varphi_1^\pm$ are the respective amplitudes of higher and lower frequency oscillation modes into which the slow oscillations of the first pendulum split.

Taking into account relations (7) and using Eq. (9), we can easily find desired adiabatic invariants $C_{1,2}$ of the diffraction problem,

$$C_{1,2} = (1 + \xi_0^{(1,2)}/\xi_0^{(2,1)})|\Phi_0^\pm|^2, \tag{10}$$

where $\Phi_0^{\pm}$ are the amplitudes of the transmitted wave that correspond to the upper and lower branches of the local dispersion surface.

It should be noted that the adiabatic invariants of the dynamic theory can be represented in different ways. For this purpose, we transform wave field amplitudes $\Phi_0^{\pm}$ using the obvious relationships $\tilde{\Phi}_h^{\pm}/\tilde{\Phi}_0^{\pm} = \pm\sqrt{\xi_0^{(1,2)}/\xi_0^{(2,1)}}$. Taking into account these relationships, which are valid within a small neighborhood of point z, we can derive from (10) adiabatic invariants $P_{1,2}$ and $R_{1,2}$ in the form

$$ P_{1,2} = \sqrt{\frac{\xi_0^{(1,2)}}{\xi_0^{(2,1)}}} \frac{|\Phi_{0,h}^{+}|}{|\Phi_{0,h}^{-}|}, \quad R_{1,2} = \frac{|\Phi_{0,h}^{+}|}{|\Phi_{h,0}^{-}|}. \quad (11) $$

As usual, the adiabatic invariants in expressions (10) and (11) are accurate to within a constant factor independent of the variable. Invariants $C_{1,2}$ can be assigned a simple physical meaning. It is easy to check that these constants coincide with normal components $S_n^{\pm}$ and $J_n^{\pm}$ of the triply averaged Poynting vector and probability flux density vector for the corresponding Bloch waves (we mean here averaging over time, lattice period, and extinction length).

Using the pendulum analogy, one can also find the conditions under which the concept of adiabatic invariance applies to dynamic diffraction. To this end, it is necessary to apply the condition of slow time variation to a certain characteristic parameter of the mechanical problem (beat frequency ω_b is best suited to this case). Next, with local relations (7), the following criterion for adiabatically smooth strains in the crystal lattice can be derived:

$$ \kappa\frac{d\beta}{dz} \ll \Lambda_h^{-2}. \quad (12) $$

It was shown [10] that this inequality coincides with the condition of applicability for the Kato ray theory [11]. Thus, we may trace a correlation between the concept of adiabatic invariance and the idea of ray (geometrical) propagation of Bloch waves in crystals with a regularly distorted lattice. In our opinion, this inference would be helpful in deducing new fundamental laws of dynamic diffraction of X rays and elementary particles in imperfect crystals.

CONCLUSIONS

The concept of adiabatic invariance as applied to the theory of diffraction appears to be worthy of noting. Indeed, in view of the adiabatic nature of intrabranch scattering, the interbranch transition between the branches can be treated as transient beats, which resonantly build up when the strains are high [12]. In addition, the adiabatic invariants calculated here are expressed in terms of the universal parameters of the dynamic theory. This, in turn, makes it possible to unify the definition of adiabatic constants when the methods of X-ray, electron, and neutron diffraction are used. This circumstance is of special significance for neutron diffraction analysis, since it is usually combined with X-ray and electron diffraction investigation. It should be noted that much interest in the theory of dynamic diffraction of thermal neutrons by thick single crystals has recently arisen owing to technological breakthroughs favoring related experiments [13, 14]. In this respect, it would be appropriate to indicate the possibility of devising a dynamic approach based on the asymptotic methods used in the theory of nonlinear vibrations [15]. The pendulum analogy of asymmetric Laue diffraction proposed here could serve as a solid basis for such an approach.

REFERENCES

1. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics* (Nauka, Moscow, 1982; Pergamon, New York, 1988).
2. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989; Pergamon, New York, 1977).
3. M. Kruskal, J. Math. Phys. **3**, 806 (1962); translated under the title *Adiabatic Invariants* (Inostrannaya Literatura, Moscow, 1962).
4. V. K. Ignatovich, *The Physics of Ultracold Neutrons* (Nauka, Moscow, 1986; Clarendon, Oxford, 1990).
5. D. Taupin, Bull. Soc. Fr. Mineral. Crystallogr. **87**, 469 (1964).
6. Z. G. Pinsker, *X-ray Crystal Optics* (Nauka, Moscow, 1982) [in Russian].
7. *Electron Microscopy of Thin Crystals,* Ed. by P. B. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley, and M. J. Whelan (Plenum, New York, 1965; Mir, Moscow, 1968).
8. M. B. Shevchenko and O. V. Pobydaylo, Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. **59**, 45 (2003).
9. I. I. Ol'khovskiĭ, *Course of Theoretical Mechanics for Physicists* (Mosk. Gos. Univ., 1978) [in Russian].
10. A. Authier, *Dynamical Theory of X-ray Diffraction* (Oxford Univ. Press, Oxford, 2003).
11. N. Kato, J. Phys. Soc. Jpn. **18**, 1785 (1963).
12. M. B. Shevchenko, Acta Crystallogr. A **59**, 481 (2003).
13. M. Shlenker and J. P. Guigay, in *X-ray and Neutron Dynamical Diffraction: Theory and Applications*, Ed. by A. Authier, S. Lagomarsino, and B. K. Tanner (Plenum, New York, 1996), pp. 63–73.
14. V. L. Nosik and M. V. Koval'chuk, Poverkhnost: Fiz. Khim. Mekh., No. 1, 120 (2001).
15. N. N. Bogolyubov and Yu. A. Mitropol'skii, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Fizmatgiz, Moscow, 1958; Gordon Breach Sci., New York, 1961).

Translated by N. Wadhwa

================================ **SOLIDS** ================================

Dynamic Phase Transitions in Moving Vortex Structures in Type II Superconductors

S. V. von Gratowski

Institute of Radio Engineering and Electronics, Russian Academy of Sciences,
pl. Vvedenskogo 1, Fryazino, Moscow Oblast, 141190 Russia
e-mail: v-meriakri@gmx.net, meriakri@ms.ire.rssi.ru
Received July 9, 2004

Abstract—A theoretical explanation for the *I–V* characteristic of type II superconductors in the vortex state is provided. The run of the experimental *I–V* curves is associated with discontinuities and depends on the experimental prehistory. It is conjectured that dynamic phase transitions occur in the network of mobile and immobile channels along which the vortices travel when a transport current is applied. This work has been done on the author's own initiative. © 2005 Pleiades Publishing, Inc.

The phenomena associated with dynamic phase transitions (DPTs) in mobile vortex structures (VSs) that are observed in type II superconductors constitute a challenging area of the physics of vorticity. Such DPTs, or crossovers, occur when the VS order changes under the action of external effects (current, magnetic field, rf field, etc.). Change of order is frequently observed in periodic structures moving in the field of a random potential. A similar situation is also common in adjacent areas of physics and is encountered, i.e., in studying Wigner crystals, charge density waves, magnetic domains, and so on. Investigation into DPTs touches upon a number of dynamic issues, specifically, strains due to vortex motion, mutual arrangement of vortex lines, and *I–V* characteristics. Strains associated with VS motion can be both elastic and plastic. Plastic strains, in turn, are due to prehistory-related hysteretic phenomena. Plastic dynamics is an intriguing aspect of DPTs. Numerous studies of flux plastic flow (FPF) carried out both numerically [1, 2] and experimentally [3–5] showed that the FPF conditions set in when a certain threshold is exceeded. This threshold correlates with the values of transport current, pinning forces, and other parameters.

FPF can also arise in a current-induced defect superstructure associated with a vortex density gradient [6]. In this regime, the *I–V* curve is, as a rule, highly nonlinear, contains steps, and may be even of a hysteresis character, depending on the prehistory, and a plethora of dynamic effects is observed. In the case of FPF, the vortex structure is partitioned into regions (channels) with different kinds of motion [7]. Not only vortex motions through the channels but also interaction between the channels may be different. Some of the channels remain stationary, while others move. Sometimes, the motion of the channels is akin to that of pieces of ice relative to each other. When moving, the

channel system may reconfigure: new channels arise and the old ones collapse. There may appear bottle-neck-like constrictions in them. Plastic dynamics has been the subject of investigation in a number of papers [8–11].

It was shown [12] that, when the flux flows near the plastic–elastic crossover in a vortex glass, the dynamic friction force increases. An increase in the transport current gives rise to the following DPTs. At extremely low currents, the pinning state is observed. As the current grows, first elastic and then plastic flux flow arises. With a further increase in the current, the lattice becomes ordered. Similar results were obtained in [13] for disordered vortex structures. In particular, it was found that first hexatic ordering appears at a certain current and then the FPF of vortex liquid arises. As the current increases further within the range of hexatic ordering, the differential resistance drops. It was found experimentally [14] that the plastic flow of defects causes the differential resistance to drop because of vortex ordering and, hence, a decrease in the dynamic friction. Hysteresis and FPF-related steps in the *I–V* curve are also observed. At high currents, dynamic ordering takes place. A microscopic motion of vortices can hardly be seen experimentally; however, numerical simulation allows studying vortex microdynamics [15]. In [15], the FPF and, at high currents, a vortex crystal were also observed. It was found that the motion in the FPF range is bimodal and the histograms of the mean velocities have two peaks. The first one corresponds to the slow motion, i.e., the motion of vortex islands, which are pinned during most of the time. The second peak, observed at higher velocities, is associated with the fast vortex motion through the channels around the pinned islands. The bimodal pattern of the velocity histogram reflects the spatial nonuniformity of vortex instantaneous velocities. Under the FPF conditions, the

width of the mean velocity distribution is much larger than zero and but equals zero at the coherent motion. The first depinning threshold separates vortices with zero and nonzero velocities of the center of mass. Above the second depinning threshold, all vortices are depinned. In the range between the first and second thresholds, some vortices move, while others are fixed. Above the second threshold, all the vortices move but the width of the mean velocity distribution is nonzero; i.e., the motion is plastic. The peak in the differential resistance appears when the current somewhat exceeds the value at which the number of pinned vortices decreases drastically and the defect density declines [16]. Nonequilibrium processes, such as hysteresis, dependence on the prehistory, and flux inhomogeneity, are also associated with specific features of the FPF and influence the *I–V* curve and the dependence $dV(I)/dI$, where *I* is the current and *V* is the voltage. Many peaks in this dependence were found experimentally [17]. Presumably, they are related to discontinuities in the vortex lattice under the FPF conditions. Voltages *V* due to vortex motion are proportional to product $N_v\langle v\rangle$ ($V \propto N_v\langle v\rangle$), where N_v is the number of vortices moving with velocity *v* and $\langle v\rangle$ means velocity averaging over space. Then, the differential resistance is given by $R_d = N_v[d\langle v\rangle/I]$, where the term in the brackets is proportional to the friction coefficient. By analogy with the critical phenomena, it can be assumed that the power-law scaling *I–V* dependence for continuous depinning, which corresponds to a vortex liquid, has the form $V \propto (I - I_c)^{\alpha^1}$, where α^1 is the critical index. Under near-FPF conditions, the friction coefficient will be a smooth function of the current and the differential resistance will vary by a saw-tooth law under the FPF conditions. If N_v also depends on the current, the saw-tooth pattern of dN_v/dI will correlate with the peaks and dips in the curve $dV(I)/dI$. Such a saw-tooth pattern would be evidence that depinning proceeds irregularly and the voltage spatial distribution is nonuniform. With increasing current, depinning is accomplished piecewise. In this case, $V \propto n_j\langle v_j\rangle$, each of the channels or teeth contributing to R_d. Every "piece" can be described in terms of the time-averaged correlation length of velocity. In the range where the *I–V* curve has steps, the dependence of the *I–V* curve on the prehistory is also observed. A higher mobility observed near the steps persists when the current decreases. This is associated with the transition between metastable states. One can trace an analogy between this process and the coexistence of different phase states at the first-order phase transition. Steps in the *I–V* curve and its dependence on the prehistory were also discovered in [18–21]. In most of those papers, such processes were also explained by the coexistence of dynamic metastable variously ordered states, FPF, channels, and two-stage depinning. It was also noted that overheated states may appear as the current decreases and that the *I–V* curve has no jumps in these conditions.
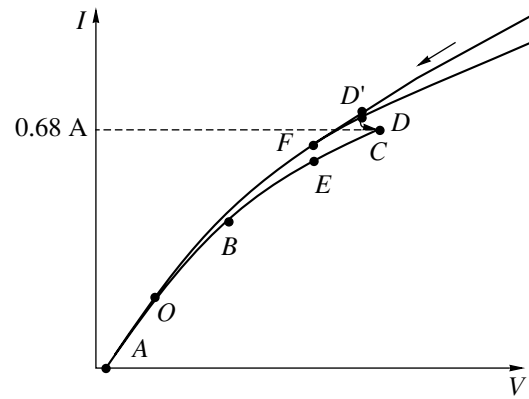
I–V curve taken of BiSrCaCuO platelets placed in a magnetic field.

In this paper, the author reports a theoretical study of *I–V* curves taken of BiSrCaCuO platelets placed in a magnetic field. Experimental data discussed below were courteously submitted by Ogrin [22]. Specifically, the *I–V* curves were taken of BiSrCaCuO platelets measuring $0.8 \times 0.3 \times 0.01$ cm, which were placed in magnetic field $H_0 = 1500$ Oe (directed normally to the surface and parallel to the *c* axis and *Z* coordinate axis) at 79 K. A direct current passed along the *Y* axis. The *I–V* curves behaved as follows. As the current was increased for the first time, the *I–V* curve (hereafter, the curve) passed through points *A*, *O*, *B*, *E*, *C*, and *D* (see figure). If the current started decreasing from a value above point *O*, the curve passed through points *D*, *F*, *O*, and *A*. When the current was increased for the second time, the result depended on the minimal current value in the first experiment. If the minimal current was lower than at point *O*, the hysteretic behavior was repeated. If the current was higher than at point *O*, the curve followed portion *OED* irrespective of whether the current was increased or decreased; in other words, the behavior of the curve depended on the experimental prehistory.

According to [23], the VS is in the liquid state under our experimental conditions. There are two typical time scales τ_{pin} and τ_{th} in an ordinary vortex liquid: the characteristic pinning-related scale and the characteristic thermodynamic scale in the liquid. Since $\tau_{pin} \gg \tau_{th}$, averaging over thermal fluctuations results in a homogenous unpinned structure. However, in a viscous liquid, which may undergo plastic deformation, one more characteristic time scale related just to plastic deformation, $\tau_{pl} \gg \tau_{pin}$, appears. In the case of averaging over plastic-deformation-related times, thermal averaging over time will be incomplete; therefore, a vortex liquid may be pinned by a random potential [24, 25]. In a plastic vortex liquid, two ohmic regimes are established: at high current densities (flux flow) and at low current densities. Between them, an intermediate highly nonlinear regime of thermally activated flux flow (TAFF) occurs. In this regime, the current is a power-type function of the voltage, with the exponent exceeding unity.

The initial part of the curve obtained in the experiments is ohmic and reversible. The reversible part is followed by the irreversible part, where the curve is nonlinear (power-type) and has a positive curvature. Such a picture corresponds to the initial part of the TAFF. Above point O, the curve exhibits steps, hysteresis behavior, and dependence on the prehistory. It can therefore be assumed that, above point O, we are dealing with the plastic TAFF of the vortex liquid and that the DFT to the plastic flow conditions takes place at point O. In the plastic flow range, the vortex liquid is metastable and exhibits many metastable states with their own energy minima. At the elastic–plastic flux flow transition, the vortex system is partitioned into channels, which are stationary, near-stationary (moving with very small velocities), or mobile. Some of the vortices that are fixed or become less mobile, pass to the system of more intense pining centers with the formation of corresponding (stationary or near-stationary) channels. The rest of the vortices, which produce mobile channels, conversely, pass to the system of less intense pinning centers. In systems with similar I–V curves, the depinning process may go in two stages. If such systems have channels with different mobility, estimation of the critical current must include two values of the effective critical current: J_{c1} for the threshold of partial depinning and J_{c2} ($J_{c2} \gg J_{c1}$) for complete depinning. Let us estimate J_{c1} and J_{c2} using the following line of reasoning. Pinning in a vortex liquid is much weaker than in the other phases; therefore, for bismuth cuprates, we shall take the mean value of the critical current $J_c = 10^4$ A/cm^2, according to [26]. In the experiments, the mean current density above point O is 2×10^2 A/cm^2; so, it can be assumed that the critical current of partial depinning is $J_{c1} = 10^3$ A/cm^2 and $J_{c2} = 10^4$ A/cm^2 is the critical current of complete depinning. If it is supposed that $J_1 < J_c$ in the mobile channels, then we must admit that they are in the region near the region of partial depinning and, that, in the stationary channels, $J_{c2} \gg J_2$ (J_2 is the current density in the stationary channels). The stationary channels are far from depinning. Now let us estimate the characteristic times in the vortex liquid. The time of thermal fluctuations is $\tau_{th} \approx 8\kappa^2 a_0^2/c^2\rho_n$, where κ is the Ginzburg–Landau parameter, ρ_n is the resistance of the normal phase, a_0 the lattice parameter, and c is the speed of light. Putting $\kappa = 50$, $a_0 = 1.1$ cm^{-5}, and $\rho_n = 50$ $\mu\Omega$cm, we obtain $\tau_{th} \sim 10^{-11}$ s. To estimate characteristic time of pinning τ_{pin}, we will reason as follows. The pinning time is $\tau_{pin} \approx r_{pin}/v_c$. In the vortex liquid, $r_{pin} \approx \xi + \langle u_{th}^2 \rangle^{1/2}$, where $\langle u_{th}^2 \rangle^{1/2}$ is the mean square of thermal fluctuations and ξ is the coherence length. The mean square of thermal fluctuations can be estimated from the Lindemann criterion for vortex lattice melting, $\langle u_{th}^2 \rangle^{1/2} \approx c_L a_0$ (c_L is the Lindemann number). Parameter v_c is determined from the critical

current values described above. Then, $\tau_{pin} \sim 10^{-10}$ s. Considering that point O manifests the transition to the FPF, so that typical times of VS motion become equal to the time of plastic deformation, τ_{pl}, and estimating the characteristic time of vortex motion at point O from the experimental data, we find that $\tau_{pl} \sim 10^{-9}$ s. It then follows that the conditions for vortex liquid pinning are satisfied; hence, the liquid under study is pinned. Such a viscous vortex liquid exists only in the range of long times, $\tau \gg \tau_{pl} \gg \tau_{pin}$. For processes characterized by the intermediate time scale, $\tau_{pl} \leq \tau_v \leq \tau_{pin}$ (τ_v is the characteristic time of vortex motion in the range of plastic flow), $\tau_v \approx a_0/v$, where v is the vortex velocity. In this time range, a vortex liquid is in many ways similar to a vortex crystal. In particular, the shear modulus of such a vortex liquid equals c_{66}. In the case under consideration, we have $\tau_v \approx \geq 10^{-9}$ s. Thus, the vortex liquid studied is similar to a vortex crystal in elastic properties.

Now let us turn to processes arising as the current increases. Such of them lead to a rise in the differential resistance. These may be (1) ordering, which results in an increase in the dynamic friction; (2) expansion of the mobile channels; and (3) an increase in the velocity in these channels. A combination of several mechanisms is also possible. The increase in the differential resistance is observed in the portion near point E, where a small increase in the current causes a drastic increase in the voltage. Then, as the current increases further, the arising dynamic configuration becomes energetically unfavorable. However, it may exist as an overheated state, since a radical reconfiguration of the system requires that the barriers for plastic flow be overcome and that the system transform into a new set of pinning centers. When the transport current becomes sufficiently high, the system is capable of overcoming the barriers. In this situation, some of the channels collapse and a bottleneck-like configuration, along with a new channel distribution, may appear. Some of the vortices become captured by the new pinning centers and stop. This process corresponds to the descending portion of the curve. As the current goes on rising, the process will be repeated but with another initial current and another initial channel distribution. Switching from one channel system to another corresponds to portion CD, and a new cycle starts at point D. After the channels have been rearranged, the current stars decreasing and the state of the system corresponds to portion $D'FO$, since the system is already in a new energy well and cannot overcome the energy barriers. As the current increases for the second time, two scenarios are possible. If its minimal value at the time of decrease exceeds the value at point O, the system remains in the same energy well, in which it was when the current decreased. This is because the energy stored in the system does not suffice to overcome the energy barriers preventing the system from being transferred to the energy well corresponding to portion OC. If the minimal value at the time of decrease is less than the value at point O, the system

falls into the range of elastic flow, where plastic deformations are absent. In other words, the system returns to the state where it had no memory. In this case, the primary process is repeated as the current grows.

At point *C*, the energy increases to the point where the system can overcome the energy barriers and pass to the state corresponding to a neighboring metastable energy minimum. Let us estimate the free energy of the system at points *C* and *D* and the related energy barrier. The system's free energy can be expressed by

$$F = \int (F_{\text{pin}}^{(1)} + F_{\text{el}}^{(1)} - F_{\text{L}}^{(1)})dV_1$$
$$+ \int (F_{\text{pin}}^{(2)} + F_{\text{el}}^{(2)} - F_{\text{L}}^{(2)})dV_2, \tag{1}$$

where upper and lower indices "1" hereafter refer to the mobile areas and indices "2," to the stationary ones; V_1 and V_2 are the current-dependent volumes occupied by the mobile and stationary areas, respectively; and F_{pin}, F_{el}, F_{L} are the densities of the pinning force energy, elastic deformation energy, and Lorentz force energy, respectively.

When estimating these energy densities, one should bear in mind the following. If a vortex is captured by a pinning center, its maximum possible displacement is on the order of the pinning center size. However, this statement is untrue in the case of thermal creep [23, 27]. During thermal creep, vortices jump to a much farther position; in this case, the displacement is much larger than when they are pinned. Using estimates made in the papers cited, we can put $u_1 = \chi u_2$, where u is the vortex displacement ($\chi \gg 1$).

When the current is applied under the TAFF conditions, the vortices jump from one metastable state to another that is energetically more favorable at the given current. A new optimal state can be determined from the condition that the energy contribution from the Lorentz force is equal to a change in the energy of deformation and pinning. For near-critical current values, this condition is satisfied for a neighboring metastable state. In such transitions, which occur at low current densities, a vortex is bound to travel a large distance. For pinned channels, where $J_{c2} \gg J_2$, adjacent metastable energy levels are spaced by the barriers [23]

$$U(J_2) \approx U_c \left(\frac{J_{c2}}{J_2}\right)^{\mu}.$$

For the plastic flow of a vortex liquid, $U_c \cong U_{\text{pl}}$, critical index μ is $\cong 1/7$ [25]. In the mobile areas, the energy barriers between adjacent metastable states are on the order of

$$U(J_1) \approx U_c \left(1 - \frac{J_1}{J_{c1}}\right)^{\alpha},$$

where α is the critical index that equals unity in the Kim–Anderson model.

To make estimates, we assume hat $J_1 \approx \chi_1 J_{c1}$ and $J_2 \approx \chi_1^2 J_{c1}$, where $\chi_1 \ll 1$. It was shown [23] that the difference in the displacements of the pinned and mobile vortices leads to the difference in pinning energy that is of the same order of magnitude. With the aforesaid taken into account, we get

$$F_{\text{pin}}^{(1)} \approx \chi U_{\text{pl}} \left(1 - \frac{J_1}{J_{c1}}\right)^{\alpha}, \quad F_{\text{pin}}^{(2)} \approx U_{\text{pl}} \left(\frac{J_{c2}}{J_2}\right)^{\mu}. \tag{2}$$

The Lorentz force energy density can be estimated as follows:

$$F_L = \frac{1}{c} JBu, \tag{3}$$

where J is the transport current density and B is the external magnetic field.

Using the same line of reasoning as in estimating the displacements of the mobile and fixed areas and taking into account that the pinning center size in a pinned vortex liquid is $r_p \approx \langle u_{\text{th}}^2 \rangle \approx c_L a_0 \leq a_0$, we obtain

$$F_L^{(1)} \approx \frac{1}{c} \chi J_1 c_L a_0 B, \quad F_L^{(2)} \approx \frac{1}{c} \chi_1 J_1 c_L a_0 B. \tag{4}$$

The mean current density at point *C*, $J^{(C)}$, is equal to 2.2×10^2 A/cm²; at point *D*, $J^{(D)} = 2.3 \times 10^2$ A/cm². The current difference between points *C* and *D* is $\Delta J = 0.1 \times 10^2$ A/cm² $\ll J^{(C, D)}$. In light of this, we suppose that the basic thermodynamic values (the current-dependent Lorentz force and pinning force energy densities), change insignificantly in passing from point *C* to point *D*. A change in the elastic energy density, which depends largely on the magnetic field, is also assumed to be negligible. Thus, in passing from point *C* to point *D*, a change in the volumes of the mobile and stationary channels due to channel rearrangement makes a major contribution to a change in the free energy. Also, in passing from point *C* to point *D*, a change in the current goes in parallel with a change in the voltage. At point *C*, the voltage is $V^{(C)} = 72$ mV; at point *D*, $V^{(D)} = 67.8$ mV. Thus, the change in the voltage equals $\Delta V^{(C, D)} = 4.2$ mV. The difference in the power between points *C* and *D* is $\Delta N \approx 1.5 \times 10^4$ erg/s. For estimation, we assume that the energy release due to a negative differential resistance reflects the fact that some of the vortices turn into a new metastable state with a lower energy. The plastic vortex flow results in the transition from state *C* to state *D*. Therefore, we suppose that time τ_N characterizing this process is associated with plastic deformation and also take into account the displacement of the vortices under creep [23]. With this in mind, we estimate τ_N as $\tau_N = \tau_{\text{pl}}\chi$ and the energy barrier between states *C* and *D* as $\Delta E = \Delta N \tau_N$. This energy barrier can be neglected, since the difference in the voltages and currents between points *D* and *D*' is much smaller than the voltage and current differences

between all other points considered. Then,

$$[-(F_L^{(1)} - F_L^{(2)}) + (F_{el}^{(1)} - F_{el}^{(2)})$$
$$+ (F_{pin}^{(1)} - F_{pin}^{(2)})]\Delta V = \Delta E. \tag{5}$$

Here, $\Delta V = V_1^{(C)} - V_1^{(D)} = -(V_2^{(C)} - V_2^{(D)})$. For estimations, we assume that the sum of the second and third terms in the parentheses in Eq. (5) is equal to the energy density of the barriers separating the metastable states. To determine this energy density, we estimate the critical volume in the vortex liquid reasoning as follows. Suppose that the critical force corresponding to the second threshold of depinning is given by

$$F_{C2} \approx \frac{U_C}{uV_C}, \quad u \approx c_L a_0.$$

Taking into account that $U_c \approx U_{pl}$ for the plastic vortex liquid, where U_{pl} can be approximated as

$$U_{pl} \approx \varepsilon_0 a_0, \quad \varepsilon_0 \approx \left(\frac{\Phi_0}{4\pi\lambda}\right)^2,$$

and neglecting the second-order terms, we arrive at

$$V_C \approx \frac{U_{pl}}{c_L a_0}\left(\frac{J_{c2}B}{c}\right)^{-1},$$

$$\Delta V = \frac{\Delta E}{\left[\frac{1}{c}J_1 c_L a_0 B\chi + \frac{U_{pl}}{V_C}\left[\left(1 - \frac{J_1}{J_{c1}}\right)^\alpha - \chi_1\left(\frac{J_{c2}}{J_2}\right)^\mu\right]\right]}. \tag{6}$$

Substituting the experimental and calculated data into Eq. (6) yields $\Delta V \approx 5 \times 10^{-5}$ cm^3. The volume of the platelet is about 2.4×10^{-3} cm^3. Comparing the platelet volume with the DPT-related change in the volume of the mobile regions shows that the volume of the stationary regions is roughly 2% of the total volume. The DPT-related change in the voltage is also about 2% of its value.

The voltage due to vortex flux flow in type II superconductors is proportional to the density of mobile vortices. Hence, a change in the voltage is bound to be proportional to a change in the number of mobile vortices, which is consistent with the estimates made in this paper and confirms the assumption that a dynamic phase transition is responsible for a voltage step in the I–V curve. Thus, it is shown that the mobile vortex system in type II superconductors exhibits a DPT and this DPT is associated with the rearrangement and collapse of the channels through which the vortices move. As a result, a descending branch of the I–V curve appears.

REFERENCES

1. H. J. Jensen, A. Brass, Y. Brechet, *et al.*, Phys. Rev. B **38**, 9235 (1988).
2. A. Brass, H. J. Jensen, and A. J. Berlinsky, Phys. Rev. B **39**, 102 (1989).
3. A. Pruymboom, P. H. Kes, E. van der Drift, *et al.*, Phys. Rev. Lett. **60**, 1430 (1988).
4. F. Nori, Science **271**, 1373 (1996).
5. Tsuyoshi Matsuda, Ken Harada, and Hiroto Kasai, Science **1393**, 204 (1996).
6. D. W. Braun, G. W. Grabtree, H. G. Kaper, *et al.*, Phys. Rev. Lett. **76**, 831 (1996).
7. H. J. Jensen, A. Brass, and A. J. Berlinsky, Phys. Rev. Lett. **60**, 1676 (1988).
8. E. M. Chydnovsky, Phys. Rev. Lett. **65**, 3060 (1990).
9. An-Chang Shi and J. Berlinsky, Phys. Rev. Lett. **67**, 1926 (1991).
10. I. Aranson and V. Vinokur, Phys. Rev. Lett. **77**, 3208 (1996).
11. I. Aranson and V. Vinokur, Phys. Rev. B **57**, 3073 (1998).
12. A. E. Koshelev and V. M. Vinokur, Phys. Rev. Lett. **73**, 3580 (1994).
13. R. Seungoh, M. Hellerqvist, S. Doniach, *et al.*, Phys. Rev. Lett. **77**, 5114 (1996).
14. M. C. Hellerqvist, D. Ephron, W. R. White, *et al.*, Phys. Rev. Lett. **76**, 4022 (1996).
15. M. C. Falesky, M. C. Marchetti, and A. A. Middleton, Phys. Rev. B **54**, 12427 (1996).
16. S. Spencer and H. J. Jensen, Phys. Rev. B **55**, 8473 (1997).
17. S. Bhachachrya and M. J. Higgins, Phys. Rev. B **52**, 64 (1995).
18. J. A. Fendrich, U. Welp, W. K. Kwok, *et al.*, Phys. Rev. Lett. **77**, 2073 (1996).
19. W. Henderson, E. Y. Andrei, M. J. Higgins, *et al.*, Phys. Rev. Lett. **77**, 2077 (1996).
20. N. R. Dilley, J. Herrmann, S. H. Han, *et al.*, Phys. Rev. B **56**, 2379 (1997).
21. S. N. Gordeev, D. Bracanovie, A. P. Rassau, *et al.*, Phys. Rev. B **57**, 645 (1998).
22. Yu. F. Ogrin, private communication.
23. G. Blatter, M. V. Feigelman, and V. B. Geshkenbein, Rev. Mod. Phys. **66**, 1125 (1994).
24. V. M. Vinokur, M. V. Fiegel'man, V. B. Geshkenbein, *et al.*, Phys. Rev. Lett. **65**, 259 (1990).
25. V. B. Geshkenbein, V. M. Vinokur, and R. Fehrenbacher, Phys. Rev. B **43**, 3748 (1991).
26. C. J. van der Beek, V. B. Geshkenbein, and V. M. Vinokur, Phys. Rev. B **48**, 3393 (1993).
27. M. Büttiker and R. Landauer, Phys. Rev. A **23**, 1397 (1981).
28. M. V. Feigelman, V. B. Geshkenbein, A. I. Larkin, *et al.*, Phys. Rev. Lett. **63**, 2303 (1989).

Translated by M. Astrov

=========================== **SOLIDS** ===========================

On the Yield Strength of Single-Crystal Zinc under Uniaxial Compression in a Plane Shock Wave

G. S. Bezruchko, G. I. Kanel, and S. V. Razorenov

Institute of Problems of Chemical Physics, Russian Academy of Sciences,
Institutskiĭ pr. 18, Chernogolovka, Moscow Oblast, 142432 Russia
e-mail: bezgs@ficp.ac.ru
Received September 7, 2004

Abstract—The shock wave pattern in zinc single crystals compressed in the direction normal to the basal plane of the crystal is traced with a high time resolution. In this geometry of the experiment, plastic deformation of the zinc is found to begin at a shock compression pressure above 15 GPa and is not accompanied by splitting of the shock wave into an elastic precursor and a plastic compression wave. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

Shock waves in elastic–plastic materials are known to be unstable because of the difference in the longitudinal compressibilities in the elastic and plastic strain ranges and, therefore, split into an elastic precursor and a shock wave of "plastic" compression in a certain stress range [1]. The elastic precursor amplitude is measured for determining the yield strength of a material at submicrosecond loading times [2]. However, when experimenting with zinc single crystals [3], we found that the shock wave remains intact at a certain geometry of the experiment. Figure 1 shows typical free-surface velocity profiles in planar single-crystal zinc specimens variously oriented relative to the loading direction that were subjected to shock compression [3]. The compression waves are seen to be radically different depending on the direction of their propagation.

Zinc has a hexagonal close-packed (hcp) structure with an anomalously large lattice parameter ratio, $c/a = 1.856$ (the "ideal" ratio for the hcp lattice is $c/a = 1.633$). Therefore, atomic bonds in a basal plane are much stronger than those between basal planes and the compressibility of zinc in the direction normal to the basal plane is well above that in the transverse directions. According to a review [4] devoted to the elastic properties of zinc, its compressibility in the axial direction is very close to the bulk value. Because of this, deviator stresses arising during uniaxial compression in the $\langle 001 \rangle$ direction build up slowly, so that deformation remains elastic over a wide stress range. The purpose of this work is to reveal the signs and conditions of the onset of plastic deformation in zinc single crystals shock-compressed in this direction.

EXPERIMENTAL

Zinc single crystals of 4*N* purity were grown by directed crystallization from high-purity graphite cruci-

bles [5] so that the normal to the surface of the crystals measuring $100 \times 16 \times (3–4)$ mm was aligned with the $\langle 001 \rangle$ direction. Pieces with a 10×16-mm cross section spark-cut from the single crystals were cleaved along basal planes in liquid nitrogen to obtain specimens 0.3–1.0 mm thick. Since the available data on the elastic moduli of zinc diverge considerably, we additionally measured the longitudinal sound velocity in the $\langle 001 \rangle$ direction; it was found to be $c_l = 2.98 \pm 0.03$ km/s. The Hugoniot adiabat of polycrystalline zinc was measured repeatedly [6–10]: in the particle velocity range 0.5–3.5 km/s, the known experimental data can be approximated by the relationship $U_s = 3.0 + 1.57u$ (where U_s is the shock wave velocity and u is the particle velocity of the shock-compressed material) with an accuracy of
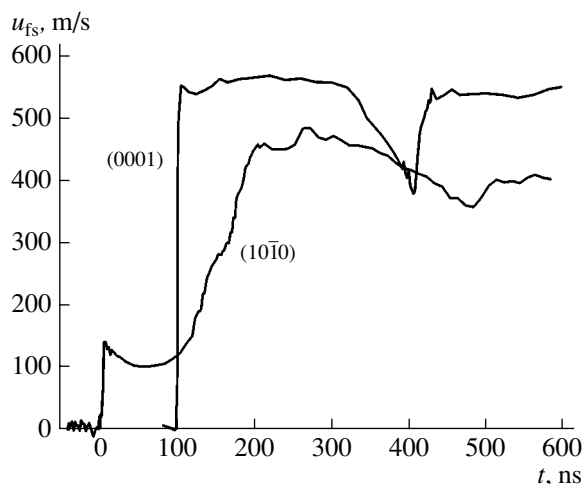

Fig. 1. Typical free-surface velocity profiles for 1.7-mm-thick single-crystal zinc specimens loaded by a 0.85-mm-thick aluminum flyer plate moving with a velocity of 650 m/s. The orientations of the planes of impact are indicated.
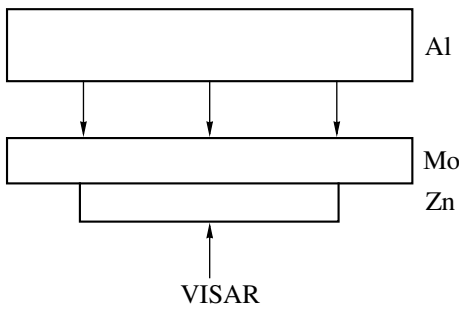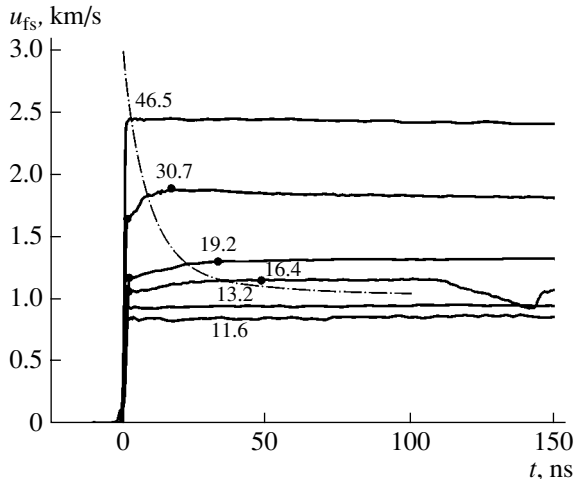
Fig. 2. Scheme of the experiment.

Fig. 3. Free-surface velocity profiles for the 0.25–0.5-mm-thick $\langle 001 \rangle$ zinc specimens at various shock compression pressures. Shock waves producing a pressure above 30 GPa were generated by an explosive detonating in contact with the base plate; in the other cases, the waves were generated by the colliding plates. Maximum compression stresses behind the shock front are indicated.

±1%. In our recent work [11], we determined the dependence of the longitudinal sound velocity in (001) zinc single crystals on the elastic compression/tensile stress. In the quasi-acoustic approximation, the results are described well by the relationships

$$a = c_0 + 2bu, \quad a = (c_0^2 + 4b\sigma_x/\rho_0)^{1/2}, \qquad (1)$$

where a is the sound velocity in the Lagrangean coordinates, σ_x is the compression/tensile stress, $c_0 = 2.99$ km/s, and $b = 1.933$.

The experimental scheme is shown in Fig. 2. The zinc specimens were impact-loaded by a planar aluminum flyer plate accelerated by explosion. The shock compression pressure was varied by changing the impact velocity and using intermediate base plates made of materials with different dynamic impedances, namely, aluminum, polymethyl methacrylate (PMMA), and molybdenum. A 0.02-mm-wide gap was provided between the molybdenum base plate and the specimen

to develop an appropriate initial shock compression and eliminate the effect of the elastic precursor in the plate on the loading history of the test specimen. In the experiments, we recorded the history of free (rear) surface velocity $u_{fs}(t)$ with a VISAR laser Doppler velocimeter [12]. The bandwidth of the measuring channel was 0–350 MHz or wider. The output signal of the velocimeter was applied to a four-channel digital oscilloscope at a sampling rate of 2.5 GHz (0.4-ns intervals between record points). The ultimate time resolution of the VISAR in the configuration used was 1 ns. To ensure reliable recording, the rear surface of the specimen was matted by grinding, polishing, and etching in hydrochloric acid.

RESULTS AND DISCUSSION

The free surface velocity profiles for different shock compression pressures are summarized in Fig. 3. At a particle velocity of up to 1 km/s (a compression stress of ≈14 GPa), the compression wave is essentially a discontinuity in the parameters with a rise time within 1.0–1.5 ns followed by a plateau where the parameters are constant. As the shock compression pressure rises, the shock front is followed by a region where the parameters increase smoothly. The duration of this dissipative region decreases with increasing intensity of the shock wave. It is natural to assume that the appearance of the dissipative region indicates the onset of plastic deformation in the shock wave. Since the rate of plastic deformation, being limited by the rate of dislocation motion and multiplication, cannot be infinitely high, plastic deformation is related to a finite time of stress relaxation and the corresponding dispersion of the shock wave. Apparently, the decrease in the stress relaxation time with increasing shock compression pressure can be explained in the same terms as the routinely observed narrowing of plastic shock waves [13].

Figure 4 compares the results in the normalized coordinates. After the shock wave has reached the surface of the specimen, a reflected rarefaction wave forms in it. This wave propagates toward the base plate and may rereflect from its boundary (Fig. 5). The time intervals in Fig. 5 take into account the time it takes for rereflected compression or rarefaction waves (which form due to different dynamic impedances ρc of the specimen and base plate) to reach the surface. Aluminum and PMMA have lower dynamic impedances than zinc. Therefore, the rarefaction wave reflects from zinc in the form of a compression wave. At the same time, the reflection of the rarefaction wave propagating in the zinc from the more rigid molybdenum plate produces another rarefaction wave. In the former case, when the rereflected wave reaches the surface, its velocity increases, whereas, in the latter case, its velocity decreases. The measurements demonstrate that the shape of the rereflected compression wave changes and the rise time of the parameters in this wave increases when the region of stress relaxation appears behind the

shock front of the first wave. Thus, at a compression stress of 14–15 GPa behind the shock front, the behavior of the material changes qualitatively, presumably manifesting the onset of plastic deformation.

The distance–time diagram shown in Fig. 5 allows us to estimate the reverberation time for elastic and plastic waves in the specimen. The estimation procedure using this diagram is the following. We separate triangle *FBD* bounded by the trailing C_- characteristic of centered rarefaction wave *FB* and the leading C_+ characteristic of rereflected tensile pulse *BD*. The states along these characteristics correspond to zero pressure; therefore, their slope yields zero-pressure sound velocity c_0. Then, the position of point *B* in the diagram can be easily determined if we know time interval $t_D - t_F$ (where t_D is the time instant the sample surface starts quieting down and t_F is the time instant the shock wave reaches the surface) provided that the value of c_0 is also known (or specified). Section *AB* of the leading C_+ characteristic of the rereflected wave is curved. Since the slopes of the C_+ and C_- characteristics in the substantial Lagrangean coordinates are symmetric, we can use the equation

$$\frac{dx}{dt} = \frac{h - x}{t - t_F}, \tag{2}$$

to calculate trajectory *AB*. In Eq. (2), the left-hand side is the slope of the leading C_+ characteristic and the right-hand side reflects the mere fact that the C_- rarefaction wave is simple and centered. Here, *h* is the target thickness and *x* is the Lagrangean coordinate, i.e., the initial (zero-pressure) distance from the surface of contact with the flyer plate. Integrating Eq. (2), we have

$$\int_0^{x_B} \frac{dx}{h - x} = \int_{t_A}^{t_B} \frac{dt}{t - t_F} \ln\left(\frac{h}{h - x_B}\right) = \ln\left(\frac{t_B - t_F}{t_A - t_F}\right) \tag{3}$$

and, eventually,

$$t_B - t_F = (t_A - t_F)\frac{h}{h - x_B}. \tag{4}$$

On the other hand,

$$t_B - t_F = (h - x_B)/c_0. \tag{5}$$

Eliminating $(h - x_B)$ from Eqs. (4) and (5) yields

$$t_B - t_F = \sqrt{(t_A - t_F)/c_0}\,h. \tag{6}$$

Time interval $t_A - t_F = h/a$ is determined in quasi-acoustic approximation (1) for the sound velocity *a* in the shock-compressed material in Lagrangean coordinates. The wave reverberation time is $t_D - t_F = 1(t_B - t_F)$.

The estimates of the front reverberation time for the elastic and plastic waves are shown in Fig. 4 (dashed lines). To calculate the reverberation time of the elastic wave, we used the coefficient values $c_0 = 2.99$ km/s and $b = 1.933$, which were determined from the longitudi-
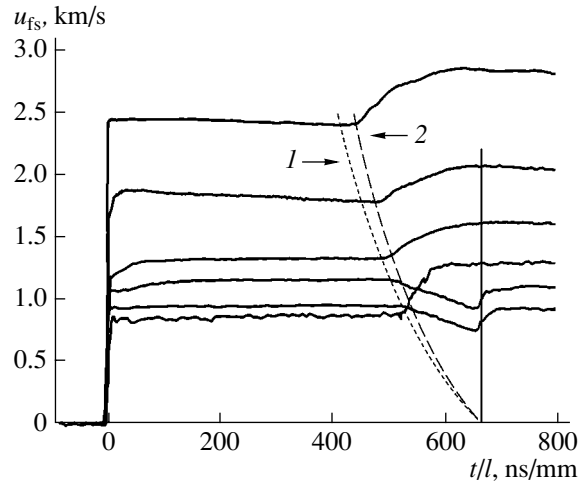
Fig. 4. Free-surface velocity profiles for the ⟨001⟩ zinc specimens. The time is normalized by dividing by the specimen thickness. The origin corresponds to the instant the shock wave front reaches the specimen surface. (*1*, *2*) Reverberation time curves estimated for the elastic- and plastic-wave fronts, respectively.
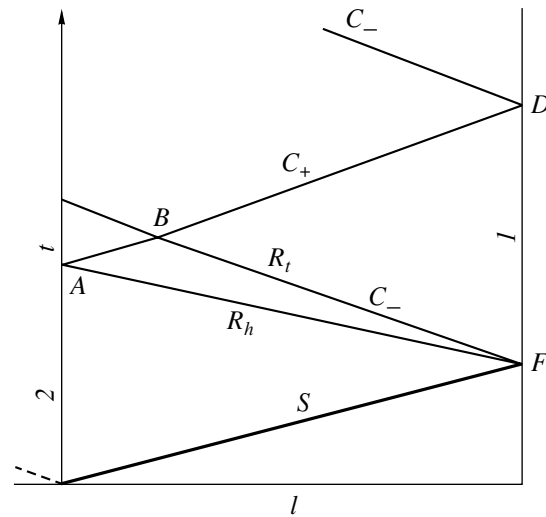

Fig. 5. Distance–time (*l*, *t*) diagram accounting for wave interactions when the waves reflect from (*1*) the free target surface and (*2*) the interface with the flyer plate or base plate. Substantial Lagrangean coordinates are used; therefore, the boundaries in the diagram are fixed.

nal sound velocity measured for compression/tensile stresses ranging from –2 to 13 GPa [11]. When estimating the reverberation time of the plastic wave, we used the coefficients appearing in the linear expression for the Hugoniot adiabat of zinc: $c_0 = 3.0$ km/s and $b = 1.57$. Figure 4 shows that, at compression stresses below 16 GPa, the actual reverberation time agrees well with the assumption of elastic response of the material to the load. However, in higher pressure experiments, the reverberation time is close to estimates for the plastic waves propagating with a volume sound velocity.
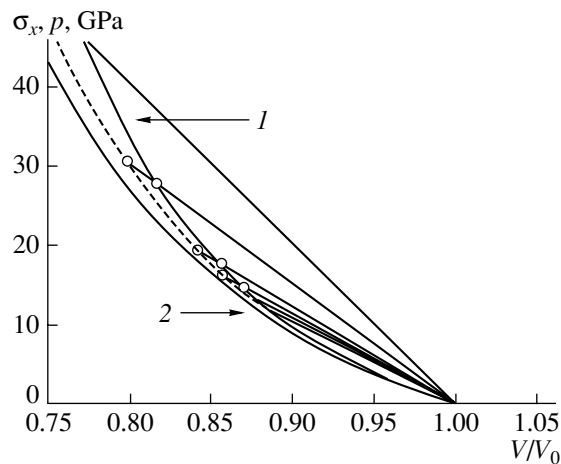
Fig. 6. Zinc state variation in shock waves of different intensity. The symbols show the final states and the states corresponding to the onset of stress relaxation. (*1*) Isentrope for longitudinal elastic compression and (*2*) isentrope for uniform compression.

Thus, at a compression stress of $\approx$14–15 GPa behind the shock front, the response of the material to the load changes qualitatively: a region of parameter relaxation appears behind the shock, the shape of the rereflected compression waves changes, and the dependence of the wave reverberation time on the shock compression pressure changes form. These findings indicate switching from purely elastic to elastic–plastic deformation in this pressure range without loss of shock wave stability.

In stationary compression waves, all intermediate states and the final state must correlate with the Rayleigh (Michelson) line,

$$\sigma_x = \rho_0 U_s^2 \frac{V_0 - V}{V_0},$$

where V_0 and V are the initial and current specific volumes of the material, respectively.

The positions of the Rayleigh lines relative to the Hugoniot adiabat of polycrystalline zinc [6–10] and the adiabat of elastic compression in the $\langle 001 \rangle$ direction [11] are shown in Fig. 6. Because of the lack of relevant data, the Hugoniot adiabat of polycrystalline zinc obtained over a wide pressure range is considered here as the adiabat of uniform compression. Marked in the Rayleigh lines are the final states and the states corresponding to the onset of stress relaxation. At an elastic compression stress of 14–15 GPa (when plastic deformation begins), the deviation of the adiabat of elastic longitudinal compression from the Hugoniot adiabat of zinc is $\approx$1.5 GPa. This value virtually coincides with the deviative stresses at the front of the elastic compression

precursor propagating in the transverse direction [3]. The final states of shock compression deviate from the Hugoniot adiabat of polycrystalline zinc by no more than 1.2–2.0 GPa.

REFERENCES

1. Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Nauka, Moscow, 1966; Academic, New York, 1967).

2. G. I. Kanel, S. V. Razorenov, A. V. Utkin, and V. E. Fortov, *Shock-Wave Phenomena in Condensed Matter* (Yanus-K, Moscow, 1996) [in Russian].

3. A. A. Bogach, G. I. Kanel, and S. V. Razorenov, Fiz. Tverd. Tela (St. Petersburg) **40**, 1849 (1998) [Phys. Solid State **40**, 1676 (1998)].

4. H. M. Ledbetter, J. Phys. Chem. Ref. Data **6**, 1181 (1977).

5. A. V. Antonov, S. V. Kopetskiǐ, L. S. Tsvindlerman, and V. G. Sursaeva, Dokl. Akad. Nauk SSSR **213**, 318 (1974) [Sov. Phys. Dokl. **18**, 736 (1974)].

6. J. M. Walsh, M. H. Rice, R. G. McQueen, *et al.*, Phys. Rev. **108**, 196 (1957).

7. L. V. Al'tshuler, A. A. Bakanova, and R. F. Trunin, Zh. Éksp. Teor. Fiz. **42**, 91 (1962) [Sov. Phys. JETP **15**, 65 (1962)].

8. R. G. McQueen and S. P. Marsh, J. Appl. Phys. **31**, 1253 (1960).

9. K. V. Volkov and V. A. Sibelev, Prikl. Mekh. Tekh. Fiz., No. 1, 125 (1984).

10. *LASL Shock Hugoniot Data*, Ed. by S. P. Marsh (Univ. California Press, Berkeley, 1980).

11. G. S. Bezruchko, G. I. Kanel, and S. V. Razorenov, Teplofiz. Vys. Temp. **42** (2), 1 (2004).

12. L. M. Barker and R. E. Hollenbach, J. Appl. Phys. **45**, 4872 (1974).

13. J. W. Swegle and D. E. Grady, J. Appl. Phys. **58**, 692 (1985).

Translated by K. Shakhlevich

OPTICS,
QUANTUM ELECTRONICS

Optical Energy Accumulation by a System Comprising an Optical Diode, Insulating Layer, and Mirror

A. A. Gevorgyan

Yerevan State University, ul. A. Manukyana 1, Yerevan, 375025 Armenia
e-mail: agevorgyan@ysu.am
Received March 24, 2004

Abstract—The optical properties of a multilayer system comprising an optical diode, insulating layer, and mirror are studied. The distribution of the electromagnetic wave energy in the system is considered. It is shown that the energy of the wave is accumulated within certain spectral intervals. Two mechanisms underlying this observation are discussed. One is due to the diode effect (i.e., nonreciprocal transmission or reflection); the other is associated with periodic layers existing in the system. Multilayer structures based on the phenomenon discovered may find application in water or gas heaters and in optical-to-electric energy converters. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

Layered periodic structures have recently attracted much attention because of the progressively expanding potentialities of epitaxial technology for creating periodic and/or aperiodic multilayer structures. Such structures are emerging artificial materials that offer specially tailored physical properties unattainable in natural insulators, metals, and semiconductors. The properties of these artificial materials depend not only on the physical parameters of the constituents but also on the geometry of the layers and the period of the structure. Structures of this sort are being widely used in advanced optics; optoelectronics; antennas; and laser, X-ray, and millimeter- and submillimeter-wave technologies.

In studies of wave propagation in such systems, the emphasis is usually on the fields of the reflected and transmitted waves, while the field inside the system remains undetermined. However, in many cases of physical interest, it is necessary to know the distribution of the field both outside and inside the system. Examples are radiation propagation in waveguides, as well as in inhomogeneous and multilayer media; optical absorption in periodic systems; etc. [1–8]. For example, in [6–8], where radiation absorption anomalies under diffraction conditions were studied, a new mechanism causing these anomalies was found that has to do with the optical field (energy) distribution just inside the system. It was shown, in particular, that, at certain wavelengths, the optical energy density inside the system may by several times exceed that outside it; that is, the optical energy is accumulated by the system. In [9], optical energy accumulation in layered periodic systems was investigated.

It was shown [10–13] that nonreciprocal optical systems may act as optical diodes under certain conditions;

i.e., they transmit light incident only on one side, making light energy accumulators feasible. This work is concerned just with this problem.

Experiments were carried out with a system consisting of an optical diode (OD) and a mirror with an isotropic insulating layer (IIL) in between. The optical diode represents a cholesteric-filled Fabry–Perot resonator. Two configurations of the system were used: OD(1)–IIL–mirror (Fig. 1) and OD(2)–IIL–mirror, where OD(1) = glass(1)–cholesteric–glass(2) and OD(2) = glass(2)–cholesteric–glass(1).

THEORY OF THE OPTICAL DIODE–INSULATING LAYER–MIRROR SYSTEM

Consider the propagation of light through an OD–IL–mirror system shown in Fig. 1. Wave propagation in inhomogeneous and multilayer media remains a challenging problem of physics and is a subject of extensive
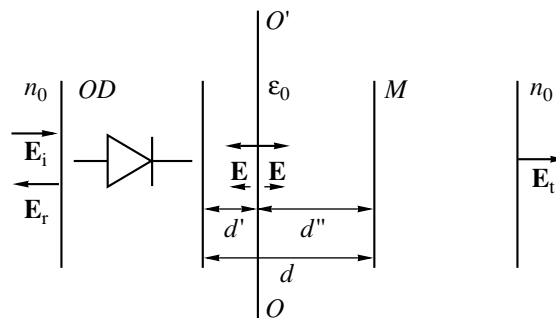

Fig. 1. Configuration of the multilayer optical system. OD(1) = glass(1)–cholesteric–glass(2), OD(2) = glass(2)–cholesteric–glass(1). ε_0 is the permittivity of the isotropic insulator, d is the thickness of the insulator, and M stands for mirror.

research (see, e.g., [14–22]). It is convenient to tackle this problem with the simple yet efficient Ambartsumyan method of addition of layers [20–22]. In this method, extra matrices $\hat{S}$ and $\hat{P}$ are introduced in order to exclude the solution of new equations that take into account problem-complicating factors: insertion of extra layers between multilayer systems or within a multilayer system, insertion of radiating planes (layers), or determination of inner fields in the optical system instead of finding transmission and/or reflection characteristics.

Let wave $\mathbf{E}_i$ incident on an OD–IIL–mirror system generate waves $\mathbf{E}_r$ and $\mathbf{E}_t$ transmitted through and reflected from the system. The complex amplitudes of the incident, reflected, and transmitted waves are expanded in circular basis polarizations,

$$\mathbf{E}_{i, r, t} = E_{i, r, t}^{+}\mathbf{n}_{+} + E_{i, r, t}^{-}\mathbf{n}_{-} = \begin{bmatrix} E_{i, r, t}^{+} \\ E_{i, r, t}^{-} \end{bmatrix}, \qquad (1)$$

where $\mathbf{n}_{+}$ and $\mathbf{n}_{-}$ are the unit vectors of circular basis polarizations.

The reflected and transmitted waves are related to the incident one as

$$\mathbf{E}_r = \hat{R}\mathbf{E}_i, \quad \mathbf{E}_t = \hat{T}\mathbf{E}_i, \qquad (2)$$

where $\hat{R}$ and $\hat{T}$ are the Jones matrices for the given system.

According to [22], if there is a system consisting of adjacent subsystems A (left) and B (right), reflection matrix $\hat{R}_{A+B}$ and transmission matrix $\hat{T}_{A+B}$ of system A + B are expressed through the respective matrices of the subsystems as

$$\hat{R}_{A+B} = \hat{R}_A + \hat{T}_A\hat{S}\hat{T}_A,$$
$$\hat{T}_{A+B} = \hat{T}_B\hat{P}\hat{T}_A, \qquad (3)$$

provided that light strikes system A + B from the left. It is assumed that subsystems A and B are, respectively, on the left and right of joining plane OO' (Fig. 1). Hence,

$$\hat{T}_A = \hat{T}_1\exp(ik_0d'), \quad \hat{R}_A = \hat{R}_1,$$

$$\tilde{\hat{T}}_A = \tilde{\hat{T}}_1\exp(ik_0d'), \quad \tilde{\hat{R}}_A = \exp(ik_0d')\tilde{\hat{R}}_1\exp(ik_0d'), \quad (4)$$

$$\hat{T}_B = \hat{T}_2\exp(ik_0d''), \quad \hat{R}_B = \exp(ik_0d'')\hat{R}_2\exp(ik_0d''),$$

where $\hat{T}_1$, $\hat{R}_1$ and $\hat{T}_2$, $\hat{R}_2$ are the transmission and reflection matrices of the OD and mirror;

$$k_0 = \frac{2\pi}{\lambda}\sqrt{\varepsilon_0};$$

ε_0 is the permittivity of the isotropic insulator, λ is the wavelength in a vacuum; and d' and d'' are the distances

from the joining plane to boundaries OO' of the left- and right-hand isotropic insulating layers, respectively. The quantities with a tilde designate the transmission/reflection characteristics of the system when light strikes the system from the right.

Matrices $\hat{S}$ and $\hat{P}$ describe resulting waves in the insulating layer on joining plane OO'. Thus,

$$\mathbf{E}_{\rightarrow} = \hat{P}\hat{T}_A\mathbf{E}_i \qquad (5)$$

is a wave arising on plane OO' and propagating to the right and

$$\mathbf{E}_{\leftarrow} = \hat{S}\hat{T}_A\mathbf{E}_i \qquad (6)$$

is a wave arising on plane OO' and propagating to the left. The total field on this plane in the insulator is therefore given by

$$\mathbf{E}_{\text{total}} = (\hat{S} + \hat{P})\hat{T}_A\mathbf{E}_i. \qquad (7)$$

According to [22], matrices $\hat{S}$ and $\hat{P}$ satisfy the set of equations

$$\hat{S} = \hat{R}_B[\hat{I} - \tilde{\hat{R}}_A\hat{R}_B]^{-1},$$
$$\hat{P} = [\hat{I} - \tilde{\hat{R}}_A\hat{R}_B]^{-1}. \qquad (8)$$

Since OD(1) and OD(2) themselves are multilayer systems, their respective transmission and reflection matrices ($\hat{T}_1$, $\hat{R}_1$ and $\hat{T}_2$, $\hat{R}_2$) can be found by the same technique (see also [23]).

Using (1)–(8), one can calculate the coefficients of reflection, $R = |E_r|^2/|E_i|^2$, and transmission, $T = |E_t|^2/|E_i|^2$; the relative intensity of the wave resulting on the joining plane, $I = |E_{\text{total}}|^2/|E_i|^2$; and other optical characteristics of the system.

NUMERICAL CALCULATION AND CONCLUSIONS

To study optical energy accumulation by a system containing a periodic layer (several layers), we will first consider a system comprising a cholesteric layer and a mirror in which the joining plane coincides with the interface. Figure 2a plots the wavelength dependence of intensity I of the wave resulting on the joining plane for different polarizations of the incident wave. It is seen that the joining plane does not accumulate optical energy ($I < 1$, incident light intensity $I_0 = 1$). This, however, does not mean that the accumulation is absent at all. Calculations show that the accumulation to an extent takes place inside the cholesteric. Now we complicate the system, introducing an isotropic insulating layer between the cholesteric and mirror. Figure 2b demonstrates the wavelength dependence of intensity I of the wave resulting on the joining plane in this case (here, joining plane OO' is in the middle of the isotropic insulator). At certain wavelengths of the incident light,
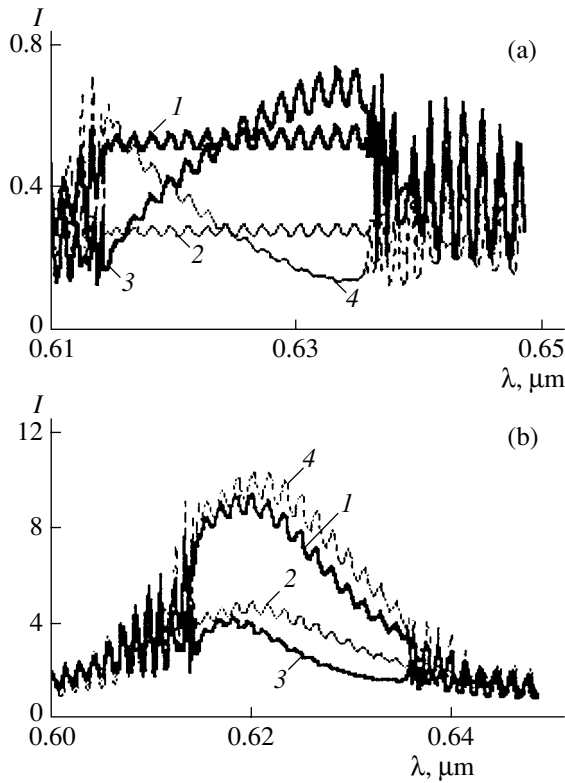
Fig. 2. Intensity I of the net wave at the IIL–mirror interface vs. the wavelength for the (a) cholestric–mirror system and (b) cholestric–IIL–mirror system when the incident wave is polarized (*1*) clockwise, (*2*) counterclockwise, (*3*) linearly along the x direction, and (*4*) linearly along the y direction. The cholesteric has a right-hand helix.

the resulting wave intensity in the system far exceeds the incident intensity (for example, by a factor of 10.5 at $\lambda = 0.6218$ μm); that is, the optical energy is compressed (accumulated). Thus, a periodic layer inserted into the system makes it possible to accumulate optical energy under certain conditions.

Let us now turn to an optical system containing a nonreciprocal element to demonstrate another mechanism of optical energy accumulation. The system comprises an OD and a mirror. Two structures of the optical diode are considered: OD(1) = glass(1)–cholesteric–glass(2) and OD(2) = glass(2)–cholesteric–glass(1). The joining plane coincides with the OD–mirror interface. Figures 3a and 3b plot the wavelength dependence of intensity I of the wave resulting on the joining plane when the light is incident on the (a) OD(1)–mirror and (b) OD(2)–mirror systems. In this case, the joining plane does not accumulate the energy again. Note that, in this case, too, this does not indicate the absence of the accumulation.

As before, we complicate the system by inserting an isotropic insulating layer between the OD and mirror. Figures 4a and 4b plot the wavelength dependence of intensity I of the wave resulting on the joining plane when the light is incident on the (a) OD(1)–IIL–mirror
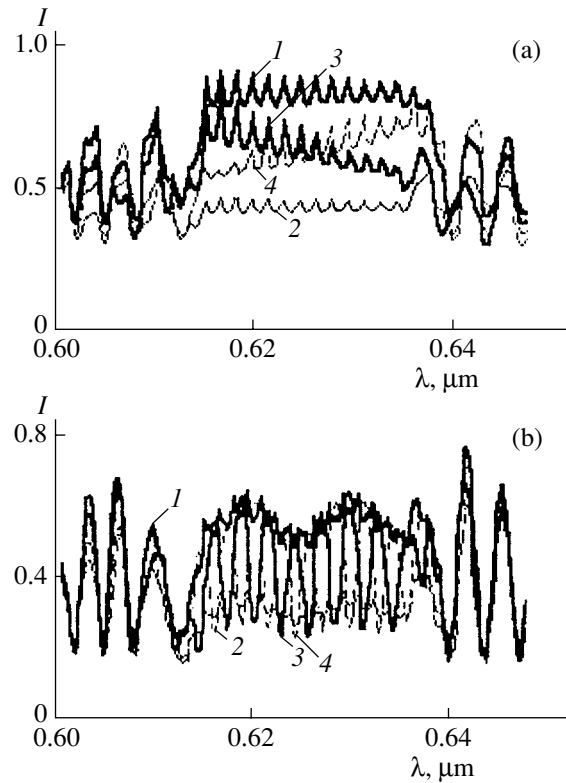
Fig. 3. Intensity I of the net wave at the OD–mirror interface vs. the wavelength for the (a) OD(1)–mirror system and (b) OD(2)–mirror system. For *1–4*, see Fig. 2.
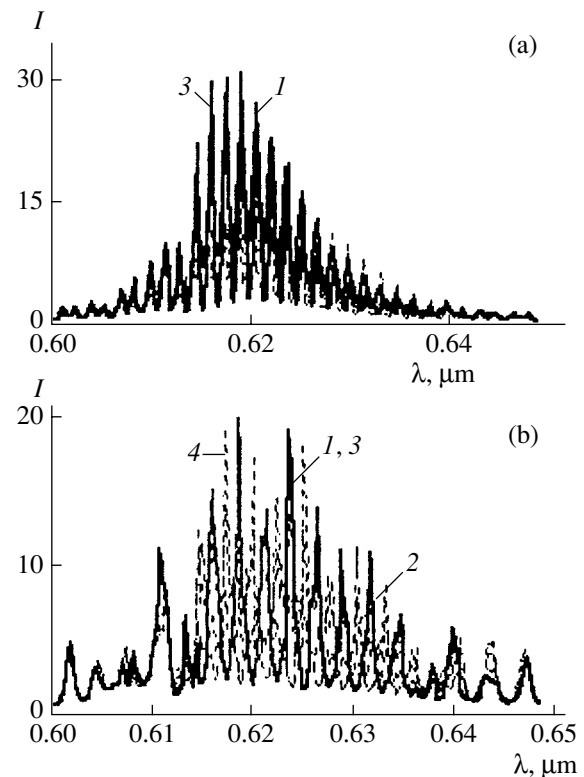
Fig. 4. Intensity I of the net wave arising in the IIL at the joining plane vs. the wavelength for the (a) OD(1)–IIL–mirror system and (b) OD(2)–IIL–mirror system. For *1–4*, see Fig. 2.
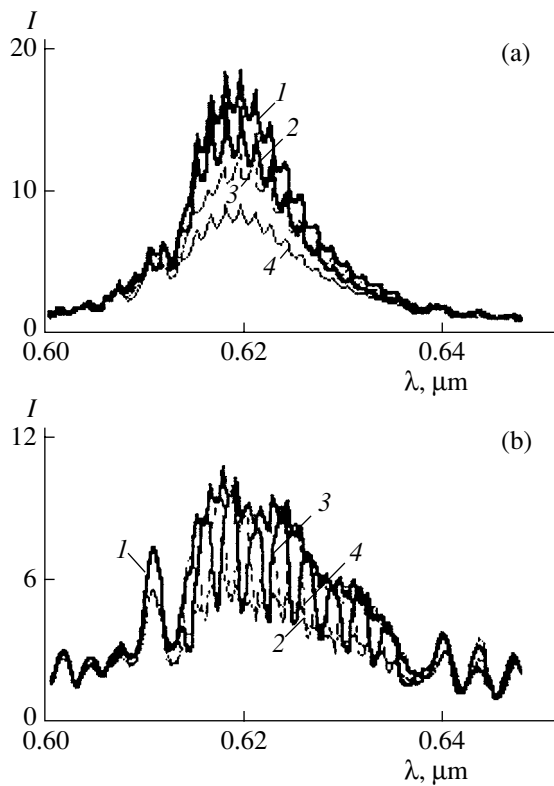
Fig. 5. Wavelength-averaged intensity I of the net wave arising in the IIL vs. the wavelength for the (a) OD(1)–IIL–mirror system and (b) OD(2)–IIL–mirror system. For 1–4, see Fig. 2.

Fig. 6. Intensity I of the net wave arising in the IIL vs. d' for the (a) OD(1)–IIL–mirror system and (b) OD(2)–IIL–mirror system. For 1–4, see Fig. 2.

and (b) OD(2)–IIL–mirror systems. From Fig. 4, it follows that (i) in the first case (in the OD(1)–IIL–mirror system), the energy compression is much higher than in the other system and (ii) the value of I reaches 30 at certain wavelengths and may be raised further.

The above observations can be explained in simple terms. In the former case, the optical energy is accumulated via two mechanisms: the first one (diode effect) is due to the asymmetric (nonreciprocal) transmission of the system's elements; the other is due to the presence of the periodic layer in the system (we mentioned previously that the periodic structures may accumulate optical energy in certain wavelength intervals). In the latter case, only the second mechanism is accomplished.

Consider now possible applications of the effect of optical energy accumulation. Note that the intensity of the resulting wave in the periodic system is usually high in narrow frequency intervals and when the system is thick. However, an inevitably existing spread in the optical system thickness averages the intensity over both the thickness and the wavelength; as a result, the intensity decreases significantly. Figures 5a and 5b show the wavelength dependences of the wavelength-averaged intensities for the (a) first and (b) second cases. It is worth noting that, generally, a finite-thick-
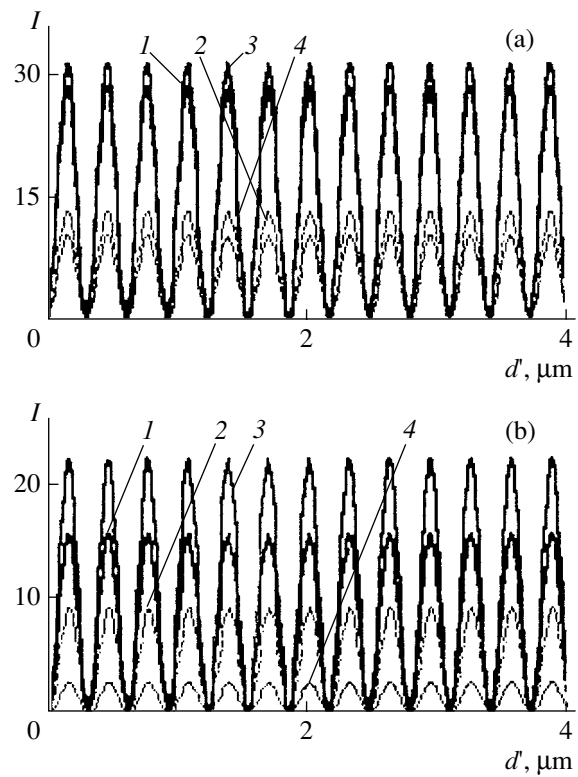
ness layer of an isotropic insulator (semiconductor, or another material) must be inserted into the system in order to extract or utilize the energy stored. That is, the possibility of light energy compression using optical systems with an IIL (the case under consideration in this work) is of great practical interest.

Figures 6a and 6b present the d' dependences of the intensity for the two cases.

Thus, our experimental and analytical data demonstrate that an OD–IIL–mirror system may efficiently and reliably compress the light energy in the insulating layer, the spectral interval of compression being rather wide and insensitive to the layer thickness spread.

Note in conclusion the possible applications of the systems. If the light energy compression is maximal in a spectral range of maximal absorption (e.g., of water), such systems may be used in heat engineering as efficient water heaters. Similarly, they can operate as gas heaters. Also, such systems may be used in optical-to-electric energy converters. Of course, other fields of application may be indicated.

Significantly, the numerical results reported in this work were obtained in a real experiment for a system containing (i) a cholesteric layer of composition cholesteryl nonanoate : cholesteryl chloride : cholesteryl acetate = 20 : 15 : 6, which is $d = 100\sigma$ thick and exhibits a pitch of helix in the optical range ($\sigma = 0.42$ μm) at

room temperature (24°C); (ii) a silver layer ($d = 2$ μm, $n = 1.44 + i3.631$) as a mirror; and (iii) a glass sheet ($d_1 = 50$ μm, $n_1 = 1.5 + i10^{-8}$) as glass(1) or a diamond sheet ($d_2 = 50$ μm, $n_2 = 2.417 + i10^{-8}$) as glass(2). The accuracy of the numerical data was estimated using the energy conservation law (in the absence of absorption, $R + T = 1$) or by comparing current data with those obtained previously.

REFERENCES

1. A. Yariv and P. Yeh, *Optical Waves in Crystals* (Wiley, New York, 1984).
2. P. Yeh, *Optical Waves in Layered Media* (Wiley, New York, 1988).
3. F. Ramos-Mendieta and P. Halevi, J. Opt. Soc. Am. B **14**, 370 (1997).
4. F. Villa, L. E. Regalado, *et al.*, Opt. Lett. **27**, 646 (2002).
5. V. V. Efimov and D. I. Sementsov, Opt. Spektrosk. **77**, 72 (1994) [Opt. Spectrosc. **77**, 61 (1994)].
6. A. H. Gevorgyan, Mol. Cryst. Liq. Cryst. **378**, 187 (2002).
7. A. A. Gevorgyan, Izv. NAN Armenii, Phys. **38**, 366 (2003).
8. A. A. Gevorgyan, Izv. NAN Armenii, Phys. **39** (2004).
9. A. A. Gevorgyan, Izv. NAN Armenii, Phys. **39**, 225 (2004).
10. A. A. Gevorgyan, Zh. Tekh. Fiz. **72** (8), 77 (2002) [Tech. Phys. **47**, 1008 (2002)].
11. A. H. Gevorgyan, Mol. Cryst. Liq. Cryst. **382**, 1 (2002).
12. A. A. Gevorgyan, Izv. NAN Armenii, Phys. **37**, 155 (2002).
13. A. A. Gevorgyan, Pis'ma Zh. Tekh. Fiz. **29** (19), 60 (2003) [Tech. Phys. Lett. **29**, 819 (2003)].
14. F. Abeles, Annales de Physique **5**, 596 (1950); **5**, 706 (1950).
15. D. W. Berreman, J. Opt. Soc. Am. **203**, 385 (1974).
16. L. M. Brekhovskikh, *Waves in Layered Media* (Nauka, Moscow, 1973; Academic, New York, 1980).
17. V. I. Klyatskin, *Stochastic Equations and Waves in Randomly Inhomogeneous Media* (Nauka, Moscow, 1980) [in Russian].
18. V. I. Klyatskin, *The Method of Embedding in the Wave Propagation Theory* (Nauka, Moscow, 1980) [in Russian].
19. D. M. Sedrakian, A. H. Gevorgyan, and A. Zh. Khachatrian, Opt. Commun. **192**, 135 (2001).
20. V. A. Ambartsumyan, Izv. Akad. Nauk Arm. SSR, Estestv. Nauki **1–2**, 31 (1944).
21. O. V. Pikichyan, Dokl. Akad. Nauk SSSR **263**, 601 (1982) [Sov. Phys. Dokl. **28**, 1033 (1983)].
22. A. A. Gevorgyan, K. V. Papoyan, and O. V. Pikichyan, Opt. Spektrosk. **88**, 647 (2000) [Opt. Spectrosc. **88**, 586 (2000)].
23. A. A. Gevorgyan, Zh. Tekh. Fiz. **70** (9), 75 (2000) [Tech. Phys. **45**, 1170 (2000)].

Translated by V. Isaakyan

OPTICS,
QUANTUM ELECTRONICS

Covariant Dispersion Relations and Evolution Tensor Operators for Optical Waveguides

A. V. Novitsky and L. M. Barkovsky

Belarussian State University, ul. Skaryna 4, Minsk, 220080 Belarus
e-mail: Barkovsky@bsu.by
Received December 30, 2003

Abstract—The operation approach is applied to solve the Maxwell equations for a sequence of circular layers the permittivity and permeability tensors of which depend on the radial coordinate. This makes it possible to extend the well-known method of stratification for cylindrical structures. Operator dispersion relations for graded-index and multilayer step-index isotropic circular fibers are derived. Numerical solutions to the dispersion relation for a multilayer waveguide with a periodically varying permittivity of the layers are obtained. © 2005 Pleiades Publishing, Inc.

INTRODUCTION

The influence of polarization on the performance of fiber-optic communications has recently been given much attention. Among polarization effects, the polarization–mode dispersion, which is attributed to a rapidly and randomly varying birefringence in the fiber, and polarization losses are the most important. The former effect causes a random change in the polarization of light, which greatly affects the signal at the exit from the fiber [1, 2]. In a number of works, the performance of fiber-optic gyros [3, 4] and lasers [5, 6] was extensively studied. A fiber-optic ring-interferometer gyro configured with a high-linear-birefringence light-guiding loop is an example [4]. Such gyros offer a high sensitivity and may be used in navigation. In [6], two lasing conditions for a ring bidirectional fiber laser with a Faraday phase shifter are treated theoretically. Of interest are techniques for determining the refractive index of a liquid from optical losses in a curved cylindrical light guide [7].

Operator methods as applied to the optics of complex media are being extensively developed today. The light reflection and transmission tensors [8, 9], the tensor refractive index [10], and the velocity operator of electromagnetic waves [11] generalize related scalar quantities with regard to the vector nature of light (photon spin) [12]. These operators describe the superposition of eigenmodes propagating with certain velocities and polarizations. In terms of the covariant formalism developed by Fedorov [13, 14], the Maxwell equations and dispersion relations are represented in a compact form suitable for analysis and calculation of layered bianisotropic structures. Here, the evolution operator and the impedance tensor are of critical importance. In [15], both are applied to attack the problems of waveguide propagation, reflection, and transmission of light. The mathematical structure of the vector equa-

tions for the field implies that the evolution operator is a block matrix. For planar waveguides, operator dispersion relations have been derived [9], which involve the impedance tensors for the cladding and the evolution operator for the core. The extension of the impedance matrix for the case of spherical layered anisotropic media, which simulate the terrestrial ionosphere, was reported [16].

In this work, we apply the operator method to determine modes in circular fibers. The electromagnetic field is described with sets of first-order differential equations. These sets are valid for those media the permittivity and permeability tensors of which are radially nonuniform. Such an approach (also called the method of stratification [18]) is, in essence, a multilayer step approximation and has been long used in analysis of planar waveguides [17]. In [9], it was extended for inhomogeneous bianisotropic waveguide layers. This method of deriving dispersion relations is demonstrated with an isotropic circular waveguide. Dispersion relations for multilayer step- and graded-index fibers are studied.

EVOLUTIONARY SOLUTIONS FOR CIRCULAR CYLINDRICAL WAVEGUIDES

Consider the propagation of electromagnetic waves with fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$ harmonically varying in time. In cylindrical coordinate system (r, φ, y), the Maxwell equations have the form

$$\left(\mathbf{b}^\times \frac{\partial}{\partial y} + \mathbf{e}_r^\times \frac{\partial}{\partial r} + \frac{1}{r}\mathbf{e}_\varphi^\times \frac{\partial}{\partial \varphi}\right)\mathbf{H}(\mathbf{r}) = -ik\varepsilon(r)\mathbf{E}(\mathbf{r}),$$

$$\left(\mathbf{b}^\times \frac{\partial}{\partial y} + \mathbf{e}_r^\times \frac{\partial}{\partial r} + \frac{1}{r}\mathbf{e}_\varphi^\times \frac{\partial}{\partial \varphi}\right)\mathbf{E}(\mathbf{r}) = ik\mu(r)\mathbf{H}(\mathbf{r}),$$

$$(1)$$

where $\mathbf{e}_r(\varphi)$, $\mathbf{e}_\varphi(\varphi)$, and $\mathbf{b}$ are the basis vectors directed, respectively, along the radius, tangent to the circle, and generatrix of the cylinder; $k = \omega/c$ is the wavenumber in a vacuum; ω is the wave frequency; and $\mathbf{b}^\times$ is the tensor dual to vector $\mathbf{b}$ $((\mathbf{b}^\times)_{ik} = \varepsilon_{ijk}b_j)$ [10, 14].

In (1), we assume that the permittivity, $\varepsilon(r)$, and permeability, $\mu(r)$, tensors depend on only radial coordinate r. This is possible if ε and μ are composed of axial-coordinate-independent (φ-independent) tensors, unit tensor $\mathbf{1}$, dyad $\mathbf{b} \otimes \mathbf{b}$, and dual tensor $\mathbf{b}^\times$. Thus, we will deal with the following permittivity (permeability) tensors: (i) $\varepsilon = \varepsilon(r)\mathbf{1}$, isotropic medium; (ii) $\varepsilon = \varepsilon_1(r)(\mathbf{1} - \mathbf{b} \otimes \mathbf{b}) + \varepsilon_2(r)\mathbf{b} \otimes \mathbf{b}$, uniaxial crystal the optical axis of which is aligned with vector $\mathbf{b}$; (iii) $\varepsilon = \varepsilon_1(r)\mathbf{1} + \chi(r)\mathbf{b}^\times$, gyrotropic medium; and (iv) $\varepsilon = \varepsilon_1(r)(\mathbf{1} - \mathbf{b} \otimes \mathbf{b}) + \varepsilon_2(r)\mathbf{b} \otimes \mathbf{b} + \chi(r)\mathbf{b}^\times$.

Taking into account the invariance of the field strengths relative to coordinates y and φ, we separate the variables,

$$\begin{pmatrix} \mathbf{H}(\mathbf{r}) \\ \mathbf{E}(\mathbf{r}) \end{pmatrix} = \exp(i\beta y + i\nu\varphi)\begin{pmatrix} \mathbf{H}(r, \varphi) \\ \mathbf{E}(r, \varphi) \end{pmatrix},$$

where β is the longitudinal wavenumber (mode propagation constant) and ν is an integer.

Since components H_y, H_φ, and H_r of vector $\mathbf{H}(r, \varphi)$ (similarly to the components of $\mathbf{E}(r, \varphi)$) do not depend on φ (i.e., $\mathbf{H}(r, \varphi) = H_y(r)\mathbf{b} + H_\varphi(r)\mathbf{e}_\varphi(\varphi) + H_r(r)\mathbf{e}_r(\varphi)$), Eqs. (1) take the form

$$\mathbf{e}_r^\times\frac{d\mathbf{H}}{dr} + \left(i\beta\mathbf{b}^\times + \frac{i\nu}{r}\mathbf{e}_\varphi^\times + \frac{1}{r}\mathbf{b} \otimes \mathbf{e}_\varphi\right)\mathbf{H} = -ik\varepsilon(r)\mathbf{E},$$

$$\mathbf{e}_r^\times\frac{d\mathbf{E}}{dr} + \left(i\beta\mathbf{b}^\times + \frac{i\nu}{r}\mathbf{e}_\varphi^\times + \frac{1}{r}\mathbf{b} \otimes \mathbf{e}_\varphi\right)\mathbf{E} = ik\mu(r)\mathbf{H}.$$

(2)

Thus, the Maxwell equations are reduced to a set of ordinary first-order differential equations in r for vector functions $\mathbf{E}$ and $\mathbf{H}$. Coordinate φ enters into vectors $\mathbf{e}_\varphi$ and $\mathbf{e}_r$, which specify the polarization of the electromagnetic waves. It is seen that the independence of the field components on φ is satisfied. Set (2) can be split into four ordinary differential equations and two algebraic equations. The algebraic equations make it possible to eliminate two of six components of $\mathbf{E}$ and $\mathbf{H}$. It is appropriate to leave the field components lying in the plane tangent to the surface of the circular cylinder (hereafter, tangential components). These components are continuous at the circular cylindrical interface and can be represented in the form $\mathbf{E}_t = I\mathbf{E}$ and $\mathbf{H}_t = I\mathbf{H}$, where $I = \mathbf{1} - \mathbf{e}_r \otimes \mathbf{e}_r = -\mathbf{e}_r^{\times 2}$ is the operator projecting onto the plane perpendicular to vector $\mathbf{e}_r$. Introducing vector $\mathbf{u} = (\beta/k)\mathbf{e}_\varphi - \nu/(kr)\mathbf{b}$ and considering the relationships $\mathbf{uH} = \mathbf{e}_r\varepsilon\mathbf{E}$ and $\mathbf{uE} = -\mathbf{e}_r\mu\mathbf{H}$, which follow from (2), we come to a relationship between the total

fields and their tangential components [8],

$$\begin{pmatrix} \mathbf{H}(r) \\ \mathbf{E}(r) \end{pmatrix} = V\begin{pmatrix} \mathbf{H}_t(r) \\ \mathbf{E}_t(r) \end{pmatrix},$$

(3)

where

$$V = \begin{pmatrix} \mathbf{1} - \mathbf{e}_r \otimes \mathbf{e}_r\mu/\mu_r & -\mathbf{e}_r \otimes \mathbf{u}/\mu_r \\ \mathbf{e}_r \otimes \mathbf{u}/\varepsilon_r & \mathbf{1} - \mathbf{e}_r \otimes \mathbf{e}_r\varepsilon/\varepsilon_r \end{pmatrix}$$

is the restoration matrix, $\varepsilon_r = \mathbf{e}_r\varepsilon\mathbf{e}_r$, and $\mu_r = \mathbf{e}_r\mu\mathbf{e}_r$.

In view of (3), we derive a set of equations for $\mathbf{E}_t$ and $\mathbf{H}_t$ from (2),

$$\frac{d}{dr}\begin{pmatrix} \mathbf{H}_t(r) \\ \mathbf{E}_t(r) \end{pmatrix} = ikM(r)\begin{pmatrix} \mathbf{H}_t(r) \\ \mathbf{E}_t(r) \end{pmatrix}, \quad M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, (4)$$

where

$$A = \frac{i}{kr}\mathbf{e}_\varphi \otimes \mathbf{e}_\varphi + \frac{1}{\varepsilon_r}\mathbf{e}_r^\times\varepsilon\mathbf{e}_r \otimes \mathbf{u} - \frac{1}{\mu_r}\mathbf{e}_r^\times\mathbf{u} \otimes \mathbf{e}_r\mu I,$$

$$B = -\frac{1}{\mu_r}\mathbf{e}_r^\times\mathbf{u} \otimes \mathbf{u} + \frac{1}{\varepsilon_r}I\tilde{\bar{\varepsilon}}\mathbf{e}_r^\times,$$

$$C = \frac{1}{\varepsilon_r}\mathbf{e}_r^\times\mathbf{u} \otimes \mathbf{u} - \frac{1}{\mu_r}I\tilde{\bar{\mu}}\mathbf{e}_r^\times,$$

(5)

$$D = \frac{i}{kr}\mathbf{e}_\varphi \otimes \mathbf{e}_\varphi + \frac{1}{\mu_r}\mathbf{e}_r^\times\mu\mathbf{e}_r \otimes \mathbf{u} - \frac{1}{\varepsilon_r}\mathbf{e}_r^\times\mathbf{u} \otimes \mathbf{e}_r\varepsilon I.$$

In (4) and (5), M is the block matrix in which blocks A, B, C, and D are planar tensors (planar tensor A meets the relationship $A\mathbf{e}_r = \mathbf{e}_rA = 0$), $\tilde{\bar{\varepsilon}}$ is the tensor adjoint to transposed tensor $\tilde{\varepsilon}$ (adjoint tensor $\bar{\alpha}$ is defined by the formula $\bar{\alpha}\,\alpha = \alpha\bar{\alpha} = |\alpha|$, where $|\alpha|$ is the determinant of α) [10, 14]. Relationships (5) turn into similar relationships for a planar-layer medium at $r \longrightarrow \infty$, $\mathbf{e}_r \longrightarrow \mathbf{q}$, and $\mathbf{e}_\varphi \longrightarrow \mathbf{a} = \mathbf{b} \times \mathbf{q}$, where $\mathbf{q}$ is the unit vector normal to the planar layers.

The fundamental solution to Eq. (4) is given by [8]

$$\begin{pmatrix} \mathbf{H}_t(r) \\ \mathbf{E}_t(r) \end{pmatrix} = \Omega_a^r[ikM(r)]\begin{pmatrix} \mathbf{H}_t(a) \\ \mathbf{E}_t(a) \end{pmatrix},$$

(6)

$$\Omega_a^r[ikM(r)] = \int_a^r (E + ikM(r)dr), \quad E = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix},$$

where $\Omega_a^r[ikM(r)]$ $(a \neq 0)$ is the evolution operator, which is represented through the multiplicative integral [19].

Relationship (6) implies that, knowing initial vectors $\mathbf{H}_t(a)$ and $\mathbf{E}_t(a)$, we can find the tangential components of the fields at any point r. Let us introduce impedance tensor $\Gamma(r)$ and define it as a quantity relat-

ing the tangential components of the magnetic and electric fields, $\mathbf{E}_t = \Gamma\mathbf{H}_t$. Eliminating vectors $\mathbf{H}_t$ and $\mathbf{E}_t$ from (4), we arrive at the Riccati tensor equation for Γ [9]

$$\frac{1}{ik}\frac{d\Gamma}{dr} + \Gamma B\Gamma + \Gamma A - D\Gamma - C = 0. \qquad (7)$$

The solution of Eq. (7) yields two impedance tensors, $\Gamma_1(r)$ and $\Gamma_2(r)$. Thus, the circular layer $r \in (a, b)$, $a \neq 0$, supports the propagation of two waves, which correspond to two independent solutions and have impedances $\Gamma_1(r)$ and $\Gamma_2(r)$ The layers covering the intervals $0 \leq r \leq a$ and $b \leq r \leq \infty$ support one wave whose amplitude is finite at points $r = 0$ and $r = \infty$, respectively. For both a homogeneous and an inhomogeneous circular layer, the impedance tensor depends on r, so that the Riccati tensor equation cannot be reduced to an algebraic equation. Differential equation (7) can be solved numerically by techniques described elsewhere [8, 9].

ISOTROPIC CIRCULAR LAYER

Consider a homogeneous isotropic medium with a constant permittivity and permeability, $\varepsilon = $ const and $\mu = $ const. Then, tensors (5) are recast as

$$A = \frac{i}{kr}\mathbf{e}_\varphi \otimes \mathbf{e}_\varphi, \quad D = \frac{i}{kr}\mathbf{e}_\varphi \otimes \mathbf{e}_\varphi,$$

$$B = \frac{\beta\nu}{k^2\mu r}(\mathbf{b} \otimes \mathbf{b} - \mathbf{e}_\varphi \otimes \mathbf{e}_\varphi) + \frac{u_1}{k^2\mu}\mathbf{b} \otimes \mathbf{e}_\varphi - \frac{u_2}{k^2\mu}\mathbf{e}_\varphi \otimes \mathbf{b},$$

$$C = -\frac{\beta\nu}{k^2\varepsilon r}(\mathbf{b} \otimes \mathbf{b} - \mathbf{e}_\varphi \otimes \mathbf{e}_\varphi) - \frac{u_1}{k^2\varepsilon}\mathbf{b} \otimes \mathbf{e}_\varphi + \frac{u_2}{k^2\varepsilon}\mathbf{e}_\varphi \otimes \mathbf{b},$$

where $u_1 = k^2\varepsilon\mu - \beta^2$ and $u_2 = k^2\varepsilon\mu - \nu^2/r^2$.

A solution to this problem is given in terms of the Bessel functions as follows:

$$\begin{pmatrix} \mathbf{H}_t(r) \\ \mathbf{E}_t(r) \end{pmatrix} = \Omega_a^r \begin{pmatrix} \mathbf{H}_t(a) \\ \mathbf{E}_t(a) \end{pmatrix}, \quad \Omega_a^r = P(r)P^-(a), \quad (8)$$

where P^- is the block matrix pseudoinverse to P ($P^-P = PP^- = E$) [10, 14].

Matrix P is expressed through planar tensors α_1, α_2, β_1, and β_2 as

$$P(r) = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{pmatrix},$$

$$\alpha_m = F_{\nu\pm}^{(m)}\left(\mathbf{b} \mp \frac{\beta\nu a^2}{u_\pm^2 r}\mathbf{e}_\varphi\right) \otimes \mathbf{b} \pm \frac{ika\varepsilon}{u_\pm}F_{\nu\pm}^{(m)'}\mathbf{e}_\varphi \otimes \mathbf{e}_\varphi,$$

$$(9)$$

$$\beta_m = \mp\frac{ika\mu}{u_\pm}F_{\nu\pm}^{(m)'}\mathbf{e}_\varphi \otimes \mathbf{b}$$

$$+ F_{\nu\pm}^{(m)}\left(\mathbf{b} \mp \frac{\beta\nu a^2}{u_\pm^2 r}\mathbf{e}_\varphi\right) \otimes \mathbf{e}_\varphi; \quad m = 1, 2.$$

Here, $u_\pm^2 = \pm a^2 u_1 = \pm k^2 a^2(\varepsilon\mu - \beta^2/k^2)$, $F_{\nu+}^{(1)} = J_\nu(u_+r/a)$ is the νth-order Bessel function of the first kind, $F_{\nu+}^{(2)} = Y_\nu(u_+r/a)$ is the Bessel function of the second kind, $F_{\nu-}^{(1)} = I_\nu(u_-r/a)$ and $F_{\nu-}^{(2)} = K_\nu(u_-r/a)$ are the modified Bessel functions, and $F_{\nu\pm}^{(m)'}(x) = dF_{\nu\pm}^{(m)}(x)/dx$ means differentiation with respect to argument. Superscripts 1 and 2 correspond to independent solutions of the Bessel equation.

For each of the waves corresponding to independent solutions to the Bessel equation, the impedance tensor is given by

$$\Gamma_m(r) = \beta_m(r)\alpha_m^-(r), \qquad (10)$$

where α_m^- is the tensor pseudoinverse to tensor α_m.

Given the initial amplitudes of waves $\mathbf{H}_{t1}(a)$ and $\mathbf{H}_{t2}(a)$, the evolution of the fields is described by the relationships

$$\mathbf{H}_{tm}(r) = \alpha_m(r)\alpha_m^-(a)\mathbf{H}_{tm}(a),$$

$$\mathbf{E}_{tm}(r) = \beta_m(r)\alpha_m^-(a)\mathbf{H}_{tm}(a). \qquad (11)$$

Evolution operator (8) of a homogeneous isotropic layer equals the product of block matrices $P(r)$ and $P^-(a)$ and is also a block matrix. Tensors α_1, α_2, β_1, and β_2 are defined by cylindrical functions $F_{\nu\pm}^{(1, 2)}$ and mode propagation constants β and ν. The layer $a \leq r \leq b$ ($a \neq 0$) supports the propagation of two waves, the field strengths of which are given by (11). In the layer $0 \leq r \leq a$ (and also $b \leq r \leq \infty$), one of the independent waves does not meet the conditions that are imposed on $\mathbf{H}$ and $\mathbf{E}$ at point $r = 0$ ($r = \infty$). Thus, only one wave can propagate in the cladding and core of the waveguide.

DISPERSION RELATIONS AND MODE POLARIZATION IN ISOTROPIC FIBERS

Now we will apply our analytical results to the propagation of electromagnetic waves in circular isotropic infinite-cladding fibers. The objects of consideration are multilayer step-index fibers and graded-index fibers with an inhomogeneous core. Our aim is to derive dispersion relations and find the polarization of waveguide modes. For a multilayer step-index fiber with radius a and constant permittivities and permeabilities of the core (ε_{co}, μ_{co}) and cladding (ε_{cl}, μ_{cl}), solutions to the Bessel equation are routinely taken in the form

$$F_\nu = \begin{cases} F_{\nu+}^{(1)} = J_\nu(u_+r/a), & r < a \\ F_{\nu-}^{(2)} = K_\nu(u_-r/a), & r \geq a. \end{cases}$$

Such a form of the solutions provides field oscillation in the core and field decay in the cladding. Then, impedance tensors (10) at the boundary $r = a$ (surface impedance tensors) can be written as

$$\Gamma_{co} = -\frac{iu}{ka\varepsilon_{co}}\frac{J_\nu(u)}{J_\nu'(u)}\left(\mathbf{b} - \frac{\beta\nu a}{u^2}\mathbf{e}_\varphi\right) \otimes \left(\mathbf{e}_\varphi + \frac{\beta\nu a}{u^2}\mathbf{b}\right)$$

$$- \frac{ika\mu_{co}}{u}\frac{J_\nu'(u)}{J_\nu(u)}\mathbf{e}_\varphi \otimes \mathbf{b},$$

$$\Gamma_{cl} = -\frac{iw}{ka\varepsilon_{cl}}\frac{K_\nu(w)}{K_\nu'(w)}\left(\mathbf{b} + \frac{\beta\nu a}{w^2}\mathbf{e}_\varphi\right) \otimes \left(\mathbf{e}_\varphi - \frac{\beta\nu a}{w^2}\mathbf{b}\right) \quad (12)$$

$$+ \frac{ika\mu_{cl}}{w}\frac{K_\nu'(w)}{K_\nu(w)}\mathbf{e}_\varphi \otimes \mathbf{b},$$

where $u^2 = k^2a^2(\varepsilon_{co}\mu_{co} - \beta^2/k^2)$ and $w^2 = k^2a^2(\beta^2/k^2 - \varepsilon_{cl}\mu_{cl})$.

Next, multiplying the boundary conditions

$$\begin{pmatrix} \mathbf{H}_t^{(co)}(a) \\ \Gamma_{co}\mathbf{H}_t^{(co)}(a) \end{pmatrix} = \begin{pmatrix} \mathbf{H}_t^{(cl)}(a) \\ \Gamma_{cl}\mathbf{H}_t^{(cl)}(a) \end{pmatrix} \quad (13)$$

by block matrix $(\Gamma_{cl} - I)$, we get the equation

$$\Theta\mathbf{H}_t^{(co)}(a) = 0, \quad (14)$$

where $\Theta = \Gamma_{cl} - \Gamma_{co}$ is a planar tensor $(\mathbf{e}_r\Theta = \Theta\mathbf{e}_r = 0)$.

Thus, tensor Θ has two eigenvectors, $\mathbf{e}_r$ and $\mathbf{H}_t^{(co)}(a)$, which correspond to the zero eigenvalue. Accordingly, Θ is a dyad and the invariant dispersion relation takes the form [9]

$$\bar{\Theta}_t = 0, \quad (15)$$

where $\bar{\Theta}_t$ is an invariant (the trace of the tensor adjoint to Θ).

Taking into account that $\bar{\Theta} = \bar{\Theta}_t\mathbf{e}_r \otimes \mathbf{e}_r = 0$ and applying the Hamilton–Cayley theorem

$$\bar{\Theta} - \bar{\Theta}_t = \Theta(\Theta - \Theta_t), \quad (16)$$

we write dispersion relation (15) in the form [9]

$$(\Theta^2)_t = (\Theta_t)^2. \quad (17)$$

Substituting impedance tensors (12) into Θ, we get a dispersion relation for a graded-index fiber that coincides with the well-known expression from the theory of circular fiber [18],

$$\left(\frac{J_\nu'(u)}{uJ_\nu(u)} + \frac{\mu_{cl}}{\mu_{co}}\frac{K_\nu'(w)}{wK_\nu(w)}\right)\left(\frac{J_\nu'(u)}{uJ_\nu(u)} + \frac{\varepsilon_{cl}}{\varepsilon_{co}}\frac{K_\nu'(w)}{wK_\nu(w)}\right)$$

$$= \frac{\beta^2\nu^2}{k^2\varepsilon_{co}\mu_{co}}\frac{(u^2 + w^2)^2}{u^4w^4}.$$

The solution of Eq. (14) also yields tangential components $\mathbf{H}_t^{(co)}(a)$ of the field vector at the interface. In fact, multiplying (16) by arbitrary vector $\mathbf{p}$ that satisfies the condition $(\Theta - \Theta_t)\mathbf{p} \neq 0$ and comparing the result with (14), we get

$$\mathbf{H}_t^{(co)}(a) = (\Theta - \Theta_t)\mathbf{p}. \quad (18)$$

The amplitudes of the independent waves at any point r in the core are determined from (11) at $\mathbf{H}_{t1}(a) = \mathbf{H}_t^{(co)}(a)$ and $\mathbf{H}_{t2}(a) = 0$. Vectors $\mathbf{H}_t(r)$ and $\mathbf{E}_t(r)$ in the cladding are found in a similar way by putting $\mathbf{H}_{t1}(a) = 0$ and $\mathbf{H}_{t2}(a) = \mathbf{H}_t^{(co)}(a)$. The total field vectors are expressed through the restoration matrix (see (3)). The polarization of waveguide modes in the core is then written in the form

$$\begin{pmatrix} \mathbf{H}^{(co)}(r) \\ \mathbf{E}^{(co)}(r) \end{pmatrix} = V\begin{pmatrix} \alpha_1(r) \\ \beta_1(r) \end{pmatrix}\alpha_1^-(a)(\Theta - \Theta_t)\mathbf{p} \quad (19)$$

(similar expressions can be written for the cladding). Propagation constants β appearing in (19) are found from dispersion relation (17). Thus, to solve the waveguide problem, one has to know tensor Θ. Knowing this tensor, we write dispersion relation (17) and determine the polarization of the modes (19).

In the case of a multilayer circular fiber with permittivities and permeabilities specified as

$$\varepsilon = \begin{cases} \varepsilon_{co}, & r < a_0 \\ \varepsilon_j, & a_{j-1} \leq r \leq a_j \\ \varepsilon_{cl}, & r \geq a_n, \end{cases}$$

$$\mu = \begin{cases} \mu_{co}, & r < a_0 \\ \mu_j, & a_{j-1} \leq r < a_j, \quad j = 1, \ldots, n \\ \mu_{cl}, & r \geq a_n, \end{cases}$$

the boundary conditions yield the relationship

$$\Theta\mathbf{H}_t^{(co)}(a_0) = 0, \quad \Theta = (\Gamma_{cl} - I)\Omega_{a_0}^{a_n}\begin{pmatrix} I \\ \Gamma_{co} \end{pmatrix}. \quad (20)$$

Evolution operator $\Omega_{a_0}^{a_n}$ is the product of the evolution operators for each of the layers (see (8)) in the following order:

$$\Omega_{a_0}^{a_n} = P_n(a_n)P_n^-(a_{n-1})P_{n-1}(a_{n-1})P_{n-1}^-(a_{n-2})\ldots$$
$$\ldots P_1(a_1)P_1^-(a_0). \quad (21)$$

For a graded-index fiber,

$$\varepsilon = \begin{cases} \varepsilon_{co}, & r < a \\ \varepsilon(r), & a \le r < b \\ \varepsilon_{cl}, & r \ge b, \end{cases} \quad \mu = \begin{cases} \mu_{co}, & r < a \\ \mu(r), & a \le r < b \\ \mu_{cl}, & r \ge b, \end{cases}$$

the boundary conditions yield a relationship similar to (20) in which $\Omega_{a_0}^{a_n}$ should be replaced by evolution operator Ω_a^b. The latter can be represented as multiplicative integral (6). Numerically, Ω_a^b can be found in two ways. The first way is to partition the interval (a, b) into n parts and set block matrix M constant and equal to $M(r_j)$ within each of the parts (r_j, r_{j+1}), $j = 0, 1, …, (n-1))$. Then, the multiplicative integral is representable as the product of exponentials with a matrix argument,

$$\Omega_a^b = \exp(ikM(r_{n-1})\Delta r_{n-1})\exp(ikM(r_{n-2})\Delta r_{n-2})… \tag{22}$$
$$…\exp(ikM(r_0)\Delta r_0),$$

where $r_n = b$, $r_0 = a$, and $\Delta r_j = r_{j+1} - r_j$.

The evolution operator of a homogeneous layer can also be calculated by formula (22). The second way implies the partition of an inhomogeneous layer into n homogeneous sublayers. In this case, the evolution operator can be found by formula (21). For the same number n of layers into which a fiber is split, formula (21) gives a more accurate value of the multiplicative integral than expression (22). On the other hand, expression (22) does not require calculation of special functions appearing in block matrix P and, thereby, may cut the time taken to compute the multiplicative integral.
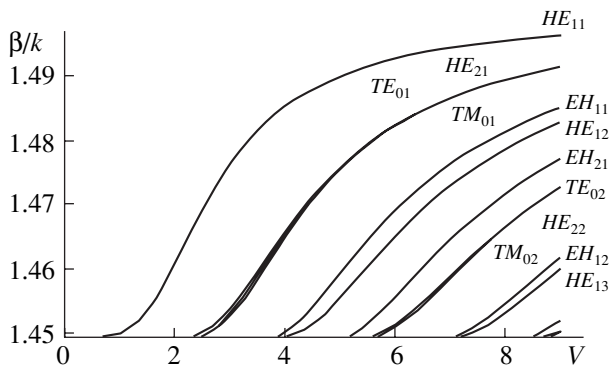
Dimensionless propagation constant β/k vs. fiber parameter $V = ka\sqrt{\varepsilon_{co} - \varepsilon_{cl}}$ for the modes of a fiber with a periodic sequence of layers at $\nu = 0, 1, 2$. The parameters of the problem are $\varepsilon_{co} = 2.25$, $\varepsilon_{cl} = 2.10$, $d = 0.4a$, and $n = 5$.

NUMERICAL CALCULATION AND CONCLUSIONS

By way of example, consider a periodic multilayer structure

$$\varepsilon = \begin{cases} \varepsilon_{co}, & r < a, \\ \varepsilon_{cl}, & a + (j-1)d \le r < a + (j-1/2)d, \\ \varepsilon_{co}, & a + (j-1/2)d \le r < a + jd, \\ \varepsilon_{cl}, & r \ge a + nd, \end{cases}$$

$$\mu = 1, \quad j = 1, …, n,$$

where a is the fiber core radius and n is the number of periods of width d.

One half of the period, $d/2$, is occupied by a medium with permittivity ε_{cl}; the second, by a medium with permittivity $\varepsilon_{co} > \varepsilon_{cl}$. The modes of such a fiber are specified by dispersion relation (17). Planar tensor Θ is calculated by formula (20), and the evolution operator equals matrix product (21).

The figure shows the solutions to the dispersion equation for $\nu = 0, 1, 2$. The case $\nu = 0$ corresponds to the *TE* and *TM* modes. For the other ν, the modes are hybrid; that is, neither longitudinal coordinate (H_y and E_y) vanishes. The type of the hybrid mode is defined by the ratio between the longitudinal components of the electric and magnetic fields [18],

$$\delta = \frac{E_y}{iH_y} = \frac{\mathbf{b}\Gamma_{co}(a)(\Theta - \Theta_t)\mathbf{p}}{i\mathbf{b}(\Theta - \Theta_t)\mathbf{p}}.$$

The case $\delta > 0$ corresponds to the *HE* mode; $\delta < 0$, to the *EH* mode. The HE_{11} mode has the lowest cutoff frequency and, hence, is the fundamental mode. The fiber is regarded as single-mode (i.e., supporting only the fundamental mode) if the fiber parameter $V = ka\sqrt{\varepsilon_{co} - \varepsilon_{cl}}$ is smaller than the cutoff frequency of the next mode ($V < 2.4$ according to the figure).

Thus, we have derived dispersion relations and found the polarizations of modes for circular multilayer isotropic fibers. The awkwardness of mathematical manipulation in deriving the dispersion relations is compensated for by the generality of the approach and feasibility of algorithmizing the problem irrespective of the complexity of media (from the class of media relevant to our problem) and the number of layers. For anisotropic fibers, the permittivity and permeability tensors of which are complex and independent of the radial coordinate, dispersion relations (17) and formulas (18) for the mode polarization are also valid. In this case, propagation constants β are complex; that is, the modes decay. If tensors $\varepsilon(r, \varphi)$ and $\mu(r, \varphi)$ are nonisotropic in the cross section of the fiber or the cross section is not circular, Eqs. (2)–(5) fail. Then, it is necessary to solve the partial Maxwell equations for the electric and magnetic fields.

REFERENCES

1. A. E. Willner, OPN Trends **1**, 16 (2002).
2. J. Zweck, I. T. Lima, Jr., Yu. Sun, *et al.*, Opt. Phot. News **14** (11), 30 (2003).
3. A. I. Andronova and G. B. Malykin, Usp. Fiz. Nauk **172**, 849 (2002) [Phys. Usp. **45**, 793 (2002)].
4. G. B. Malykin and V. I. Pozdnyakova, Opt. Spektrosk. **95**, 646 (2003) [Opt. Spectrosc. **95**, 603 (2003)].
5. L. A. Zenteno and D. T. Walton, Opt. Photon. News **14** (3), 38 (2003).
6. R. V. Kiyan, A. A. Fotiadi, and O. V. Shakin, Pis'ma Zh. Tekh. Fiz. **29** (9), 24 (2003) [Tech. Phys. Lett. **29**, 450 (2003)].
7. A. G. Borisov, T. V. Volkova, V. A. Gan'shin, *et al.*, Opt. Zh. **70** (9), 9 (2003).
8. L. M. Barkovskii, G. N. Borzdov, and A. V. Lavrinenko, J. Phys. A **20**, 1095 (1987).
9. G. N. Borzdov, J. Math. Phys. **38**, 6326 (1997).
10. L. M. Barkovsky and A. N. Furs, *Operational Methods of Describing Optical Fields in Compound Media* (Belarus. Navuka, Minsk, 2003) [in Russian].
11. L. M. Barkovsky and A. N. Furs, Opt. Spektrosk. **90**, 632 (2001) [Opt. Spectrosc. **90**, 561 (2001)].
12. R. Loudon, OPN Trends **3**, 6 (2003).
13. F. I. Fedorov, *Optics of Anisotropic Media* (Izd. Akad. Nauk BSSR, Minsk, 1958) [in Russian].
14. F. I. Fedorov, *Theory of Gyrotropy* (Nauka i Tekhnika, Minsk, 1976) [in Russian].
15. L. M. Barkovskii, G. N. Borzdov, and F. I. Fedorov, J. Mod. Opt. **37**, 85 (1990).
16. P. E. Krasnushkin and R. B. Baibulatov, Dokl. Akad. Nauk SSSR **182**, 294 (1968) [Sov. Phys. Dokl. **13**, 859 (1968)].
17. A. A. Stratonnikov, A. P. Bogatov, A. E. Drakin, *et al.*, J. Opt. A **4**, 535 (2002).
18. M. Adams, *An Introduction to Optical Waveguides* (Wiley, New York, 1981; Mir, Moscow, 1984).
19. F. R. Gantmacher, *The Theory of Matrices* (Nauka, Moscow, 1967; Chelsea, New York, 1959).

Translated by V. Isaakyan

=========== **RADIOPHYSICS** ===========

Excitation of Oscillations in a Hemispherical Dielectric Resonator by a Radial Magnetic Dipole

Yu. V. Prokopenko, Yu. F. Filippov, and V. M. Yakovenko

Usikov Institute of Radiophysics and Electronics, National Academy of Sciences of Ukraine,
ul. Akad. Proskura 12, Kharkov, 61085 Ukraine
e-mail: prokopen@ire.kharkov.ua

Received October 11, 2004

Abstract—Forced oscillations excited by a radial magnetic dipole in a dielectric hemisphere (resonator) placed on a perfectly conducting plane are studied. It is shown that the dipole excites H modes. When the dipole radiation frequency equals the eigenfrequency of the resonator, an amplitude resonance is observed in the spectrum. The excitation efficiency is high when the magnetic dipole is placed at the maximum of the radial field component of the resonator's eigenmode. © 2005 Pleiades Publishing, Inc.

INTRODUCTION

The study of oscillation excitation of a homogeneous dielectric sphere was pioneered by Mie [1] and Debye [2], who considered scattering of electromagnetic waves by the sphere. External and internal excitation of a finite-conductivity sphere by electric and magnetic currents distributed in a homogeneous isotropic space was addressed in [3, 4]. Today's interest in this problem is associated with the wide application of open dielectric resonators in nondestructive complex-permittivity measurements [5–7], in measurements of the surface impedance of metals and thin-film high-temperature superconductors [7–9], as well as in signal filtering [10] and frequency stabilization techniques [11]. When studying these resonators, researchers focus on eigenmodes of whispering-gallery type, which are characterized by high polar index n. A low leakage of these modes from the resonator is related to the almost total reflection of the resonance field from the smooth spherical surface. This effect was first demonstrated and explained with acoustic waves propagating in a cylindrical gallery [12]. The hemispherical dielectric resonator placed on a conducting plane offers much more extended functionality. In particular, it can be excited through slots cut in a metal screen. In the electromagnetic field distribution over the spherical surface of this resonator operating in whispering-gallery modes, beltlike patterns were observed experimentally [13]. It is of interest to theoretically study the field and power distributions of oscillations excited in a hemispherical dielectric resonator with a conducting planar base.

THEORY

Consider a dielectric hemisphere of radius r_0 that is placed on a perfectly conducting plane (Fig. 1). The hemisphere is made of a homogeneous isotropic substance characterized by complex permittivity ε_d and permeability μ_d. The environment has complex permittivity ε_e and permeability μ_e. The structure is excited by a magnetic dipole in the form of a radial slot in the conducting plane, which has magnetic current density

$$j_r^M(r, \theta, \varphi, t) = j_r^M(r, \theta, \varphi)\exp(-i\omega t),$$

where

$$j_r^M(r, \theta, \varphi) = \frac{J_0\delta(r - r_1)}{r^2\sin\theta}\delta(\theta - \pi/2)\delta(\varphi),$$

J_0 is the total magnetic current, ω is the dipole radiation frequency, $\delta(x)$ is the delta function, and r_1 is the radial position of the source (dipole).

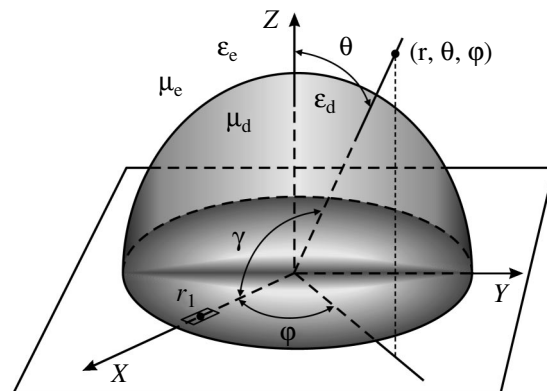**Fig. 1.** Hemispherical dielectric resonator on a conducting base.

The system of Maxwell equations for the field components has the form

$$\frac{\partial}{\partial r}rE_\theta - \frac{\partial}{\partial \theta}E_r = ikr\mu_\nu H_\varphi;$$

$$\frac{1}{\sin\theta}\left(\frac{\partial}{\partial\theta}\sin\theta E_\varphi - \frac{\partial}{\partial\varphi}E_\theta\right)$$

$$= ikr\mu_\nu H_r + \frac{4\pi r}{c}j_r^M(r,\theta,\varphi,t);$$

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\varphi}E_r - \frac{\partial}{\partial r}rE_\varphi = ikr\mu_\nu H_\theta; \quad (1)$$

$$\frac{1}{\sin\theta}\left(\frac{\partial}{\partial\theta}\sin\theta H_\varphi - \frac{\partial}{\partial\varphi}H_\theta\right) = -ikr\varepsilon_\nu E_r;$$

$$\frac{\partial}{\partial r}rH_\theta - \frac{\partial}{\partial\theta}H_r = -ikr\varepsilon_\nu E_\varphi;$$

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\varphi}H_r - \frac{\partial}{\partial r}rH_\varphi = -ikr\varepsilon_\nu E_\theta,$$

where subscript $\nu =$ "d" for $r \le r_0$ and $\nu =$ "e" for $r > r_0$, $k = \omega/c$, and c is the speed of light. All the field components contain time-varying factor $\exp(-i\omega t)$, which is hereinafter omitted.

After eliminating components E_θ, E_φ, H_θ, and H_φ from Eqs. (1), the Maxwell equations reduce to the following differential equations for radial components H_r and E_r:

$$\left[\left(\frac{\partial^2}{\partial r^2} + \chi_\nu^2\right)r^2 + \Delta_\perp\right]H_r(r,\theta,\varphi)$$

$$= \frac{4\pi i}{\omega\mu_\nu}\left(\frac{\partial^2}{\partial r^2} + \chi_\nu^2\right)r^2 j_r^M(r,\theta,\varphi); \quad (2)$$

$$\left[\left(\frac{\partial^2}{\partial r^2} + \chi_\nu^2\right)r^2 + \Delta_\perp\right]E_r(r,\theta,\varphi) = 0.$$

Here, the operator $\Delta_\perp$ is defined as

$$\Delta_\perp = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}$$

and

$$\chi_\nu = k\sqrt{\varepsilon_\nu\mu_\nu}.$$

An isotropic dielectric hemisphere supports H ($E_r = 0$) and E ($H_r = 0$) modes [14]. As follows from Eqs. (2), magnetic current $j_r^M(r,\theta,\varphi)$ excites only H modes.

Let us introduce function $U(r,\theta,\varphi)$ that is related to the radial magnetic field component as

$$H_r(r,\theta,\varphi) = \left(\frac{\partial^2}{\partial r^2} + \chi_\nu^2\right)U(r,\theta,\varphi). \quad (3)$$

Then, the equation in (2) that specifies H modes takes the form

$$\left(\frac{\partial^2}{\partial r^2} + \chi_\nu^2 + \frac{\Delta_\perp}{r^2}\right)U(r,\theta,\varphi) = \frac{4\pi i}{\omega\mu_\nu}j_r^M(r,\theta,\varphi). \quad (4)$$

The tangential field components of H modes can be written in terms of function $U(r,\theta,\varphi)$ as

$$rH_\theta(r,\theta,\varphi) = \frac{\partial^2 U(r,\theta,\varphi)}{\partial\theta\partial r};$$

$$rH_\varphi(r,\theta,\varphi) = \frac{1}{\sin\theta}\frac{\partial^2 U(r,\theta,\varphi)}{\partial\varphi\partial r};$$

$$rE_\theta(r,\theta,\varphi) = \frac{ik\mu_\nu}{\sin\theta}\frac{\partial U(r,\theta,\varphi)}{\partial\varphi}; \quad (5)$$

$$rE_\varphi(r,\theta,\varphi) = -ik\mu_n\frac{\partial U(r,\theta,\varphi)}{\partial\theta}.$$

We seek a solution to Eq. (4) in the form of the expansion

$$U(r,\theta,\varphi) = \sum_{n=0}^\infty \sum_{m=-n}^n R_{nm}(\chi_\nu r)Y_{nm}(\theta,\varphi). \quad (6)$$

Functions $Y_{nm}(\theta,\varphi) = P_n^m(\cos\theta)\exp(im\varphi)$ are the eigenfunctions of operator $\Delta_\perp$ with $P_n^m(\cos\theta)$ being the associated Legendre functions, which are nonzero when $m \le n$. Functions $R_{nm}(\chi_\nu r)$ satisfy the inhomogeneous differential equation

$$\left(\frac{\partial^2}{\partial r^2} + \chi_\nu^2 - \frac{n(n+1)}{r^2}\right)R_{nm}(\chi_\nu r)$$

$$= \frac{4\pi i}{\omega\mu_\nu}\frac{P_n^m(0)}{q_{nm}}\frac{J_0\delta(r-r_1)}{r^2}, \quad (7)$$

where $q_{nm} = \pi(n+m)!/(2n+1)(n-m)!$.

Equation (7) was derived by substituting expansion (6) into Eq. (4), multiplying the obtained result by $Y_{n'm'}(\theta,\varphi)$, and integrating the product over angles φ and θ from 0 to 2π and from 0 to $\pi/2$, respectively. In so doing, the orthogonality condition is taken into account [15],

$$\int_0^{2\pi}d\varphi\int_0^{\pi/2}Y_{nm}(\theta,\varphi)Y_{n'm'}(\theta,\varphi)\sin\theta d\theta = q_{nm}\delta_{nn'}\delta_{mm'},$$

where δ_{ij} is the Kronecker delta.

On the conducting surface ($\theta = \pi/2$), functions $P_n^m(\cos\theta)$ and $(d/d\theta)P_n^m(\cos\theta)$ satisfy the following relationships:

$$P_n^m(0) = \frac{2^m}{\sqrt{\pi}}\cos\left(\frac{\pi}{2}(n+m)\right)\frac{\Gamma((n+m+1)/2)}{\Gamma((n-m+2)/2)},$$

$$\frac{d}{d\theta}P_n^m(\cos\theta)\Big|_{\theta=\pi/2} \qquad (8)$$

$$= \frac{2^{m+1}}{\sqrt{\pi}}\sin\left[\frac{\pi}{2}(n+m)\right]\frac{\Gamma((2+n+m)/2)}{\Gamma((1+n-m)/2)},$$

where $\Gamma(x)$ is the gamma function.

On the conducting surface of the resonator, field component E_φ vanishes. As follows from Eqs. (5) and (8), this boundary condition is satisfied when the sum of indices $n+m$ is even. Consequently, azimuth-homogeneous H modes ($m=0$) with odd polar index n cannot be excited in the resonator.

Function

$$R_{nm}(\chi_v r) = \frac{P_n^m(0)}{q_{nm}}R_n(\chi_v r),$$

which is a solution to differential equation (7), is finite at the center of the hemisphere and satisfies the radiation condition at infinity. Function $R_n(\chi_v r)$ depends on source location r_1 and is defined as follows:

for $r_1 \le r_0$,

$$R_n(\chi_v r) = \begin{cases} T_n(\chi_d r) + A_n^{(1)}j_n(\chi_d r), & r \le r_0 \\ B_n^{(1)}h_n^{(1)}(\chi_e r), & r > r_0; \end{cases}$$

for $r_1 > r_0$,

$$R_n(\chi_v r) = \begin{bmatrix} A_n^{(2)}j_n(\chi_d r), & r \le r_0 \\ T_n(\chi_e r) + B_n^{(2)}h_n^{(1)}(\chi_d r), & r > r_0. \end{bmatrix}$$

Here, $j_n(z) = \sqrt{\pi z/2}\, J_{n+1/2}(z)$ and $h_n^{(1)}(z) = \sqrt{\pi z/2}\, H_{n+1/2}^{(1)}(z)$, where $J_{n+1/2}(z)$ and $H_{n+1/2}^{(1)}(z)$ are the cylindrical Bessel function and Hankel function of the first kind, respectively. Function

$$T_n(\chi_v r) = \begin{bmatrix} \dfrac{4\pi}{c}\dfrac{J_0}{\mu_v\sqrt{\varepsilon_v\mu_v}k^2 r_1^2}h_n^{(1)}(\chi_v r_1)j_n(\chi_n r), & r \le r_1, \\[3mm] \dfrac{4\pi}{c}\dfrac{J_0}{\mu_v\sqrt{\varepsilon_v\mu_v}k^2 r_1^2}j_n(\chi_v r_1)h_n^{(1)}(\chi_v r), & r \ge r_1 \end{bmatrix}$$

describes the effect of the excitation source on the field inside and outside the resonator. Amplitudes $A_n^{(1)}$, $A_n^{(2)}$, $B_n^{(1)}$, and $B_n^{(2)}$ of the fields excited by the magnetic dipole are found from the boundary conditions of tangential field component continuity on the surface of the hemisphere, $r = r_0$, and have the form

$$A_n^{(1)} = [\sqrt{\varepsilon_e\mu_d}T_n(\chi_d r_0)h_n^{(1)'}(\chi_e r_0)$$
$$- \sqrt{\varepsilon_d\mu_e}T_n'(\chi_d r_0)h_n^{(1)}(\chi_e r_0)]/\Delta_n(\omega),$$

$$A_n^{(2)} = \mu_e\sqrt{\varepsilon_e}[T_n'(\chi_e r_0)h_n^{(1)}(\chi_e r_0)$$
$$- T_n(\chi_e r_0)h_n^{(1)'}(\chi_e r_0)]/\sqrt{\mu_d}\Delta_n(\omega),$$

$$B_n^{(1)} = \mu_d\sqrt{\varepsilon_d}[T_n(\chi_d r_0)j_n'(\chi_d r_0)$$
$$- T_n'(\chi_d r_0)j_n(\chi_d r_0)]/\sqrt{\mu_e}\Delta_n(\omega), \qquad (9)$$

$$B_n^{(2)} = [\sqrt{\varepsilon_e\mu_d}T_n'(\chi_e r_0)j_n(\chi_d r_0)$$
$$- \sqrt{\varepsilon_d\mu_e}T_n(\chi_e r_0)j_n'(\chi_d r_0)]/\Delta_n(\omega).$$

Expressions (9) for the amplitudes have the denominator

$$\Delta_n(\omega) = \sqrt{\varepsilon_d\mu_e}j_n'(\chi_d r_0)h_n^{(1)}(\chi_e r_0)$$
$$- \sqrt{\varepsilon_e\mu_d}j_n(\chi_d r_0)h_n^{(1)'}(\chi_e r_0).$$

The prime by the cylindrical functions means differentiation with respect to the argument. Solutions to the equation

$$\Delta_n(\omega_p) = 0 \qquad (10)$$

are eigenfrequencies ω_p of the resonator [16]. Because of radiation losses and dielectric losses, the eigenfrequencies of the resonator are complex, $\omega_p = \omega_p' - i\omega_p''$ ($\omega_p'' \ge 0$). The imaginary part of eigenfrequency ω_p determines the damping constant of a mode with index p. In general, each pth oscillation mode is characterized by three indices: n, m, and s. Polar index n specifies the number of field variations in polar coordinate θ for azimuth-symmetric oscillations ($m=0$). Azimuth index m determines the number of field variations for azimuth oscillations in φ. Radial index s, which is the serial number of the root of dispersion relation (10), is responsible for the number of field variations in radial coordinate r.

At the same time, frequency ω of the forced oscillations is real. When frequency ω of the forced oscillations is equal to the real part, ω_p', of one of the eigenfrequencies, an amplitude resonance is observed. As a result, the frequencies and polar indices of the forced and eigen H modes of the resonator are equal to each other.

After applying the summation theorem [15],

$$P_n(\cos\gamma) = \sum_{m=-n}^{n}\frac{(n-m)!}{(n+m)!}P_n^m(\cos\theta)P_n^m(0)\exp(im\varphi),$$

the field components of the forced H modes, which are determined by relationships (3) and (5) in view of solution (6), take the form

$$E_r = 0;$$

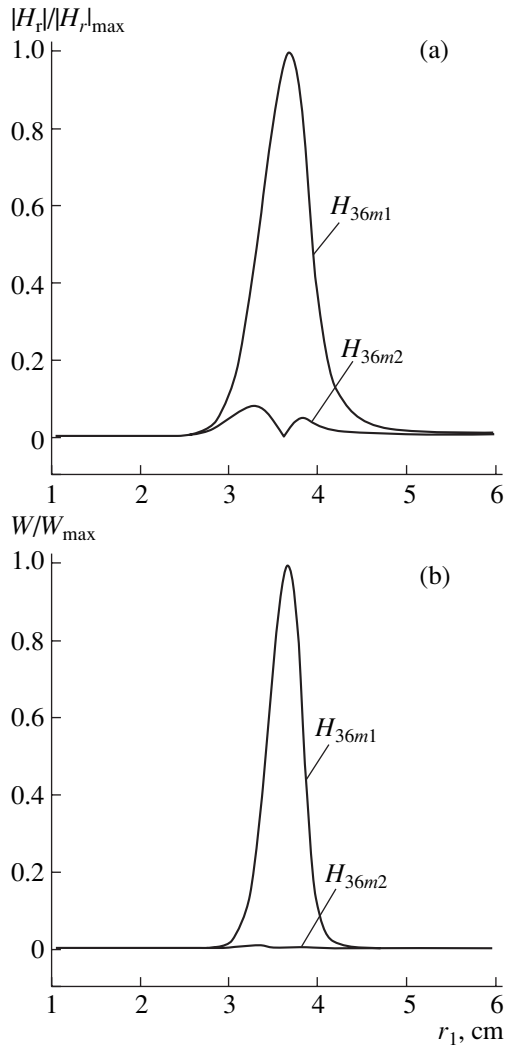$$H_r = \frac{1}{\pi r^2}\sum_{n=0}^{\infty}n(n+1)(2n+1)R_n(\chi_v r)P_n(\cos\gamma),$$

Fig. 2. Normalized (a) magnitude, $|H_r|/|H_r|_{max}$, of component H_r of the field and (b) power density of the H_{36ms} mode at the point with coordinates $r = 3.9$ cm and $\theta = \varphi = \pi/4$ vs. the location of the source.

Fig. 3. Power density distribution of the H_{36m1} mode excited by the source located at $r_1 =$ (a) 3.7 and (b) 3.9 cm.

$$H_\theta = \frac{\cos\theta\cos\varphi}{\pi r}\chi_\nu \sum_{n=0}^{\infty}(2n+1)R'_n(\chi_\nu r)P'_n(\cos\gamma),$$

$$E_\theta = -ik\mu_\nu\frac{\sin\varphi}{\pi r}\sum_{n=0}^{\infty}(2n+1)R_n(\chi_\nu r)P'_n(\cos\gamma),$$

$$H_\varphi = -\frac{\sin\varphi}{\pi r}\chi_\nu\sum_{n=0}^{\infty}(2n+1)R'_n(\chi_\nu r)P'_n(\cos\gamma),$$

$$E_\varphi = -ik\mu_\nu\frac{\cos\theta\cos\varphi}{\pi r}\sum_{n=0}^{\infty}(2n+1)R_n(\chi_\nu r)P'_n(\cos\gamma),$$

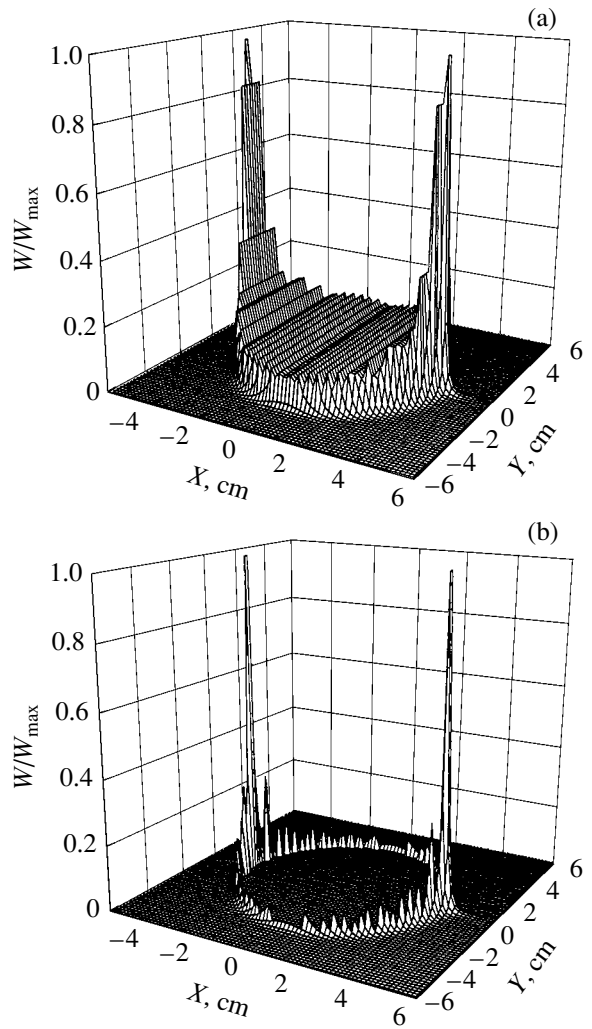where $P_n(\cos\gamma)$ is the Legendre polynomial, $\cos\gamma = $ sinθcosφ, and γ is the angle between the source and observation point (Fig. 1).

Thus, each of the modes of a hemispherical isotropic dielectric resonator with a conducting base exhibits $(n+1)$-fold frequency degeneracy in azimuth index m. The mode degeneracy is also observed for forced oscillations in the resonator excited by a radial magnetic dipole.

The power density of the H modes excited at any observation point is given by the relationship [17]

$$W = (\varepsilon_\nu\mathbf{E}\cdot\mathbf{E}^* + \mu_\nu\mathbf{H}\cdot\mathbf{H}^* + \text{c.c.})/16\pi,$$

where the asterisk and c.c. mean complex conjugates.

NUMERICAL EXPERIMENT

Numerical simulations were performed with a Teflon ($\varepsilon_d = 2.04(1 + 1.7 \times 10^{-4}i)$, $\mu_d = 1$) hemispherical
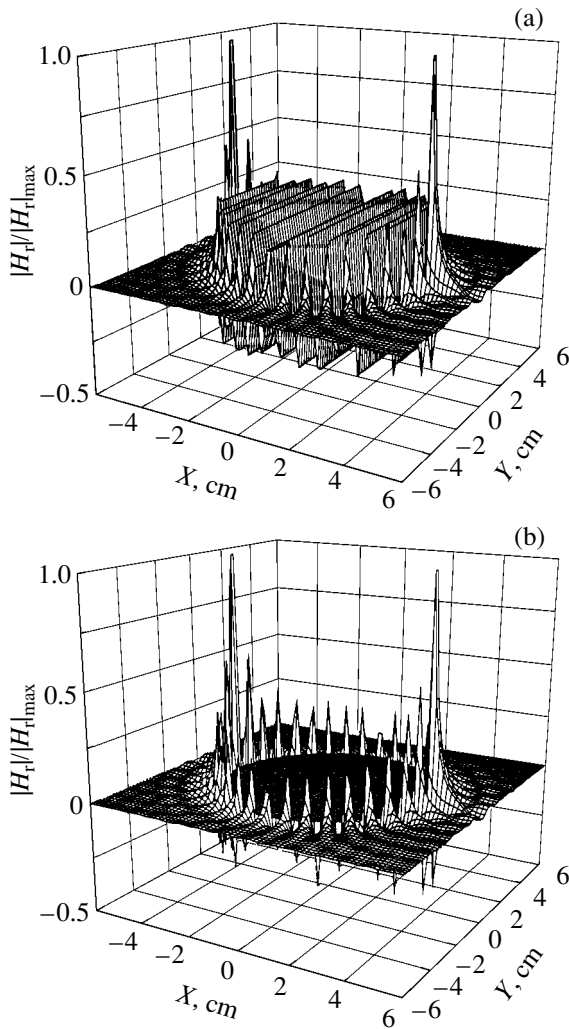
Fig. 4. Distribution of the H_r component of the H_{36m1} mode for the source located at $r_1 =$ (a) 3.7 and (b) 3.9 cm.

resonator of radius $r_0 = 3.9$ cm placed on a perfectly conducting plane. The resonator was surrounded by a medium with $\varepsilon_e = \mu_e = 1$. Excitation of H_{36ms} oscillations was considered. The resonator excitation efficiency varied with distance of the magnetic dipole to the resonator center (Fig. 2). The excitation was the most efficient when the point of excitation (dipole's position) coincided with the maximum of the resonator eigenmode power distribution. Experimentally, this means that the excitation efficiency depends on the degree of coupling between the source and hemisphere. In our numerical simulations, the coupling coefficient was proportional to the radial distance between the source and the maximum of the field distribution of the corresponding whispering gallery mode.

The distributions of the power density and radial field component of the H_{36m1} mode on the hemisphere and on the conducting plane outside the dielectric are shown in Figs. 3 and 4. This mode was excited at a fre-

quency of 35.445 GHz. The Q factor of the resonator thus excited is 5013.7. A (point) source was placed at two points: one, with angular coordinates $\theta = \pi/2$ and $\varphi = 0$ and radial coordinate $r_1 = 3.7$ cm, coincided with the maximum of the eigenmode power distribution; the other, with the same angular coordinates and $r_1 = 3.9$ cm, lay on the surface of the hemisphere. At $r_1 = r_0$, the energy of the mode excited concentrates in the form of a belt close and parallel to the conducting surface (Fig. 3b). Such a field distribution is typical of the azimuth mode with index $m = n$.

The maxima of the power density distribution for the oscillation excited (Fig. 3) coincide with the maxima of the resonator's eigenmode power distribution and are in the same radial plane where the source is placed. The power density and field distributions (Fig. 3) show a beltlike pattern, which is especially distinct on the conducting surface. The belt passes through the excitation point and pole of the hemisphere. Such a distribution is typical of the azimuth-homogeneous oscillation with index $m = 0$. By appropriately choosing the source location, one can somewhat amplify a mode with particular index m.

A similar field distribution is observed for mode H_{36m2}. The only difference is that the efficiency of excitation of this mode is much lower (Fig. 2) despite its higher frequency (39.743 GHz). The unloaded Q factor of the resonator for this mode is 498.65.

CONCLUSIONS

Forced oscillations in an isotropic dielectric hemisphere placed on a perfectly conducting plane are studied. It is shown that a radial magnetic dipole excites only H modes. As follows from the boundary condition on the conducting plane, the sum $n + m$ of the polar and azimuth indices for the H modes is even. As a result, the forced, as well as natural, H modes exhibit $(n + 1)$-fold degeneracy in azimuth index m. Azimuth-homogeneous H modes ($m = 0$) with odd polar index n cannot be excited in this resonator.

When the dipole radiation frequency approaches the frequency of a resonator's eigenmode, an amplitude resonance is observed. The radial magnetic dipole is demonstrated to form a beltlike power distribution of the whispering-gallery modes in the hemisphere—the effect experimentally found in [13].

ACKNOWLEDGMENTS

REFERENCES

1. G. Mie, Ann. Phys., No. 25, 77 (1908).
2. P. Debye, Ann. Phys., No. 30, 17 (1909).
3. G. T. Markov, Zh. Tekh. Fiz. **23**, 838 (1953).

4. G. T. Markov and A. F. Chaplin, *Generation of Electromagnetic Waves* (Radio i Svyaz', Moscow, 1983) [in Russian].

5. V. N. Derkach, Yu. F. Filippov, A. S. Plevako, *et al.*, Int. J. Infrared Millim. Waves **25**, 139 (2004).

6. Yu. V. Prokopenko, T. A. Smirnova, Yu. F. Filippov, *et al.*, Vopr. At. Nauki Tekh., No. 4, 87 (2004).

7. Yu. V. Prokopenko and Yu. F. Filippov, Zh. Tekh. Fiz. **72** (6), 79 (2002) [Tech. Phys. **47**, 731 (2002)].

8. A. A. Barannik, Yu. F. Prokopenko, Yu. V. Filipov, *et al.*, Zh. Tekh. Fiz. **73** (5), 99 (2003) [Tech. Phys. **48**, 621 (2003)].

9. N. T. Cherpak, A. A. Barannik, Yu. V. Prokopenko, *et al.*, IEEE Trans. Appl. Supercond. **33**, 3570 (2003).

10. S. V. Boriskina and A. I. Nosich, Radiofiz. Radio Astron. **2**, 333 (1997).

11. S. Kharkovsky, A. Kirichenko, and A. Kogut, Microwave Opt. Technol. Lett. **12**, 210 (1996).

12. J. W. Strutt (Lord Rayleigh), Philos. Mag. **20**, 1001 (1910).

13. S. N. Kharkovsky, A. E. Kogut, and V. A. Solodovnik, Pis'ma Zh. Tekh. Fiz. **21** (18), 38 (1995) [Tech. Phys. Lett. **21**, 741 (1995)].

14. K. G. Sullivan and D. G. Hall, Phys. Rev. A **50**, 2701 (1994).

15. H. Bateman and A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953; Nauka, Moscow, 1966), Vol. 2.

16. Yu. V. Prokopenko, T. A. Smirnova, and Yu. F. Filippov, Zh. Tekh. Fiz. **74** (4), 82 (2004) [Tech. Phys. **49**, 459 (2004)].

17. *Fundamental Formulas of Physics*, Ed. by D. H. Menzel (Dover, New York, 1960; Inostrannaya Literatura, Moscow, 1957).

Translated by A. Khzmalyan

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Efficiency of Pulsed Electrode Conditioning in a Vacuum

A. A. Emelyanov

Orel State Technical University, Orel, 302020 Russia
e-mail: emel@ostu.ru
Received February 27, 2004; in final form, September 8, 2004

Abstract—It is shown that the efficiency of pulsed electrode conditioning in a vacuum, which is estimated from the state of the electrode surface, rises as conditioning pulses get shorter and may exceed the efficiency of conditioning by dc breakdowns by two orders of magnitude or more. A criterion with which one can judge the cathodic mechanism of breakdown initiation in the steady-state regime is suggested. Under optimal electrode processing conditions (pulse width $t_p < 10^{-8}$ s, field strength $E_0 > 10^8$ V/m), the ultimate dielectric strength corresponding to the cathodic mechanism of breakdown initiation in a vacuum can be reached. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

The condition of the electrode surface has a pronounced effect on the vacuum insulation quality. Microprotrusions, loosely bonded particles, oxide films, contamination, dielectric inclusions, and gas bubbles present on the surface all influence the values of prebreakdown current and breakdown voltage. Training by breakdowns is the most popular method of electrode conditioning, training by milli- or microsecond pulses being little different in efficiency from training by steady breakdown [1]. Application of high-voltage nanosecond pulses cuts the number of vacuum breakdown initiation mechanisms, leaving the cathodic mechanism as the most plausible. In this mechanism, Joule heating due to the passing field-emission current causes the explosive breakdown of the emitter.

Joule heating of the emitter makes it possible to consider breakdown delay time t_d as a function of electric field strength E and physical constants of the emitter material. The processing of electrodes by high-voltage pulses whose width equals the breakdown time delay, $t_p = t_d$, is the optimal regime of conditioning. Under such conditions, the cathode surface is formed with the lowest value of field amplification factor β. Comparing (at $t_\alpha = $ const) analytical dependences of the time delay on the field microstrength at the cathode, $t_d = f(E)$, with experimental time delay versus macrostrength curves, $t_d = f(E_0)$, makes it possible to estimate the efficiency of optimal pulsed conditioning regimes from a change in the state of the surface. The conditioning efficiency may also be estimated from a change in the electrical performance of vacuum insulation.

BREAKDOWN DELAY IN A VACUUM

Application of voltage pulses with a duration

$$t_r \ll t_p \ll h^2 \rho c / \lambda \tag{1}$$

(t_r is the time of thermal relaxation; h is the emitter height; and ρ, c, and λ are, respectively, the density, specific heat, and thermal conductivity of the emitter material) reduces the boundary-value problem of emitter temperature distribution [2] to the form [3]

$$\begin{cases} \rho c \dfrac{dT}{dt} = j_0^2 \left(\dfrac{\pi T / 2T^*}{\sin(\pi T / 2T^*)} \right)^2 k_0 T \\ T|_{t=0} = T_0, \quad T|_{t=t_d} = T_{cr}, \end{cases} \tag{2}$$

where j_0 is the field-emission current density; k_0 is a proportionality coefficient in the temperature dependence of the resistivity of the emitter material, $k(T) = k_0 T$; T^* is the inversion temperature, $T^* = 5.67 \times 10^{-7} E \varphi^{-1/2}$; and φ is the work function.

Equation (2) written in integral form [4],

$$\int_0^{t_d} j_0^2(t)\, dt = a \frac{\rho c}{k_0}, \tag{3}$$

where

$$a = \left[\text{Ci}(x) - \frac{\sin x}{x} - \frac{1}{2} \frac{\sin^2(x/2)}{(x/2)^2} \right]\Bigg|_{x=\pi T_0/T^*}^{x=\pi T_a/T^*}$$

with

$$\text{Ci}(x) = -\int_x^{\infty} \frac{\cos y}{y}\, dy$$

represents a criterion of vacuum breakdown initiation. According to this criterion, the energy sufficient to break the emitter is released for a time that is equal to breakdown delay time $t_p = t_d$. Quantity $a\rho c / k_0$ is the specific energy of emitter breakdown and is constant for a given metal.

If the voltage pulses are rectangular, one can easily express from (3) the time delay as a function of the emitter breakdown specific energy, work function, and electric field microstrength,

$$t_d = 4.2 \times 10^{11} \frac{a\rho c}{k_0} \varphi^2 \exp\{-21.6\varphi^{-1/2}\}$$
$$\times \frac{\exp\{1.32 \times 10^{10} \varphi^{3/2} E^{-1}\}}{E^4}. \qquad (4)$$

OPTIMALITY CRITERION FOR PULSED CONDITIONING

Electrode conditioning in the prebreakdown regime, $t_p < t_d$, is inefficient, since the energy is insufficient for emitter breakdown. The conditioning efficiency rises with voltage pulse duration and reaches a maximum when $t_p = t_d$ (the critical breakdown-initiating condition). The conditioning by pulses with $t_p = t_d$ smoothes out the relief on the cathode surface and even polishes it at $t_p \leq 1$ ns [5].

Under the conditions of explosive electron emission, $t_p > t_d$, the pulse energy not only breaks the emitter but also favors breakdown mechanism switching in the electrode gap. The conditioning efficiency rises when explosive emission current pulses narrow and reaches a maximum (surface cathode polishing) as the pulse width approaches the critical value $t_p = t_d$. The polishing is observed for $t_p \approx 1$ ns [6, 7].

The optimal regime of conditioning corresponds to the critical condition $t_p = t_d$, when the pulse energy breaks the emitter but cannot initiate switching. The processing of the emitter by pulses with $t_p = t_d$ smoothes out microprotrusions without generating new emission centers. The optimal conditioning regime minimizes the emitter surface roughness. In this case, the optimality criterion takes the form [8]

$$\int_0^{t_p} j_0^2(t)dt = a\frac{\rho c}{k_0}. \qquad (5)$$

According to (5), the energy being released in the emitter for pulse time $t_p = t_d$ remains constant and equal to the emitter breakdown energy. As the power of the pulses in the optimal regime grows, their width shrinks. As the pulse width decreases, so do the distance the thermal wave travels and the depth of emitter damage. Accordingly, the conditioning efficiency rises up to polishing.

Experiments on determining the breakdown time delay in a vacuum in which the emitter is processed by pulses with $t_p \approx t_d$ correspond to the optimal conditioning regime and make it possible to estimate the efficiency of optimal conditioning regimes.

EMITTER SURFACE

The condition of the emitter surface is characterized by electric field amplification factor β. This factor is classically determined from the transconductance of the Fowler–Nordheim curve. In the case of pulses with $t_p = t_d$, factor β can be found from the analytical dependence of the time delay on the electric field microstrength $t_d = f(E)$ (see (4)) and the experimental dependence of the time delay on the field macrostrength $t_d = f(E_0)$.

Experimental data for the vacuum breakdown time delay [9–14] were obtained for copper, aluminum, nickel, molybdenum, and iron electrodes with a developed surface. The pulse width was varied from the subnanosecond range to the millisecond range. Comparing the experimental curves $t_d = f(E_0)$ with the calculated curves $t_d = f(E)$ at $t = \text{const}$ made it possible to estimate the field amplification factor for these materials,

$$\beta = E/E_0, \qquad (6)$$

achieved after conditioning by pulses with $t_p \approx t_d$ and construct the dependences of β on the time, $\beta = f(t_p)$, and intensity, $\beta = f(E_0)$, of optimal action.

Figure 1 plots the field amplification factor against the pulse width $t_p \approx t_d$ for copper electrodes. The curve $\beta = f(t_p)$ shows how the emitter surface quality varies with duration $t_p \approx t_d$ of processing pulses. From this curve, it follows that, at $t_p < 10^{-7}$ s, a decrease in the
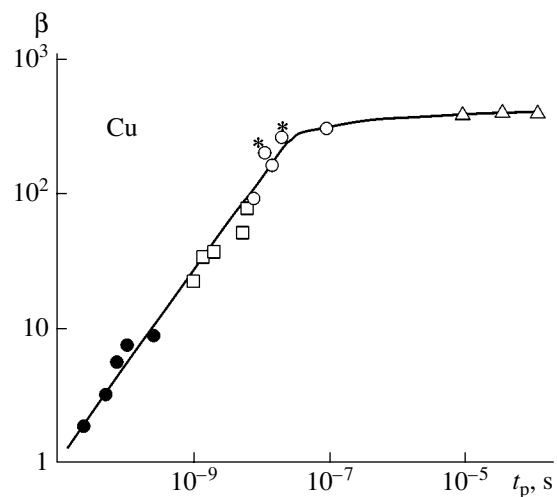

Fig. 1. Field amplification factor vs. the duration of conditioning pulses with $t_p \approx t_d$. Data points are taken from [5] (●), [9] (○), [10] (△), [11] (□), and [12] (∗).
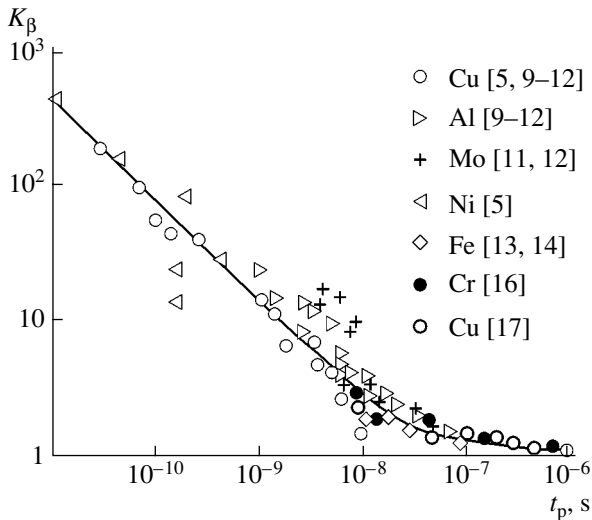
Fig. 2. Emitter surface conditioning efficiency vs. pulse duration $t_p \approx t_d$. (○) Efficiency of raising the dielectric strength of the gap between all-metal electrodes [17] and (●) the voltage of local glows in the gap between the evaporated electrodes [16].

Fig. 3. Field amplification factor after conditioning by pulses with $t_p \approx t_d$ vs. the breakdown-initiating field macrostrength.

pulse width and a corresponding increase in the pulse power reduce β and improve the emitter surface quality. At $t_p \sim 10^{-7}$ s, the run of the curve changes. At $t_p > 10^{-7}$ s, the width does not influence both the field amplification factor and the emitter surface quality and the efficiency of training by milli- and microsecond pulses differs insignificantly from the efficiency of training by dc breakdowns. This is in accordance with the results of [1] and can be explained by cooling of the emitter via heat conduction.

Considering that amplification factor β_0 obtained at $t_p > 10^{-7}$ s corresponds to training by dc breakdowns, one can estimate the efficiency of pulsed conditioning from the relative change in the emitter surface quality,

$$K_\beta = \beta_0 / \beta_p, \qquad (7)$$

where β_p is the amplification factor after pulsed conditioning.

Based on the curve $\beta = f(t_p)$, we can construct the dependence of pulsed conditioning efficiency K_β on pulse duration $t_p \approx t_d$ in the optimal regime.

Experimental data for aluminum [9–12], molybdenum [11, 12], nickel [5], and iron [13, 14] electrodes were also fitted by the curves $t_d = f(E)$ calculated under the assumption of explosive emission. In this way, we estimated factor β for different conditions of training by pulses with $t_p \approx t_d$ and could construct the dependences $K_\beta = f(t_p)$ for these materials. Figure 2 plots the variation of K_β against the time of processing (i.e., in the process of conditioning regime optimization) and data points for the delay time for Cu, Al, Mo, Ni, Fe, and Cr.
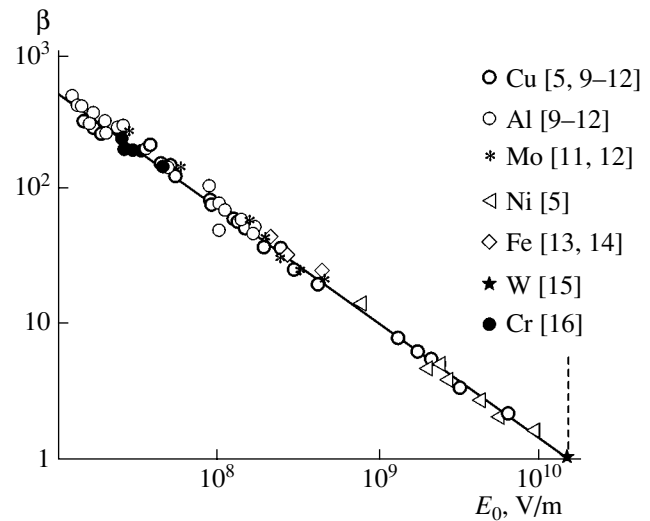
From the curve $K_\beta = f(t_p)$ (Fig. 2) generalizing the available experimental data for the breakdown delay time, it follows that the conditioning efficiency rises as the pulse duration decreases and the pulse power grows. At $t_p < 10^{-8}$ s, the efficiency rises by a factor of three or more. At $t_p < 10^{-10}$ s, the efficiency of pulsed processing exceeds the efficiency of conditioning by dc breakdowns by more than two orders of magnitude. In the limit, conditioning with such pulses may result in a perfectly smooth surface with $\beta = 1$. At $t_p > 10^{-7}$ s, the pulse duration influences the training efficiency only slightly and the efficiency differs little from the efficiency of conditioning in the dc regime.

Achievement of a smooth emitter surface requires not only the pulse duration in the subnanosecond range but also a high electric field strength. Strength E_0 corresponding to the optimal regime with $t_p \approx t_d$ was estimated by comparing the experimental, $t_d = f(E_0)$, and calculated, $t_d = f(E)$, curves at $t_d = $ const. The dependence of amplification factor β on the macrostrength ($\beta = f(E_0)$) that provides the optimal regime of conditioning Cu, Al, Mo, Ni, Fe, W, and Cr electrodes [9–17] is given in Fig. 3.

The unified dependence $\beta = f(E_0)$ in Fig. 3 fits well the experimental data for processing of various materials by pulses with $t_p \approx t_d$. This means that, under the optimal conditioning regime, the field amplification factor is independent of the electrode material and depends only on breakdown-initiating electric field macrostrength E_0. From the curve $\beta = f(E_0)$, it follows
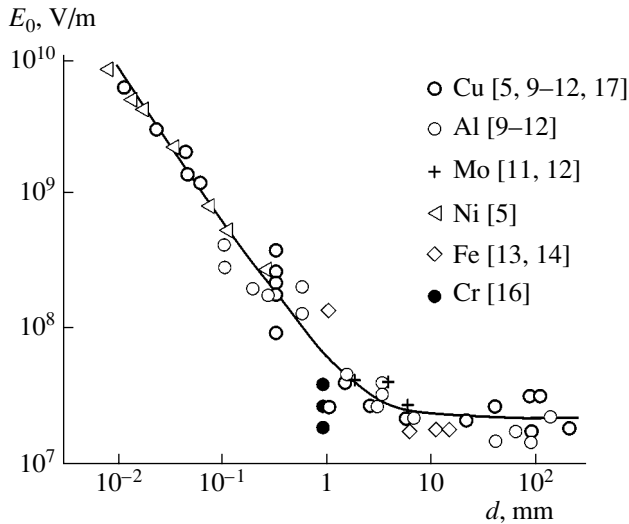
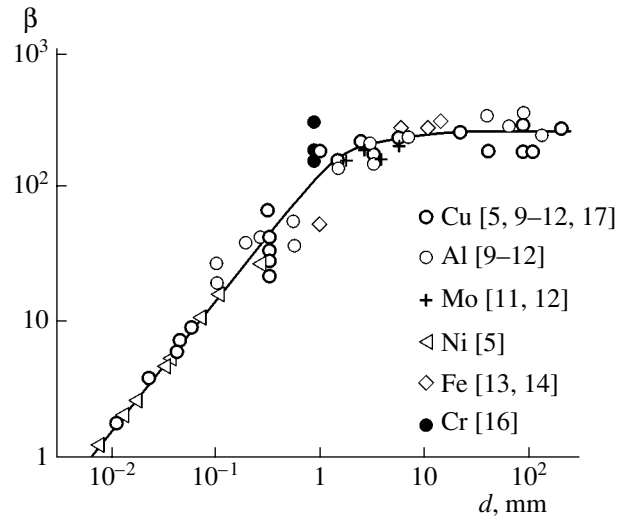**Fig. 4.** Electric field macrostrength at conditioning by pulses with $t_p \approx t_d$ vs. the electrode gap.

Fig. 5. Field amplification factor upon conditioning by pulses with $t_p \approx t_d$ vs. the electrode gap.

that β can be approximated by the power-law dependence

$$\beta = \left(\frac{E_{cr}}{E_0}\right)^{0.9}\Bigg|_{t_p = t_d}, \qquad (8)$$

where $E_{cr} = 1.32 \times 10^{10}$ V/m.

The curve $\beta = f(E_0)$ was constructed using the experimental data for electrodes with a developed surface and gaps varying from 3 μm to 20 cm. Thus, in the experiments, the change in the electric field by three orders of magnitude (from $1.5 \times 10^7 \leq E_0 \leq 1.3 \times 10^{10}$ V/m) and the corresponding change in factor β by three orders of magnitude were provided largely by the change in gap d by five orders of magnitude. The dependence of macrostrength E_0 that provides the optimality of processing on gap d used in conditioning by pulses with $t_p \approx t_d$ is demonstrated in Fig. 4.

From the curve $E_0 = f(d)$ in Fig. 4, it follows that, for $d > 3$ mm, the strength varies insignificantly and equals $E_0 \approx 2 \times 10^7$ V/m. This value corresponds to pulse duration $t_p > 10^{-7}$ s and amplification factor $\beta_0 \approx 270$, which is reached when wide electrodes are processed in the dc regime. At $d < 1$ mm and $t_p < 10^{-10}$ s, the increase in the strength to $E_0 \sim 10^{10}$ V/m, which is accompanied by the decrease in β nearly to the minimal value $\beta = 1$ and the rise in the conditioning efficiency by more than two orders of magnitude, was achieved by shrinking the electrode gap to a micrometer size.

Figure 5 plots amplification factor β obtained under the optimal processing conditions against the electrode gap used in experiments [5, 9–17]. From the curve $\beta = f(d)$ (Fig. 5), it can be concluded that, as the gap narrows, the efficiency of conditioning by pulses with $t_p \approx$ t_d rises; that is, β tends to a minimum as the gap shrinks to micrometer sizes.

Thus, the efficiency of pulsed electrode surface conditioning in the optimal regime grows with decreasing pulse duration and increasing field strength. Field strength E_0 is raised by narrowing the electrode gap. The most efficient regime of conditioning is observed at processing by pulses with $t_p < 10^{-8}$ s and field strengths $E_0 > 10^8$ V/m. Such high values of E_0 are achievable in gaps $d < 1$ mm. In micrometer-wide gaps at $E_0 \sim 10^{10}$ V/m, processing of the electrode by pulses with $t_p < 10^{-10}$ s produces a perfectly smooth surface with $\beta \approx 1$. However, the fact that the emitter surface is perfectly smooth does not mean that the ultimate dielectric strength of the vacuum insulation is reached, since breakdown initiation mechanisms may change when the voltage is applied for a long time.

DIELECTRIC STRENGTH

In the steady-state regime, the cathodic mechanism of breakdown initiation in the electrode gap comes into play when the electric field microstrength reaches the critical value [18]

$$E = \beta E_0 = \text{const.} \qquad (9)$$

Under the conditions of the cathodic breakdown-initiating mechanism, any action changing the emitter surface condition changes field amplification factor β and, accordingly, electric field E_0 and voltage U at which the first breakdown occurs,

$$U \sim E_0 \sim 1/\beta. \qquad (10)$$

Optimized pulsed conditioning makes it possible to bring the emitter surface to a desired quality. According

to (10), a decrease in β is accompanied by an increase in the dielectric strength and first-breakdown voltage under the steady-state conditions of cathodic initiation. Consequently, factor K_β, which characterizes a change in the emitter surface condition, must be equal to a change in the dielectric strength,

$$K_U = U_p/U_0, \qquad (11)$$

where U_p and U_0 are, respectively, the dc voltage of the initial (after processing) breakdown and the steady-state breakdown voltage before processing.

The criterion of the cathodic mechanism can be represented in the form

$$\frac{K_\beta}{K_U} = 1. \qquad (12)$$

Comparing relative changes in the surface quality and in the dielectric strength of vacuum insulation eliminates a systematic error and improves the accuracy of determining the onset of the cathodic mechanism compared with criterion (9). If $K_\beta/K_U \neq 1$, breakdown is caused by other mechanisms.

Criterion (12) was tested in experiments where all-metal copper electrodes ($C \approx 110\pi F$, $d = 0.5$ mm) were processed by rectangular pulses ($10 \leq t_p \leq 800$ ns, $1.5 \times 10^7 \leq E_0 \leq 10^8$ V/m) [19] and in experiments with compound electrodes comprising a cathodoluminescent screen coated by an evaporated microchannel Cr plate ($C \approx 10\pi F$, $d = 0.9$ mm) [16].

In the case of the all-metal electrodes, the efficiency of pulsed conditioning in the optimal regime was estimated from the relative change of the first-breakdown voltage (i.e., from K_U); in the case of the compound (evaporated) electrodes, from the relative change of the voltage at which local glows on the screen were observed. In the latter case, the estimating parameter is

$$K^* = U_p^*/U_0^*, \qquad (13)$$

where U_p^* and U_0^* are the voltages at which local glows were observed, respectively, after and before pulsed conditioning.

Screen glows are associated with the field emission from the exit surface of the microchannel plate, i.e., with the cathodic mechanism, while the breakdown of the gaps in the microchannel plate is due to other mechanisms. That is why the dielectric strength remained constant, $E_0 \sim 10^7$ V/m, irrespective of the pulsed conditioning regime at $t_p \approx t_d$.

The data points for copper and chromium in Fig. 2 show the variation of the dielectric strength and the local glow voltage with the pulse duration in the optimal regime ($K_U = f(t_p)$ and $K^* = f(t_p)$, respectively). For both the all-metal and evaporated electrodes, a decrease in the duration of conditioning pulses improves the performance of the vacuum insulation due to cathodic processes. The experimental curves $K_U = f(t_p)$ and $K^* =$

$f(t_p)$, which reflect the variation of the dielectric strength and screen glow voltage at conditioning by pulses with $10 \leq t_p \leq 800$ ns and $1.5 \times 10^7 \leq E_0 \leq 10^8$ V/m, coincide with the curve $K_\beta = f(t_p)$, which characterizes the change in the emitter surface condition after processing in the optimal regime. When the field amplification factor decreases twofold as a result of emitter conditioning by pulses with $t_p \sim 10^{-8}$ s and $E_0 \sim 10^8$ V/m, the dielectric strength of the gap in the all-metal electrodes increases twofold and so does the local glow voltage in the case of the evaporated electrodes. This coincidence indicates that cathodic processes initiate the breakdown (for the all-metal electrodes) and local screen glows (for the evaporated electrodes), supporting the validity of criterion (12). Conversely, the independence of the dielectric strength of vacuum gaps between the evaporated electrodes from the pulsed conditioning regime leads us to infer that the breakdown-initiating mechanism is other than field emission in this case.

The conditioning of stainless steel Rogowski electrodes ($S = 100$ mm^2, $d = 0.12$ mm) by pulses with $t_p \approx t_d = 10$ ns made it possible to reach the dielectric strength $E_0 = 2.1 \times 10^8$ V/m, which is close to the ultimate value obtained on molybdenum electrodes after long-term heating and conditioning in an argon glow discharge [20].

The run of the curve $K_\beta = f(t_p)$, which describes the improvement of the emitter surface quality with a decrease in the duration $t_p \approx t_d$ of conditioning pulses, suggests that, basically, the dielectric strength observed at the cathodic mechanism of breakdown initiation may be more than two orders of magnitude higher than in the case of dc (steady-state) conditioning. However, when the ultimate dielectric strength is attained in the steady-state regime, the effective breakdown-initiating mechanism may change. If $K_\beta/K_U \neq 1$, the criterion suggested may be used for estimating the ultimate dielectric strength achievable at the cathodic mechanism of breakdown initiation.

CONCLUSIONS

Optimization of pulsed electrode conditioning in a vacuum was considered, and the efficiency of the optimal regime was estimated. The efficiency of emitter surface conditioning by pulses with $t_p \approx t_d$ estimated from the relative change in the emitter surface quality ($K_\beta = f(t_p)$) in experiments on determining the breakdown delay time rises with a decrease in the pulse duration and may exceed the efficiency of conditioning by dc breakdowns by more than two orders of magnitude. In the optimal regime ($t_p < 10^{-10}$ s, $E_0 > 10^{10}$ V/m), the emitter surface may become perfectly smooth with field amplification $\beta = 1$. To attain an electric field strength on the order of $E_0 \sim 10^{10}$ V/m, the vacuum electrode gap should be shrunk to micrometer sizes.

In he case of the cathodic breakdown-initiating mechanism, a change in the emitter surface condition goes in parallel with a change in the dielectric strength of the insulation. This circumstance is used to suggest a new criterion for the cathodic initiation of breakdown. This criterion was tested experimentally in the nanosecond range of pulse durations for all-metal and evaporated electrodes with a developed surface. Application of pulses with $t_p < 10^{-8}$ s and $E_0 > 10^8$ V/m for electrode conditioning makes it possible to estimate the ultimate dielectric strength achievable at cathodic breakdown initiation.

REFERENCES

1. N. V. Cherepnin, *Sorption in Vacuum Technology* (Sov. Radio, Moscow, 1973) [in Russian].

2. E. A. Litvinov, G. A. Mesyats, and A. F. Shubin, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 4, 149 (1970).

3. A. A. Yemelyanov, G. M. Kassirov, and G. V. Smirnov, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 4, 142 (1976).

4. A. A. Yemelyanov and G. M. Kassirov, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 9, 105 (1976).

5. B. Jüttner, W. Rohrbeck, and H. Wolff, in *Proceedings of the 9th International Conference on Phenomena in Ionized Gases, Bucharest, 1969*, p. 140.

6. G. A. Mesyats, D. I. Proskourovsky, and E. B. Yankelevitch, in *Proceedings of the 7th International Symposium on Discharges and Electrical Insulation in Vacuum, Novosibirsk, 1976*, pp. 230–233.

7. G. A. Mesyats, D. I. Proskourovsky, E. B. Yankelevitch, and V. F. Tregubov, Dokl. Akad. Nauk SSSR **227**, 1335 (1976) [Sov. Phys. Dokl. **21**, 228 (1976)].

8. A. A. Emelyanov, Zh. Tekh. Fiz. **73** (9), 113 (2003) [Tech. Phys. **48**, 1192 (2003)].

9. G. M. Kassirov, Zh. Tekh. Fiz. **36**, 1883 (1966) [Sov. Phys. Tech. Phys. **9** (1966)].

10. I. I. Kalyatskiĭ, G. M. Kassirov, G. V. Smirnov, and N. N. Frolov, Zh. Tekh. Fiz. **45**, 1547 (1975) [Sov. Phys. Tech. Phys. **20**, 988 (1975)].

11. G. A. Mesyats, S. P. Bugaev, D. I. Proskourovsky, *et al.*, Radiotekh. Élektron. (Moscow) **14**, 2222 (1969).

12. S. P. Vavilov and G. A. Mesyats, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 8, 90 (1970).

13. G. K. Olendzskaya and M. A. Sal'man, Zh. Tekh. Fiz. **40**, 333 (1970) [Sov. Phys. Tech. Phys. **15**, 242 (1970)].

14. I. D. Chalmers and B. D. Phukan, Vacuum **32**, 145 (1982).

15. G. K. Kartsev, G. A. Mesyats, D. I. Proskourovsky, *et al.*, Dokl. Akad. Nauk SSSR **192**, 309 (1970) [Sov. Phys. Dokl. **15**, 475 (1976)].

16. A. A. Yemelyanov, Prib. Tekh. Éksp., No. 6, 90 (1998).

17. A. A. Yemelyanov, I. I. Kalyatskiy, and G. M. Kassirov, in *Proceedings of the 7th International Symposium on Discharges and Electrical Insulation in Vacuum, Novosibirsk, 1976*, pp. 130–133.

18. D. Alpert, D. A. Lee, F. M. Lyman, *et al.*, J. Vac. Sci. Tech. **1**, 35 (1964).

19. V. P. Buts, A. A. Yemelyanov, G. M. Kassirov, *et al.*, Elektron. Tekh., Ser. 4: Elektrovak. Gazorazryadnye Prib., No. 7, 115 (1978).

20. P. N. Chistyakov and N. P. Dubinin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **22**, 1020 (1979).

Translated by V. Isaakyan

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Tribological and Electret–Thermal Analysis of Medicinal Preparations for Local Treatment of Joint Diseases

Yu. M. Chernyakova*, L. S. Pinchuk**, A. G. Kravtsov**, Zh. V. Kadolich**, and V. I. Nikolaev*

* Gomel State Medical University, Gomel, Belarus
** Belyĭ Institute of Mechanics of Metal-Containing Polymer Systems,
National Academy of Sciences of Belarus, Gomel, 246050 Belarus
e-mail: mpri@mail.ru
Received September 20, 2004

Abstract—Main medicinal preparations (MPs) administered by injections into a joint cavity are studied by means of electret–thermal analysis and tribometry. In the latter case, the MPs are used as lubricants in a friction pair (a model of an artificial joint) exposed to an electromagnetic field simulating the biological field. Changes in the thermally stimulated current spectra from the MPs exposed to the electromagnetic field and those in the friction coefficient of the pair provide a deeper insight into the physicochemical mechanisms of MP action. Comprehensive tests of MPs involving electret–thermal and tribological analyses may be of great practical significance for optimizing local treatment of diseased joints. © 2005 Pleiades Publishing, Inc.

INTRODUCTION

Of the variety of approaches to the treatment of patients with diseases or injuries of joints, injections of medicinal preparations (MPs) directly into the joint cavity are considered to be the most effective [1, 2]. MPs are chosen individually with due regard for the general state of the patient, anamnesis, and the course of disease. Due to recent advances in pharmacology, a broad range of MPs for treating joint diseases is available today [3]. For example, antibiotics and corticosteroid drugs are used to suppress inflammation, and progressing cartilage destruction with degenerative or dystrophic processes in the joints is treated by means of MPs functioning as protectors and lubricants, i.e., similarly to the natural synovial fluid. Being injected into the joint cavity, MPs not only exert a directed therapeutic effect but also, mixing together with the synovial fluid, qualitatively change its biomechanical properties, primarily the lubricating ability. Preparations with the same pharmacological mechanism of action may have different effects on the parameters of friction in joints.

As does any biological system, the tissues of a joint generate electromagnetic field (EMF) and respond to its action [4, 5]. The EMF energy influences the tissue structure and biophysical mechanisms governing the specific functions of the synovial fluid [6, 7]. There are reasons to believe that the electrophysical and biomechanical properties of MPs injected into the joint cavity are interrelated, and elucidation of this relation may help to optimize the treatment of joint diseases.

The purpose of this study is to perform for the first time tribological "monitoring" of some MPs routinely used in rheumatological and orthopedic practice and to reveal the interplay between their tribological and polarization parameters.

INVESTIGATION METHODS

We experimented with MPs of different pharmacological groups (table) chosen by the following criteria: knowledge of the procedure of treatment among practicing physicians, accessibility, and pronounced clinical effect.

For the test conditions to be as close to the friction in a natural joint as possible and for a greater accuracy of measurements, we used a pendulum-type tribometer with a single (test) friction pair. The coefficient of friction was estimated from the logarithmic decrement of damping of pendulum's oscillations. The friction pair consisted of a grooved plate made of superhigh-molecular-weight polyethylene certified for orthopedic applications and a pendulum installed on a trihedral prism made of 12Kh18N9 chrome–nickel steel. The base (bearing) of the prism was rounded with a radius of 2 mm. The load created in the contact zone by the 2-kg pendulum moving at a (sliding) rate of 1.0 m/s was equal to the average physiological load on the human knee joint. To simulate the biophysical field of a joint, the bearing of the pendulum was placed into a solenoid (the outer diameter 21 mm, 600 ± 2 coils of 0.07-mm copper wire) connected to a dc power source. The magnetic field intensity at the friction surface averaged 1.2 kA/m, being within the optimal range of magnetic field intensities used in joint magnetotherapy [8].

To estimate the polarization of MPs as pseudoelectrets consisting of coordination-bonded polar groups,

Medicinal preparations used in the study

Name	Composition	Manufacturer	Mechanism of action
Hydrocortisone	Hydrocortisone acetate, 125 mg; lidocaine hydrochloride, 25 mg	Gedeon Richter AG, Hungary	Antiinflammatory
Kenalog-40	Triamcinolone acetonide, 40 mg; benzyl alcohol, 9.9 mg	Bristol-Myers Squibb SpA, Italy	"
Dyprospan	Betamethasone dipropionate, 6.43 mg; betamethasone sodium phosphate, 2.63 mg	Shering-Plau, Germany	"
Lincomycin hydrochloride	Lincomycin hydrochloride, 30% solution	Borisov Pharmaceutical Plant, Belarus	Antimicrobial
Synvisk	Hylan A and Hylan B, 8.0 mg/ml	Biomatrix, Inc., United States	Synovial fluid substitute
Hyalart	Hyaluronic acid in 0.9% sodium hydrochloride solution, 2 ml	Bayer AG, Germany	"
Diasynol	1.5% carboxymethylcellulose sodium salt and 1.0% Tecon-20 in 0.9% sodium hydrochloride solution	Belbiofarm Plant of Diagnostic and Pharmaceutical Preparations, Belarus	"

they were subjected to electret–thermal analysis (ETA). A 0.2-ml sample of an MP was placed on a degreased brass electrode and covered by an insulating Teflon spacer, on which the second electrode was placed. During heating of the sample at a rate of 5°C/min, a current of about 10^{-12} A was detected in the electrode circuit. The spectra of these thermally stimulated currents (TSCs) were mathematically processed (digital filtration with the Origin 5.0 application program) and displayed as $I(T)$ plots.

RESULTS AND DISCUSSION

Tribometry data showed that, all other conditions being equal, the initial value of friction coefficient μ_0 in the test pair depends on the chemical composition and density of the MP.

The lowest value of μ_0 was observed for Hyalart, the synovial fluid "endoprosthesis" (substitute), which is used mainly for treating the degenerative joint disease. Synvisk, another MP of the same group, had higher μ_0 and, therefore, was inferior to Hyalart as a lubricant. Differences in their lubricating ability could be found visually when the sample was placed in the tribometer. Synvisk, a thick viscous gellylike substance, did not spread over the bearing and was unevenly distributed in the groove even after unloading. Hyalart, being more watery, uniformly wetted the friction surface made of extremely hydrophobic superhigh-molecular-weight polyethylene, thus favoring sliding of the pendulum. It appears that the clinical effect of treatment with Synvisk [9] is due to its viscoelastic properties (providing for cartilage protection against peak mechanical loads) and stimulation of synovial fluid production, rather than to improved conditions for sliding in the joint.

Diasynol is another substitute of the synovial fluid that was prepared by Belarussian scientists [7]. Its consistency is similar to that of Hyalart. The formula of

Diasynol includes thermotropic liquid-crystal cholesterol compounds, which makes it a good lubricant [10].

Steroid MPs with a strong antiinflammatory effect (Kenalog-40, Dyprospan, and Hydrocortisone) have high μ_0. This may be attributed to the fact that these suspensions are subject to phase separation immediately at the instant they are drawn from ampoules with a syringe. In the ascending order of the time of crystalline residue formation upon drying, these MPs are arranged as Hydrocortisone < Kenalog-40 < Dyprospan. The highest friction coefficient was observed for Hydrocortisone. This MP exhibited rapid phase separation followed by the displacement of the liquid phase from the contact zone (groove) of the friction pair, whereupon the remaining coating consisted mainly of the crystal-

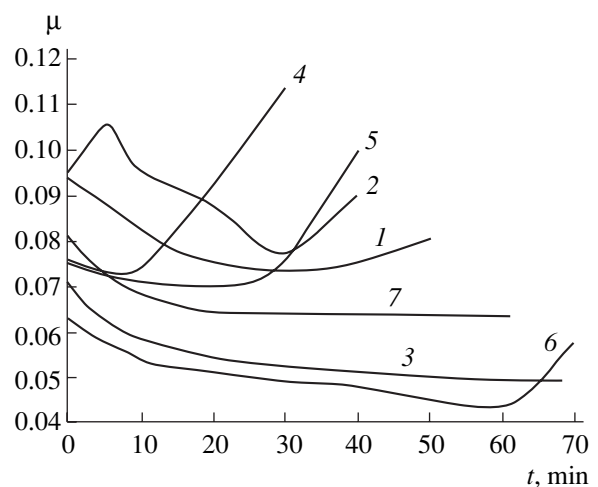**Fig. 1.** Time variation of friction coefficient μ in experiments with MPs exposed to the EMF in the friction pair of the tribometer: (*1*) Hydrocortisone, (*2*) Kenalog-40, (*3*) Dyprospan, (*4*) Lincomycin, (*5*) Synvisk, (*6*) Hyalart, and (*7*) Diasynol.
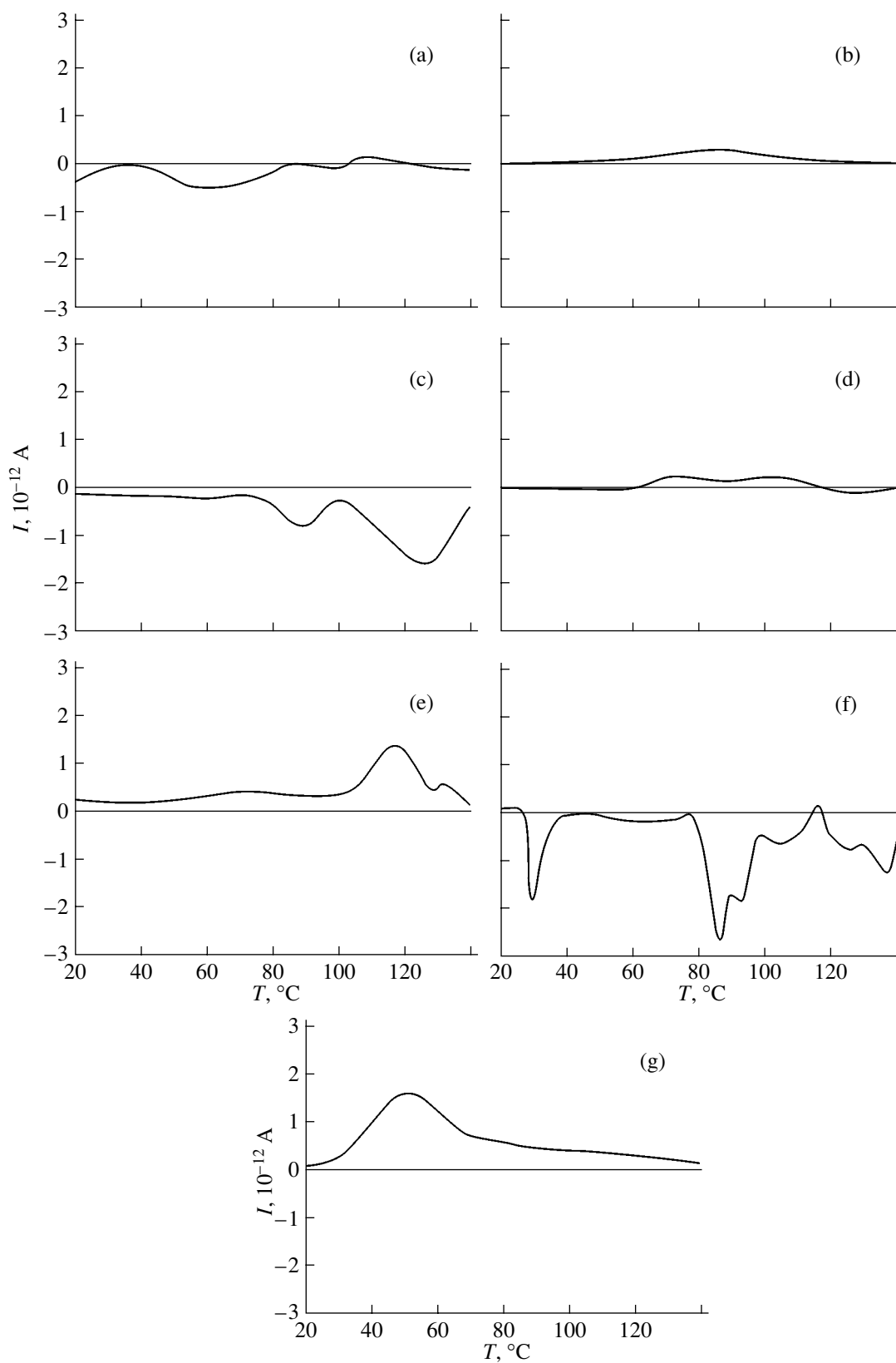
Fig. 2. TST spectra taken of (a) Hydrocortisone, (b) Kenalog-40, (c) Dyprospan, (d) Lincomycin, (e) Synvisk, (f) Hyalart, and (g) Diasynol.

line phase acting as an abrasive rather than as a lubricant. A similar picture was observed when the bearing of the tribometer was lubricated with Kenalog-40. Dyprospan, a suspension of finer particles with a markedly lower sedimentation rate, had a relatively high lubricating ability.

Lincomycin hydrochloride, an antibiotic prescribed in cases of inflammation caused by bacterial infection, is prepared in the form of a saline solution, which rapidly crystallized in the course of tribometry. One may suppose that Lincomycin, being mixed with the synovial fluid *in vivo*, will not have a high abrasive effect on the articular surfaces, since it contains no water-insoluble components and readily enters the circulatory system through the synovial membrane.

Figure 1 shows the time variation of the friction coefficient in the EMF for different MPs. In the case of Hyalart and Dyprospan, the exponential decrease in this coefficient is apparently due to a high sensitivity of their polar groups to the EMF and ordering of these groups along the lines of force, which may improve the lubricating properties.

As was noted above, the friction coefficient in the case of Diasynol is low, because a liquid-crystal layer with a high shear strength forms in the contact zone [7]. EMF-induced ordering raises the lubricating ability of this layer, lowering μ still further.

The lubricating ability of Kenalog-40 and Hydrocortisone exposed to the EMF remained the same, presumably because of a low sensitivity of their large and heavy particles to the field. Yet, these MPs respond to the field, indicating that they incorporate polar groups and electrically nonequilibrium structures.

The above differences can be explained on the basis of specific electrophysiological properties noticed in the course of ETA. We managed for the first time to take and identify the TSC spectra of MPs administered by injections into the joint cavity. The phenomenon of the quasi-electret state was previously demonstrated for biological protein systems, such as blood [11] and synovial fluid [12]. The results shown in Fig. 2 confirm that the quasi-electret state is also characteristic of the above MPs.

In the experiment with Hyalart (Fig. 2f), three TSC peaks were recorded: a low-temperature peak at 30°C, a medium-temperature peak at 85°C, and a high-temperature peak at about 140°C. We assign the first one to the generation of charge carriers upon thermal destruction of hydration sheaths around the polar groups of hyaluronic acid molecules. The medium- and high-temperature peaks seem likely to arise because of the decomposition of supermolecular structures and the thermal destruction of the molecules themselves. The quasi-electret properties of the biopolymer components entering into Hyalart suggest their increased sensitivity to the EMF; in particular, the field may pass these components to a tribologically optimal state.

The quasi-electret state is also characteristic of Synvisk (Fig. 2e), but its TSC spectrum has only one high-temperature peak ($\approx$115°C). Apparently, its biopolymer components are characterized by a low polarity, poor ability to form coordination compounds with water molecules, and low sensitivity to the EMF under friction. This is confirmed by the degradation of Synvisk lubricating ability under the EMF in the tribological experiment (Fig. 1).

The TSC spectrum of Diasynol (Fig. 2g) has one low-temperature peak ($\approx$50°C) reflecting a heat-induced reconfiguration in the liquid-crystalline phase, the basic lubricating component of this MP. This correlates with the behavior of the friction coefficient for Diasynol in the EMF (Fig. 1, curve *7*): μ decreases proportionally to the time of EMF influence on the lubricating coating. The liquid-crystal cholesterol compounds seem likely to reconfigure into a tribologically optimal structure under the action of the field.

Unlike Hydrocortisone and Kenalog-40, not exhibiting the quasi-electret effect, Dyprospan has two peaks in the TSC spectrum: at 90 and 130°C (Fig. 2c). They are accounted for by thermal destruction of two types of betamethasone derivatives (table). Presumably, the pattern of TSC spectra for steroid antiinflammatory MPs depends on the polarity and binding energy of their molecular components: the higher their polarity, the more probable the coordination interaction between these components and, consequently, the more pronounced the quasi-electret properties of these MPs.

No TSC appears in the Lincomycin hydrochloride solution upon heating (Fig. 2d), because this MP contains no polarized molecules and molecules capable of forming coordination structures.

CONCLUSIONS

A comprehensive analysis of basic MPs administered by injections into the synovial joints has confirmed the hypothesis for a correlation between their lubricating ability and coordination bonding between their components, which impart quasi-electret properties to these substances. The reparations containing EMF-sensitive components are the best lubricants, irrespective of the pharmacological groups they belong to. The presence of such electrically nonequilibrium structures in these substances has been established by means of ETA and is described for the first time. In our opinion, further experimental studies on the tribological properties of preparations administered in this way will help to optimize methodological approaches to the treatment of joints. A compromise between electret–thermal research and tribological analysis, which is routine in mechanics but new in medicine, would allow specialists to produce new-generation MPs optimally combining a specific therapeutic effect with a good lubricating ability.

REFERENCES

1. N. F. Soroka, *On Rheumatic Arthritis: Questions and Answers* (Belarus', Minsk, 1992) [in Russian].

2. S. Shutsyanu, V. Ionescu-Blaja, and M. Moange, *Clinical Diagnostics and Medical Treatment of Rheumatic Diseases* (Med. Izd., Bucharest, 1983) [in Russian].

3. M. D. Mashkovskiĭ, *Medicinal Agents: A Handbook* (Meditsina, Moscow, 1993), Vol. 1 [in Russian].

4. A. I. Lakomkin and I. F. Myagkov, *Electrophysiology* (Vysshaya Shkola, Moscow, 1977) [in Russian].

5. A. A. Remizov, *Medical and Biological Physics* (Vysshaya Shkola, Moscow, 1999) [in Russian].

6. G. E. Grigoryan, *Magnetoreception and the Mechanisms of Magnetic Effects on Biosystems* (Gitutyun, Erevan, 1999) [in Russian].

7. S. F. Ermakov, V. G. Rodnenkov, E. D. Beloenko, *et al.*, *Liquid Crystals in Engineering and Medicine* (Asar, Minsk, 2002) [in Russian].

8. V. S. Ulashchik, *Introduction to the Theory of Physiotherapy* (Nauka, Minsk, 1981) [in Russian].

9. *Clinical Aspects of Synvisc Application* (Margo Anterior, 2002), Vol. 4, pp. 5–8.

10. S. F. Ermakov, Doctoral Dissertation (Gomel, 2001).

11. L. S. Pinchuk, A. G. Kravtsov, and S. V. Zotov, Zh. Tekh. Fiz. **71** (5), 115 (2001) [Tech. Phys. **46**, 620 (2001)].

12. L. S. Pinchuk, V. I. Nikolaev, and E. A. Tsvetkova, *Endoprosthetics of Joints: Technical and Medical–Biological Aspects* (IMMS NANB, Gomel, 2003) [in Russian].

Translated by N. Gorgolyuk

Excitation of Low-Frequency Oscillations
Via Interaction of a High-Current Relativistic Electron Beam
with a Radial Ion Flux

V. A. Balakirev, N. I. Onishchenko, and I. N. Onishchenko

Kharkov Institute of Physics and Technology, National Scientific Center,
Akademicheskaya ul. 1, Kharkov, 61108 Ukraine
e-mail: meddv@grv.ifmo.ru
Received March 16, 2004; in final form, September 21, 2004

Abstract—The nonlinear excitation of low-frequency oscillations in the case when an ion flux is radially injected into the drift chamber where a tubular relativistic electron beam propagates is studied. A mechanism behind low-frequency ion oscillations is discussed. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

Low-frequency (LF) ion oscillations play a great part in the dynamics of high-current relativistic electron beams (REBs) of long duration (a microsecond or more). These processes are the most pronounced in plasma-assisted microwave oscillators (e.g., plasma-anode vircators [1], pasotrons [2, 3], and ion-drag accelerators [4]). In plasma-anode vircators, intense ion fluxes are generated directly at the anode. Depending on the acceleration mechanism in ion-drag accelerators, the ions may enter the operating system in the longitudinal direction (Luce diodes); be generated in the operating space via residual gas ionization; or enter the drift chamber in the radial direction from the edge plasma sheath, which forms when a part of the REB strikes the side wall of the drift chamber [5]. In an ion-drag accelerator based on space–time REB modulation [5], the edge plasma acts as a source of ions moving in the radial direction. When moving in the field of a spatial REB, the ions execute radial oscillations, which, in turn, generate LF oscillations of the electric field and ion density. Eventually, there arises a need for LF modulation of the REB, which is the case in the collective method of acceleration [5]. In this work, using a simple physical model, we study mechanisms behind LF ion oscillations when the ion flux is radially injected into the space of REB propagation.

PROBLEM DEFINITION AND BASIC EQUATIONS

Let a tubular REB of inner radius r_1 and outer radius r_2 propagate in an infinitely long cylindrical metallic drift chamber of radius a. The entire system is placed in a magnetic field. The beam electrons are magnetized, and the effect of the magnetic field on the ion motion is

neglected. The latter implies that the condition

$$n_{e0} M c^2 \gg \frac{H_0^2}{4\pi}$$

is satisfied. Here, n_{e0} is the REB density, M is the ion mass, H_0 is the applied magnetic field strength, and c is the speed of light. It is easy to check that inequality (5) is met in the experiments performed in [5].

Let $j_i(t)$ be the density of the radial ion flux on the surface of the drift chamber ($r = a$). Under the action of forces due to the space charge of the beam, the ions will execute radial oscillations. The electric field of the space charge will be written under the assumption that the density and velocity of the beam are given. The dynamics of the ionic component will be described on a totally self-consistent basis. The radial electric field of the space charge of a tubular REB is given by

$$E_r = -\frac{2I_b}{v_e a}\frac{1}{R}F(R),$$

where I_b is the REB current, v_e is the REB velocity, and

$$F(R) = \begin{cases} 1, & R \geq R_2 \\ \dfrac{R^2 - R_1^2}{R_2^2 - R_1^2}, & R_2 \geq R \geq R_1, \quad R = \dfrac{r}{a}, \quad R_{1,2} = \dfrac{r_{1,2}}{a} \\ 0, & R \leq R_1. \end{cases}$$

The field of the ion space charge will be described in terms of Lagrangean variables. The charge density of an infinitely thin cylindrical ionic sheath entering into the drift chamber at time t_0 is given by

$$d\rho_i = I_i(t_0)dt_0\frac{\delta(r - r_L(t, t_0))}{2\pi r_L(t, t_0)},$$
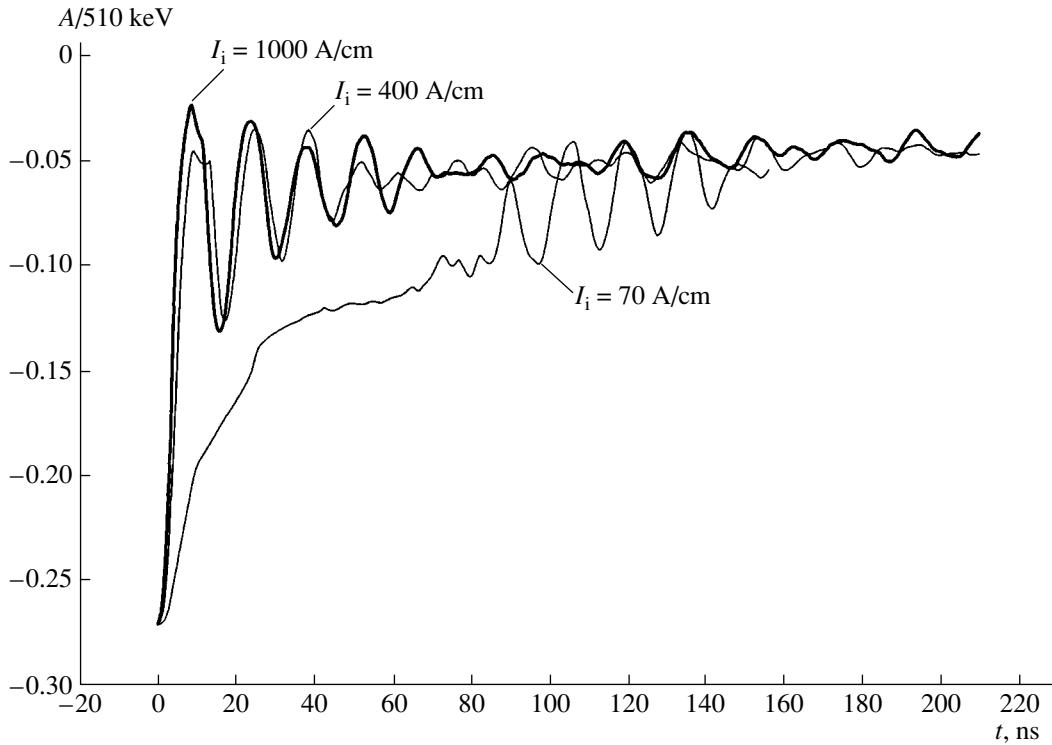
Fig. 1. Time dependence of the potential at the outer boundary of the tubular REB for different values of the ion current. $I_b = 4.6$ kA, $E_e = 280$ keV, $E_i = 25$ keV, $r_1 = 1.4$ cm, $r_2 = 1.7$ cm, and $a = 1.5$ cm.

where $I_i = j_i(t_0)2\pi a$ is the total injection current of the ion flux per unit length of the system and $r_L(t, t_0)$ is the trajectory of an ionic sheath entering into the drift chamber at time $t = t_0$ ($r_L(t_0, t_0) = a$).

The radial electric field produced by an infinitely thin ionic sheath is described by the equation

$$\frac{1}{r}\frac{d}{dr}rE_{s,r} = 2I_i(t_0)dt_0\frac{\delta(r - r_L(t, t_0))}{r_L(t, t_0)},$$

from which we obtain

$$E_{s,r} = \frac{2I_i(t_0)dt_0}{r}\chi(r - r_L). \qquad (1)$$

An expression for the total electric field is found by integration of (1) over the injection time,

$$E = \frac{2I_{0i}}{r}\int_0^t\Psi(t_0)\chi(r - r_L(t, t_0))dt_0, \qquad (2)$$

where

$$\chi(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases}$$

is the unit Heaviside function.

The ion current can be represented as $I_i = I_{0i}\Psi(t_0)$, where $\Psi(t_0)$ is a function that takes into account the shape of the injected ion current pulse ($\max\Psi(t_0) = 1$).

The dimensionless set of equations for the ion motion in the self-consistent field of the ion flux and in the field of the tubular REB has the form

$$\frac{d^2R_L}{d\tau^2} + \frac{1}{R_L}F(R_L) = \frac{\alpha}{R_L}\int_0^\tau\Psi(\tau_0')\chi(R_L - R_L')d\tau_0', \quad (3)$$

where $\tau = \omega_0 t$ is the dimensionless time,

$$\omega_0 = \frac{c}{a}\sqrt{\frac{2mI_B}{M I_A}\frac{c}{v_e}}$$

is the characteristic ion oscillation frequency in the electric field of the REB space charge, c is the speed of light, $I_A = 17$ kA,

$$\alpha = \frac{I_{0i}}{I_b}\frac{v_e}{\omega_0};$$

$R_L' \equiv R_L(\tau, \tau_0')$, and m is the electron mass.

The boundary conditions to Eq. (3) are as follows:

$$R_L(\tau_0, \tau_0) = 1, \quad \frac{dR_L}{d\tau}\bigg|_{\tau = \tau_0} = -U_{0i},$$
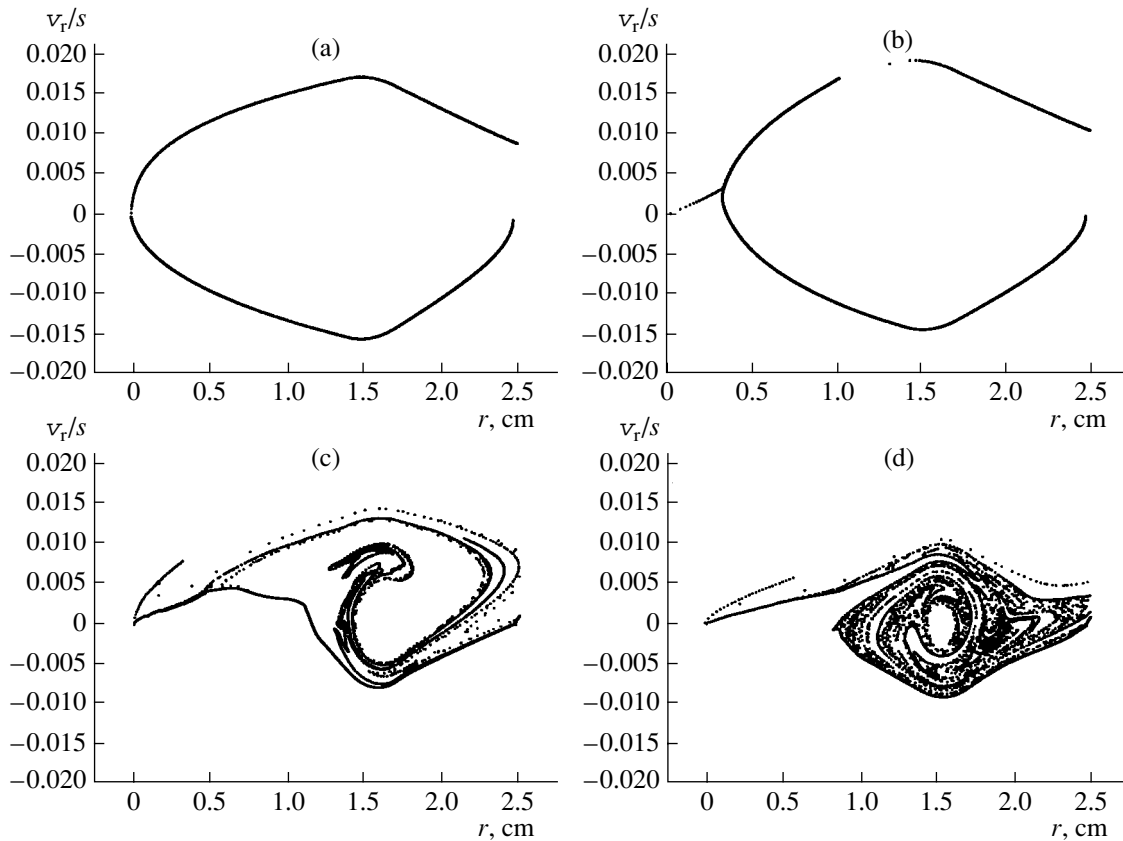
Fig. 2. Phase portraits of the ions at $t =$ (a) 19.2, (b) 23.5, (c) 107, and (d) 156 ns. $I_{0i} = 70$ A/cm.

where $U_{0i} = v_{0i}/a\omega_0$ is the dimensional initial radial velocity of the ions.

Knowing ion trajectories $R_L(\tau, \tau_0)$, one can find the dimensionless potential in the drift chamber by the formula

$$\frac{e\Phi}{mc^2} = -\frac{c}{\omega_0}\frac{I_{0i}}{I_A}\int_0^\tau d\tau_0 L(R, R_L(\tau, \tau_0)),\qquad(4)$$

where

$$L(R, R_L) = \begin{cases} \ln R, & R > R_L \\ \ln R_L, & R < R_L. \end{cases}$$

ANALYSIS OF NUMERICAL RESULTS

Equations of motion (3) and electric potential (4) were obtained numerically for different ion injection currents. The REB parameters were fixed: $I_b = 4.6$ kA, electron energy $E_e = 280$ keV, REB inner radius $r_1 = 1.4$ cm, and REB outer radius $r_2 = 1.7$ cm. To be definite, we considered hydrogen ions. The initial ion energy was $E_i = 25$ keV; the drift chamber radius, $a = 2.5$ cm. The ion current injected into the system is constant, $\Psi(\tau_0) = 1$.

Figure 1 shows the time dependence of the potential at the outer boundary of the beam for three values of ion current I_{0i} (A/cm). For the low-current ion flux ($I_{0i} = 70$ A/cm), oscillations are excited roughly 80 ns after the beginning of injection. For the high-current flux ($I_{0i} = 400$ and 1000 A/cm), oscillations start well before than in the case of low-current flux. For the ion current values, the oscillations decay with time. The decay time and the steady-state level are almost independent of the ion flux current. At the high currents, the time behavior of the LF oscillations is the nearly same. To gain greater insight into the behavior of the LF oscillations at different ion flux currents, let us turn to phase portraits (v_r/c, r) of the ions, where v_r is the ion radial velocity.

Figure 2 shows the phase portraits of the ions for the low-current beam. From the very beginning of injection, the ions move in the field of the REB space charge toward the axis of the system. After crossing the axis, the ion flux is divided into convergent and divergent radial fluxes. At the axis of the system, their velocities are equal in magnitude and opposite in sign. The deceleration of the ions in the self-field of the space charge near the axis decreases the velocities of the fluxes at the axis. Eventually, at $t = 19.3$ ns, the velocities of the convergent and divergent fluxes at the axis vanish (Fig. 2a).
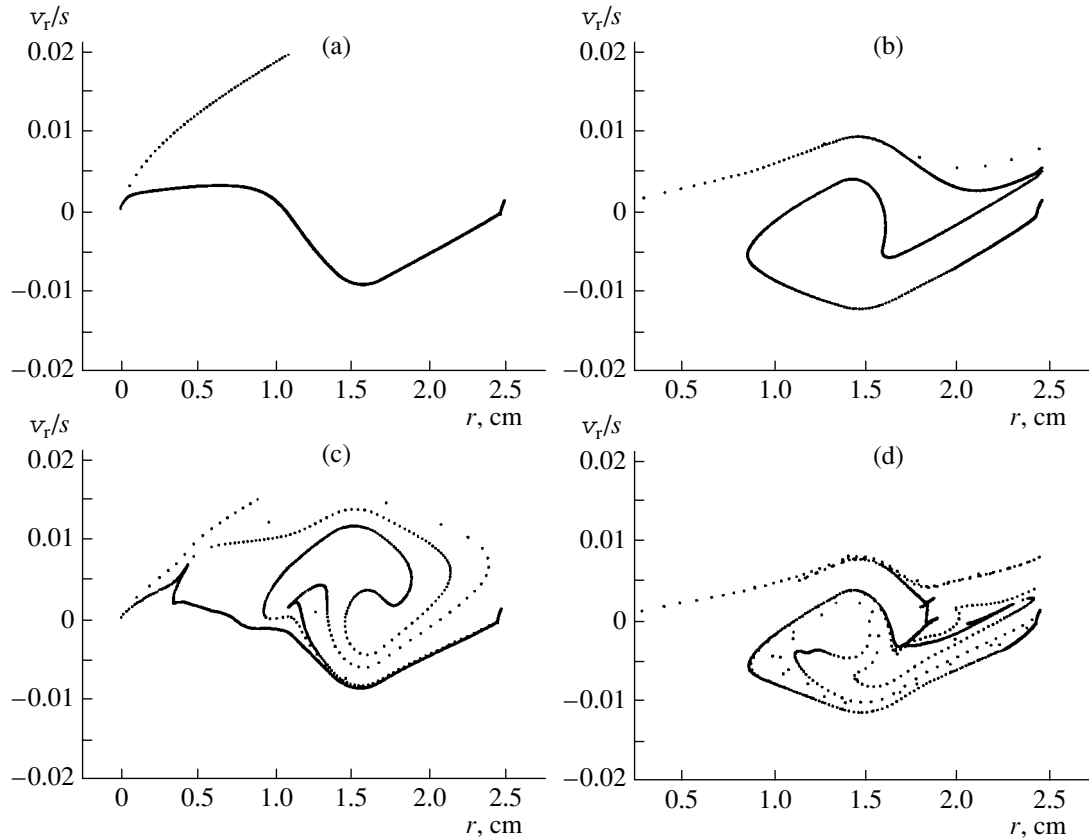
Fig. 3. Phase portraits of the ions at $t =$ (a) 10.6, (b) 17.2, (c) 25, and (d) 32 ns. $I_{0i} = 400$ A/cm.

Because of the continuing deceleration of the convergent flux, the point where the incident flux stops becomes displaced from the axis. The ions that are between the axis and the stopping point are accelerated and give particles to the divergent flux. As a result, a three-stream flux forms at the axis on the phase plane: two divergent fluxes and one convergent flux (Fig. 2b). Having approached the drift chamber, the ions of the divergent fluxes leave the system and the particles of the convergent flux are accumulated at the axis, causing a rapid growth of the potential. Subsequently, the ions of the divergent fluxes produce a bunch, which raises the potential at the site of ion flux injection when approaching the chamber edge. Eventually, at $t = 74$ ns, a virtual ionic anode (VA) arises in the ion flux injected. This cathode is nonstationary. The transmitted and reflected currents make its position oscillate with a frequency that is much higher than the potential oscillation frequency. The ion bunch is partially lost at the wall of the drift chamber, while the rest of ions reflect from the potential barrier and start moving toward the axis of the system (circulation in the phase plane). The potential reduces again, and the VA disappears. The circulation of the ion bunch is accompanied by ion losses at the wall of the drift chamber and simultaneously give rise to new bunches. The ions near the wall (a part of which leaves the system) and the continuously injected

ions produce a potential barrier near the wall. At the same time, the ions that are permanently present near the axis produce a potential barrier at the center of drift chamber.

Ion losses at the wall are compensated for by an increase in the emission current from the ionic VA. These processes go in parallel with the LF oscillations of the electric potential. The circulation of the bunches in the phase plane, which is accompanied by a continuous generation of new bunches (bunch disintegration), causes the formation of a complicated multistream flux (Figs. 2c and 2d) and, eventually, turbulizes the radial ion flux. This turbulence is the reason for the phase mixing of the particles and for the decay of the coherent ion oscillations of the potential. The steady state is established, which is characterized, for example, by a constant number of ions in the system. The ion flux toward the wall is counterbalanced by the flux from the stationary VA, which is invariably present at the wall (Fig. 2d).

In the case of the high-current ion beam ($I_{0i} = 400$ A/cm), an ionic VA forms very rapidly and is then present all the time. The phase portraits for the high-current ion flux, which are shown in Fig. 3, are taken at time instants when the first two maxima and minima appear in the time dependence of the potential. The

boundaries of the tubular REB correspond to $r_1 = 1.4$ cm and $r_2 = 1.7$ cm on the abscissa. The first maximum of the potential ($t = 10.6$ ns) appears at the stage of bunch formation, when most of the ions are inside the REB. At the time of potential minimum ($t = 17.2$ ns), two bunches are distinctly seen in the phase plane, with their major part being outside the REB. The particles of the bunches are partially lost at the walls. The circulation of the bunches causes them to shift toward the center of the drift chamber (Fig. 3c) and increases the REB potential. At $t = 25$ ns, the potential at the REB outer boundary reaches a maximum. At the time of the second minimum ($t = 32$ ns), the bunches are on the circumference again and some of the ions are lost at the wall. It should be noted that the escape of the ions from the drift chamber is accompanied by an increase in the ion current entering into the system and, accordingly, by a decrease in the VA-reflected ion current. The phase mixing of the particles results in the steady state, as in the case of the low-current beam.

CONCLUSIONS

We considered the excitation of LF ion oscillations when an ion flux is radially injected from the drift chamber wall into the space where a high-current REB propagates. For a low-current ion current ($I_{0i} = 70$ A/cm), an ionic VA at the wall does not form and a two-stream flux arises. The interaction of the ion fluxes opposing each other results in the formation of ion bunches, which, when circulating in the phase plane, continuously disintegrate. Eventually, the accumulation of the ions in the system generates an ionic VA near the lateral surface. The reason for the occurrence of the LF oscillations is the formation of ion bunches and their coherent radial oscillations in the potential well. The decay of the LF oscillations is associated with bunch disintegration during the radial oscillations of the bunches and also with the phase mixing of the ions. Ultimately, the steady state is established, which is characterized by a constant number of ions in the system. The ion losses at the wall are compensated for by the ions injected into the system from the VA.

In the case of high-current ion fluxes ($I_{0i} = 400$ and 1000 A/cm), a VA forms very rapidly. The pattern of LF oscillations under the VA conditions is virtually independent of the ion injection current. The fact is that an ionic VA is an emitter whose current is limited by the space charge. The transmitted ion current depends on the REB current and the dimensions of the drift chamber. When the ion flux current is high, the time taken to establish steady-state oscillations is significantly shorter and the oscillation amplitude is higher than in the low-current case.

REFERENCES

1. A. L. Babkin, A. E. Dubinov, V. S. Zhdanov, *et al.*, Plasma Phys. Rep. **23**, 316 (1997).
2. Yu. P. Bliokh and G. S. Nusinovich, IEEE Trans. Plasma Sci. **29**, 951 (2001).
3. Yu. P. Bliokh, G. S. Nusinovich, J. Felsteiner, and V. L. Granatstein, Phys. Rev. E **66**, 056503 (2002).
4. A. E. Dubinov, I. Yu. Kornilova, and V. D. Selemir, Usp. Fiz. Nauk **172**, 1225 (2002) [Phys. Usp. **45**, 1109 (2002)].
5. V. A. Balakirev, A. M. Gorban', I. I. Magda, *et al.*, Fiz. Plazmy **23**, 350 (1997) [Plasma Phys. Rep. **23**, 323 (1997)].

Translated by V. Isaakyan

SHORT
COMMUNICATIONS

Doped ZnS Phosphors with a Constant Spectral Density in the 500–750 nm Range

Yu. Yu. Bacherikov and N. V. Kitsyuk

Institute of Semiconductor Physics, National Academy of Sciences of Ukraine,
pr. Nauki 45, Kiev, 03028 Ukraine

e-mail: Yuyu@isp.kiev.ua

Received April 21, 2004

Abstract—A ZnS : (In, Cu, Cl) electrophosphor that emits with an almost constant spectral density at wavelengths λ in the range $550 < \lambda < 750$ nm and a ZnS : In photophosphor exhibiting the same property in the range $500 < \lambda < 700$ nm are prepared. © 2005 Pleiades Publishing, Inc.

Doping of ZnS by various impurities makes it possible to obtain electrophosphors emitting in the green, red, blue, and other parts of the spectrum [1, 2]. Superposition of these emissions allows creation of an electroluminescent white indicator. However, today, the problem of a phosphor material whose spectrum contains three basic colors, which, when combined, give white light, still remains on the agenda. A much more challenging problem in this field is to find a radiation source offering a continuous spectrum with a constant spectral density throughout the visible range. Solving this problem would make it possible to remedy a number of engineering bottlenecks arising in designing light-absorbing and light-emitting devices [1, 3]. In this paper, we report on electroluminescence from a ZnS : (In, Cu, Cl) powder with a constant spectral density in the range 550–750 nm and photoluminescence from ZnS : In powder with a constant spectral density in the range 500–700 nm.

Doping of as-prepared ZnS powders was carried out by thermal annealing under a low air pressure in the presence of metallic In and CuCl for 3 h at 800°C [4]. After annealing, the powders were passivated in air for 30 h.

The luminescence spectra were recorded with a KSVU-23 instrument at room temperature. Photoluminescence (PL) was excited by radiation from an LGI-21 nitrogen laser ($\lambda = 337.1$ nm), and electroluminescence (EL) was excited by applying a sinusoidal voltage ($U = 250$ V, $\nu = 5000$ Hz). The electroluminescence properties were studied on electroluminescent indicators (ELIs) prepared by the standard powder phosphor technology [1].

Figure 1a shows the PL spectra taken of the ZnS : In powder before and after passivation in air. The PL spectrum of the unpassivated ZnS : In (curve *1*) contains at least three bands with peaks at $\lambda_{max} = 515$, 600, and 640 nm. The origin of the luminescence band with a

peak near 515 nm is unclear. In [2, 5], it is attributed to oxygen and copper impurities in zinc sulfide. In [6], this band ($\lambda_{max} = 515$ nm) is associated with the radiation from self-activated ZnS and with the formation of anion vacancies in the sulfur sublattice. As for the band with $\lambda_{max} = 600$ nm, Morozova and Kuznetsova [2] assign it to the presence of intrinsic In, which enters into ZnS without the help of activators and coactivators.
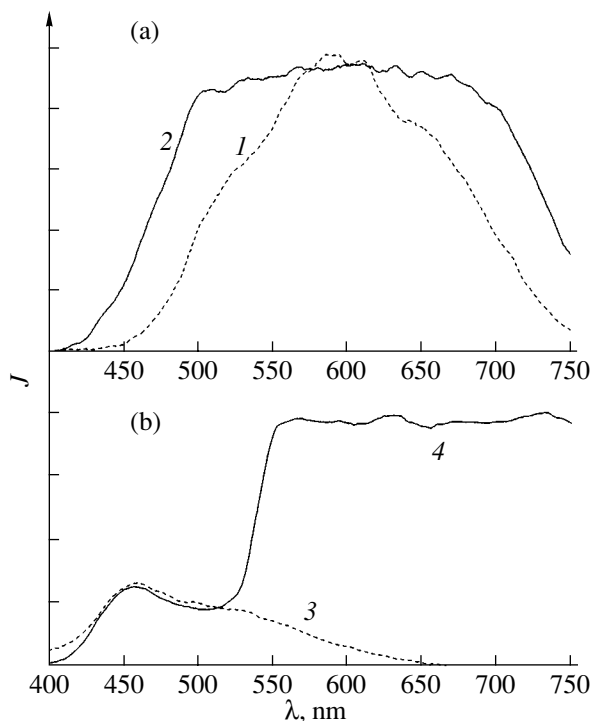

Fig. 1. (a) PL spectra of the ZnS powder annealed at 800°C in the presence of metallic In (*1*) before and (*2*) after passivation in air and (b) EL spectra ($U = 250$ V, $f = 5000$ Hz) from the ELI made of the ZnS : (In, Cu, Cl) powder (*3*) before and (*4*) after passivation in air.

It can be seen from Fig. 1a that exposure of the ZnS : In powder to air leads to the formation of a plateau extending from 500 to 700 nm (curve *2*); i.e., the spectral density of luminescence in this range becomes almost constant (within 8%). Thus, the passivation of the In-doped ZnS powder in air (oxygen seems likely to be a key passivating agent) causes either or both additional luminescence lines to emerge in this wavelength range and the already existing lines to broaden (or overlap to a greater extent). In any of the cases, additional energy levels are assumed to be involved in radiative recombination that earlier (before passivation) participated in nonradiative transitions or were absent at all.

The situation with EL is slightly different. While annealing of the ZnS powder in the presence of In, as well as subsequent passivation, does not lead to any tangible signs of EL, combined doping of zinc sulfide by metallic In and CuCl leads to the emergence of EL bands. Figure 1b (curve *3*) shows the EL spectrum of the ELI taken prior to passivation of the phosphor in air. The spectrum consists of bands with λ_{max} = 450 and 530 nm, which were repeatedly reported in the literature [2, 5–7]. These bands were attributed [2, 5, 7] to copper impurity in ZnS. As for the case of PL in ZnS : In, passivation in air causes considerable changes in the EL spectrum. Curve *4* in Fig. 1b shows the EL spectrum for the ELI made of air-passivated ZnS : (In, Cu, Cl) powder. The main feature of this spectrum is a plateau covering the 550–750 nm range, where the intensity remains constant within 5%. Note that our objective was to study EL in the visible range; for this reason, the long-wave edge of the 550–750 nm plateau in the EL spectrum should be refined.

The results obtained here are certainly insufficient that final conclusions are drawn. Sound interpretation of such nontrivial luminescence characteristics of the powders studied calls for further investigation. However, our findings indicate the possibility of preparing a continuous-spectrum ZnS : (Cu, Cl, In) electrophosphor with a constant spectral density in the range $550 < \lambda \leq 750$ nm and a ZnS : In photophosphor with the same luminescence properties in the range $500 < \lambda < 700$ nm.

REFERENCES

1. I. K. Vereshchagin, B. A. Kovalev, L. A. Kosyachenko, *et al.*, in *Electroluminescent Sources of Light*, Ed. by I. K. Vereshchagin (Énergoatomizdat, Moscow, 1990) [in Russian].

2. N. K. Morozova and V. A. Kuznetsova, *Zinc Sulfide: Production and Optical Properties* (Nauka, Moscow, 1987) [in Russian].

3. *Display Systems Engineering*, Ed. by H. Luxenberg and R. Kuehn (New York, 1968; Mir, Moscow, 1970).

4. Yu. Yu. Bacherikov, I. S. Golovina, N. V. Kitsyuk, *et al.*, in *Proceedings of the Conference on Structural Relaxation in Solids, Vinnitsa, 2003*, pp. 180–182.

5. M. Aven and J. S. Prener, *Physics and Chemistry of II–VI Compounds* (North-Holland, Amsterdam, 1967; Mir, Moscow, 1970).

6. R. Grasser, F. Scharmann, and W. Schwedes, Z. Phys. B **20**, 235 (1975).

7. L. A. Gromov and V. A. Trofimov, Zh. Fiz. Khim. **55**, 2629 (1981).

Translated by N. Wadhwa

SHORT
COMMUNICATIONS

Estimating the Energy Performance of an Insulator with Regard to the Relaxation Time Distribution

O. A. Emelyanov

St. Petersburg State Polytechnical University,
ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia
e-mail: oae@mail.wplus.net
Received July 19, 2004

Abstract—The energy relationships in the macroscopic electrodynamics of an insulator are analyzed with regard to the polarization relaxation time distribution. Expressions for the discharge power and discharge energy flux densities in an insulator are derived for an electric field exponentially depending on time. The performance of polyethylene terephthalate in capacitive energy storage systems is estimated in terms of energy. © 2005 Pleiades Publishing, Inc.

To perform comparative analysis of insulators used in today's capacitive energy storage systems, the author has derived the relationships between stored energy W_{st}, released (effective) energy W_{eff}, and lost energy W_l of the electric field under the condition of the Debye mechanism of polarization of a dielectric medium. The performance of an insulator can be estimated in terms of efficiency as follows:

$$\eta = \frac{W_{eff}}{W_{st}} = 1 - \frac{W_l}{W_{st}}, \qquad (1)$$

where $W_{eff} = W_{st} - W_l$ reflects the electric field energy balance.

A wide class of polymer insulators and layered systems used for capacitor insulation is characterized by the relaxation time spectrum, which may be both continuous and sparse. This paper is aimed at extending the energy concepts of the electrodynamics of insulators by taking into account the electric polarization relaxation time distribution.

Let us consider the differential equation for elementary increment $d_\tau P$ of relaxation polarization in relaxation time interval $d\tau$,

$$\frac{d(d_\tau P)}{dt} + \frac{d_\tau P}{\tau} = \frac{\varepsilon_0 y(\tau)}{\tau} E(t) d\tau, \qquad (2)$$

where $y(\tau) = d(\varepsilon_s(\tau) - \varepsilon_\infty)/d\tau$ is the relaxation time distribution function; ε_∞ and ε_s are the permittivities for the polarization instantaneously reaching the steady state and static polarization at given time instant τ, respectively; and E is the electric field.

With regard to the prehistory of the polarization process, a solution to Eq. (2) can be written in the form

$$d_\tau P(t, \tau) = \varepsilon_0(\varepsilon_s(\tau) - \varepsilon_\infty) \int_{-\infty}^{t} \frac{E(u) \exp\left(-\frac{t-u}{\tau}\right)}{\tau} du. \qquad (3)$$

Assuming that the relaxation process terminates at static value P_s within the time interval $[-\infty, 0]$ and integrating expression (3) over all possible values of τ, we obtain the following expression for the total relaxation polarization:

$$P_r(t) = P_s \int_0^\infty \frac{y(\tau) \exp\left(-\frac{t}{\tau}\right)}{\tau} d\tau$$

$$+ \varepsilon_0 \int_0^\infty \exp\left(-\frac{t}{\tau}\right) \int_0^t \frac{y(\tau) \exp\left(\frac{u}{\tau}\right) E(u)}{\tau} d\tau du. \qquad (4)$$

According to [1], the flux densities of the stored energy, $\partial W_{st}/\partial t$; effective energy, $\partial W_{eff}/\partial t$; and loss energy, Q, for given τ are expressed as

$$\frac{\partial W_{st}}{\partial t} = \frac{1}{\varepsilon_0(\varepsilon_s - \varepsilon_\infty)} P_r(t) \frac{\partial P_r(t)}{\partial t} + \frac{\partial}{\partial t} \frac{\varepsilon_0 \varepsilon_\infty E^2}{2},$$

$$Q(t) = \frac{\tau}{\varepsilon_0(\varepsilon_s - \varepsilon_\infty)} \left[\frac{\partial P_r(t)}{\partial t}\right]^2, \qquad (5)$$

$$\frac{\partial W_{eff}}{\partial t} = E \frac{\partial P_r(t)}{\partial t}.$$

For an exponentially decreasing field, $E = E_0 \exp\left(-\dfrac{t}{\tau_E}\right)$, we have

$$d_\tau R(t, \tau) = \frac{d_\tau P_s(\tau)}{\alpha - 1}\left[\alpha \exp\left(-\frac{t}{\tau_E}\right) - \exp\left(-\frac{t}{\tau}\right)\right],$$

$$\frac{d}{dt}[d_\tau P(t, \tau)] = \frac{d_\tau P_s(\tau)}{\tau(\alpha - 1)}\left[\exp\left(-\frac{t}{\tau}\right) - \exp\left(-\frac{t}{\tau_E}\right)\right], \quad (6)$$

where $\alpha = \tau_E/\tau$ is the ratio between the characteristic times of variation of the electric field and relaxation polarization process.

Then, the elementary increments of the corresponding powers in the course of depolarization are given by

$$\frac{\partial}{\partial t}[d_\tau W_{st}(t, \tau)] = \frac{\varepsilon_0 d(\varepsilon_s(\tau) - \varepsilon_\infty)}{2}\frac{E_0^2}{(\alpha - 1)^2}$$

$$\times \frac{\partial}{\partial t}\left[\alpha \exp\left(-\frac{t}{\tau_E}\right) - \exp\left(-\frac{t}{\tau}\right)\right]^2,$$

$$d_\tau Q(t, \tau) = \frac{\tau}{\varepsilon_0 d(\varepsilon_s(\tau) - \varepsilon_\infty)}\frac{(d_\tau P_s(\tau))^2}{(\tau_E - \tau)^2}$$

$$\times \left[\exp\left(-\frac{t}{\tau}\right) - \exp\left(-\frac{t}{\tau_E}\right)\right], \quad (7)$$

$$\frac{\partial}{\partial t}[d_\tau W_{eff}(t, \tau)] = \frac{\varepsilon_0 d(\varepsilon_s(\tau) - \varepsilon_\infty)E_0^2}{\tau_E - \tau}$$

$$\times \left[\exp\left(-\frac{t}{\tau}\right) - \exp\left(-\frac{t}{\tau_E}\right)\right]\left[\exp\left(-\frac{t}{\tau_E}\right)\right].$$

Integrating relationships (7) over time t between 0 and ∞ and over relaxation times τ within the same limits, we arrive at the following expressions for stored energy W_{st}, loss energy W_l, and effective energy W_{eff} being released in the load:

$$W_{st} = \frac{\varepsilon_0(\varepsilon_s - \varepsilon_\infty)E_0^2}{2},$$

$$W_l = \frac{\varepsilon_0 E_0^2}{2}\int_0^\infty \frac{\tau y(\tau)}{\tau_E + \tau}d\tau, \quad (8)$$

$$W_{eff} = \frac{\varepsilon_0 E_0^2}{2}\int_0^\infty \frac{\tau_E y(\tau)}{\tau_E + \tau}d\tau.$$

To obtain a final estimate of the energy characteristics, we must know a specific form of relaxation time distribution function $y(\tau)$. In a first approximation, relaxation loss factor ε'' can be determined with the Fouss–Kirkwood relationship

$$\varepsilon''(\omega) = \varepsilon''_{max}\,\text{sech}[\lambda \ln(\omega\tau)], \quad (9)$$

where ω is the frequency of the applied field, ε''_{max} is the maximum value of the loss factor, and λ is the relaxation time distribution parameter [2].

On the other hand, the frequency dependence of the loss factor with regard to the relaxation time distribution has the form [3]

$$\varepsilon''(\omega) = \int_0^\infty \frac{\omega\tau y(\tau)}{1 + \omega^2\tau^2}d\tau. \quad (10)$$

Knowing parameter λ and applying the inverse Fourier transformation to expression (10), we can find the form of function $y(\tau)$. For arbitrary dependences $\varepsilon''(\omega)$, an effective procedure for determining $y(\tau)$ was proposed in [4]. In particular, at moderate temperatures (20–30°C), parameter λ for polyethylene terephthalate (PET) assumes values from 0.48 to 0.53 [5]. Setting λ equal to 0.5, we come to the following expression for the distribution function:

$$y(\tau) = \frac{(\varepsilon_s - \varepsilon_\infty)\sqrt{\tau_0}}{\tau\sqrt{\tau}(\tau + \tau_0)},$$

$$\int_0^\infty y(\tau)d\tau = \varepsilon_s - \varepsilon_\infty, \quad (11)$$

where τ_0 is the most probable relaxation time.

Taking into account relationship (11) and the inertia-free polarization component yields final expressions for the corresponding energies,

$$W_{st} = \frac{\varepsilon_0(\varepsilon_s - \varepsilon_\infty)E_0^2}{2} + \frac{\varepsilon_0\varepsilon_\infty E_0^2}{2},$$

$$W_l = \frac{\varepsilon_0(\varepsilon_s - \varepsilon_\infty)E_0^2}{2} - \frac{\varepsilon_0 E_0^2(\varepsilon_s - \varepsilon_\infty)\sqrt{\tau_E}}{2}\frac{\sqrt{\tau_E}}{\sqrt{\tau_E} + \sqrt{\tau_0}}, \quad (12)$$

$$W_{eff} = \frac{\varepsilon_0 E_0^2(\varepsilon_s - \varepsilon_\infty)}{2}\frac{\sqrt{\tau_E}}{\sqrt{\tau_E} + \sqrt{\tau_0}} + \frac{\varepsilon_0\varepsilon_\infty E_0^2}{2},$$

hence, $W_{eff} = W_{st} - W_l$, as expected.

Let us estimate efficiency η for an exponential discharge at $E_0 = 300$ MV/m and $\tau_E = 10^{-4}$ s [6]. Under normal conditions, the value of τ_0 for PTE is equal to $\approx 5 \times 10^{-4}$ s [7], $\varepsilon_s \approx 3.2$, and $\varepsilon_\infty \approx 2.46$. Calculations by formulas (12) yield $W_{st} = 1.27$ MJ/m³, $W_l = 0.20$ MJ/m³, and $W_{eff} = 1.07$ MJ/m³; accordingly,

$$\eta = \frac{W_{eff}}{W_{st}} = 0.84. \quad (13)$$

For single (short-term) discharges, energy W_l being released as heat warms the insulator insignificantly. However, with increasing discharge time τ_E, heating may become appreciable (which is especially true for rf

discharges) and cause thermal instability in the insulator [8].

REFERENCES

1. O. A. Emelyanov, Pis'ma Zh. Tekh. Fiz. **30** (19), 40 (2004) [Tech. Phys. Lett. **30**, 964 (2004)].

2. *Electrical Properties of Polymers*, Ed. by B. I. Sazhin (Khimiya, Leningrad, 1986) [in Russian].

3. S. N. Koikov, *Physics of Insulators* (Leningr. Politekh. Inst., Leningrad, 1967) [in Russian].

4. S. A. Ktitorov, Pis'ma Zh. Tekh. Fiz. **29** (22), 74 (2003) [Tech. Phys. Lett. **29**, 956 (2003)].

5. K. Miyairi and N. Morimitsu, IEEE Trans. Dielectr. Electr. Insul. **8**, 874 (2001).

6. H. G. Wisken and Th. H. C. Weise, IEEE Trans. Magn. **39**, 446 (2003).

7. G. S. Kuchinskiy and N. I. Nazarov, *High-Power Electrical Capacitors* (Atomizdat, Moscow, 1992) [in Russian].

8. O. A. Emelyanov, Pis'ma Zh. Tekh. Fiz. **27** (9), 76 (2001) [Tech. Phys. Lett. **27**, 679 (2001)].

Translated by N. Wadhwa

SHORT COMMUNICATIONS

Chemical and Isotope Composition of Baltic Iron–Manganese Concretions

G. S. Anufriev*, L. N. Blinov**, B. S. Boltenkov*, and M. Arif**

* Ioffe Physicotechnical Institute, Russian Academy of Sciences,
ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia

** St. Petersburg State Polytechnical University,
ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia

Received August 30, 2004

Abstract—Methods of processing iron–manganese concretions from the Gulf of Finland in the Baltic Sea are reported. The relative content of their basic components is found. The concentration of helium isotopes in and the rate of growth of these concretions (7.5 mm/thousand years) are determined. It is shown that the content of the cosmic matter in the concretions can be raised by means of special chemical processing. © 2005 Pleiades Publishing, Inc.

INTRODUCTION

Iron–manganese concretions (IMCs) containing Mn, Fe, Co, Ni, Cu, and other metals in amounts appropriate for industrial processes may well be viewed as promising polymetallic ores for the chemical industry and metallurgy in the near term. Concretions taken from different seas and oceans differ significantly in chemical composition, which depends on a variety of factors, such as the geography of a basin, salinity, depth, rate of sedimentation in general and rate of growth of concretions in particular, redox properties of the medium, and interaction of d elements with oxygen- and sulfur-containing natural water. In concretions, iron and manganese are present largely as Fe^{3+} and Mn^{4+} and, to a minor extent, as Fe^{2+} and Mn^{2+}. In Baltic concretions, some metals are present in lower amounts than in deep-sea (oceanic) concretions [1–3]. However, Baltic concretions are more attractive, because they are readily available (the seam depth is as small as several tens of meters from the sea surface) and are in the immediate vicinity of industrially developed regions and technology centers.

Industrial usage of IMCs implies a preliminary estimation of their rate of reproduction in nature, which, in turn, requires the rate of growth of concretions to be known. When estimating the latter parameter by the conventional methods of nuclear geochronology [4], one runs into serious difficulties. Namely, (i) these methods are applicable under the assumption that the rate of growth of concretions is constant, which is doubtful in the case of Baltic concretions, and (ii) geologic signs indicate that Baltic concretions of a size of 2 or 3 cm grow relatively rapidly (≈ 1 cm/thousand years). The latter means that radioisotopes with a half-life of 10^4–10^6 years, which are common in nuclear geochronology, fail in predicting the rate of growth of

Baltic IMCs. This problem (determining the rate of growth of rapidly growing IMCs) may be solved using the cosmic tracer method [5, 6]. This method is based on the accretion of cosmic dust by the Earth and accumulation of the dust by sedimentary rocks. When in outer space, dust particles attract ions of the solar wind and "transport" solar helium with an isotope ratio $^3He/^4He \sim 10^{-4}$ (which is higher than in terrestrial helium, where this ratio is on the order of 10^{-8}–10^{-7} [8]) to terrestrial sedimentary rocks. Therefore, the ^3He isotope in sedimentary rocks is basically of cosmic origin and the rate of growth (r) of sedimentary rocks (including IMCs) can be found from the relationship [7]

$$r = F/\rho\, ^3He_c, \qquad (1)$$

where $F = 1.2 \times 10^{-15}$ cm^3/(cm^2 year) is the ^3He_c isotope flux [5], $\rho = 1.6$ g/cm^3 is the IMC density, and ^3He_c is the concentration of cosmic helium light isotope.

Since Baltic concretions grow very rapidly, the ^3He concentration measured in the sample is not entirely cosmic helium. In terms of a two-component model, the fraction of cosmic ^3He can be found by the formula

$$\frac{^3He_c}{^3He} = \frac{1 - \dfrac{(^3He/^4He)_{rad}}{(^3He/^4He)}}{1 - \dfrac{(^3He/^4He)_{rad}}{(^3He/^4He)_c}}, \qquad (2)$$

where $(^3He/^4He)_{rad} = 2 \times 10^{-8}$ is the isotope ratio in terrestrial (radiogenic [8]) helium, $(^3He/^4He)_c = 4 \times 10^{-4}$ is the isotope ratio in cosmic (solar) helium [9], and $(^3He/^4He)$ is the helium isotope ratio measured.

EXPERIMENTAL

IMC samples were prepared as follows. Four spherical concretions of roughly equal size ($\approx$25 mm in diameter) were mechanically crushed to produce a powder with a grain size of a fraction of millimeter. To uniformly distribute cosmic dust particles over the volume, the powder was intimately stirred. The total weight of the powder was about 20 g. Samples to be analyzed were taken from this preparation (Table 2). They were processed in a dilute hydrochloric acid + hydrogen peroxide mixture with and without adding hydrofluoric acid.

The chemical composition of the concretions was determined by means of a standard chemical technique using (at the final stage of analysis) a PC 1000 Leeman Labs echelle-based spectrometer configured with an inductively coupled plasma source. In addition, a 216 Perkin–Elmer analyzer was applied to carry out atomic absorption analysis. The relative error of the techniques does not exceed 4% [10, 11].

Isotopic analysis was performed with a high-resolution magnetic resonance mass spectrometer [12]. Helium was released from the samples upon heating them in a vacuum followed by gettering aimed at removing reactive gases [13].

RESULTS AND DISCUSSION

The amount of Fe and Mn in Baltic concretions was determined in five samples. On average, the concentra-

Table 1. Elemental composition of the concretion sample before and after the chemical processing

Element	Content, wt %		Relative content of acid-soluble form, %	
	before processing	after processing	before processing	after processing
Manganese	32.01	28.61	100	100
Iron	9.15	12.32	97.3	98.2
Silicon	5.14	4.71	14.4	5.5
Aluminum	2.7	2.4	48.1	45.8
Sodium	0.56	0.50	100	100
Calcium	1.47	1.45	88.4	89.6
Magnesium	1.28	0.93	93.8	93.5
Titanium	0.105	0.09	38.1	44.4
Potassium	0.81	0.73	100	100
Vanadium	0.012	0.004	100	100
Copper	0.010	0.016	80	44
Nickel	0.03	0.03	100	100
Zinc	0.116	0.075	60.3	93.3
Lead	<0.01	<0.01	–	–
Barium	0.235	0.227	97.9	96.9

tions of Mn and Fe in the samples were, respectively, 32 and 9% (in other words, the Mn/Fe ratio was $\approx$3.5). This value allows us to classify Baltic concretions with iron–manganese concretions [1, 2]. To study the content of ^{3}He in the concretions, we analyzed the samples for Mn, Fe, and other elements before and after chemical processing in the reagent mixture. The processing time was varied from 1 to 120 min, the weighed portion was $\approx$1 g, and the volume of the 2% ($HCl + H_2O_2$) solution equaled 200 ml. After the processing, the content of some of the metals in the fine-grain samples decreased, while that of others increased. By way of example, Table 1 presents the elemental composition of one of the samples before and after the processing (the data obtained with the PC 1000 spectrometer) and also the content of acid-soluble forms of the same elements. It is seen that the relative content of Mn, Si, Al, Mg, and V in the concretions decreased markedly, while that of Fe and Cu rose. At the same time, the content of Na, Ca, K, Ti, and Ba changed insignificantly. These quantitative results are not only of great value for developing new technological approaches. They also form a basis for tailoring the most appropriate ("soft") chemical processing conditions under which the ^{3}He isotope contained in cosmic silicate dust, which enters into IMCs, will not be lost.

It is known that the ^{3}He isotope is much less abundant in terrestrial rocks than ^{4}He [14]. That is why we eventually selected the HCl solution (at room temperature) as a processing agent. Alkaline solutions and HF here are inappropriate, since they dissolve the silicates and ^{3}He is lost.

Calculation using (2) and the experimental data in Table 1 shows that the true content of cosmic He amounts to 80% of the value measured. With this in mind, we can now find from (1) the rate of growth of the concretions in the Gulf of Finland of the Baltic Sea,

$$r = 7.5 \text{ mm/thousand years.} \qquad (3)$$

As far as we know, this is the first estimate of the rate of growth of concretions in the Gulf of Finland (the random error in determining the composition is about 10%). Previously, the rate of growth of Baltic concretions was calculated for those in the western part of the sea near the coastline of Denmark [15, 16]. The calculations were performed by various techniques, and the rate of growth was measured to be 7 and 20 mm/thousand years. Our estimate coincides with these values in order of magnitude, being closer to the lower one.

The second task to be tackled in this work was to find a method of chemical processing that carries the main constituents of the concretions (Mn and Fe hydroxides) to the solution and, thereby, increases the concentration of the cosmic matter in the insoluble residue. The basic goal here was to obtain the cosmic material as pure as possible for detailed analysis. The experimental data for the He isotope content are listed in Table 2.

Table 2. Helium isotope content in chemically processed samples

Run no.	Weighed portion, g	$^4He \times 10^6$, cm^3/g	$^3He \times 10^{12}$, cm^3/g	$^3He/^4He \times 10^7$	Note
1	0.5435	13.7 ± 1.4	1.3 ± 0.13	0.95 ± 0.15	As-prepared mixture of unprocessed crushed concretions
2	0.4442	46.8 ± 4.7	10.9 ± 1.1	2.3 ± 0.34	Etching in HCl + H$_2$O$_2$, $K = 4.35$
3	0.3229	31.8 ± 3.2	1.01 ± 0.10	0.32 ± 0.04	Etching in HCl + HF, $K = 7.05$
4	0.5908	8.5 ± 0.85	1.4 ± 0.14	1.70 ± 0.25	Etching in HCl + H$_2$O$_2$ with HNO$_3$ added, $K = 2.05$
5	0.4011	13 ± 1.3	1.3 ± 0.13	1.00 ± 0.14	As-prepared IMC powder (repeat of run 1)

Note: K is the ratio of the initial weight of the sample to the weight of the insoluble residue taken for isotopic analysis

It is seen that the concentration of the ^3He isotope in concretions from the shallow Gulf of Finland is nearly the same as in those from the deep Pacific Ocean [5]. Presumably, this indicates that the cosmic dust flux toward the Earth depends on the geographic coordinates only slightly.

It also follows from Table 2 that, when hydrochloric acid as a processing agent is free of hydrofluoric acid, the insoluble residue is much more enriched by the ^3He isotope. Further optimization of the processing agent is on the agenda.

The results of this work are in fairly good agreement with the data in [17]. Note that our experimentally found optimal concretion-processing solution turned out to be the same as that found by Kanungo [18, 19]. However, unlike Kanungo, we used dilute hydrochloric acid to which hydrogen peroxide was added.

CONCLUSIONS

The content of Mn and Fe in concretions of the Baltic Sea was found by physicochemical methods. The rate of growth of concretions in the Gulf of Finland was determined for the first time. The value of the Mn/Fe ratio obtained in the experiments indicates that the concretions studied can be classified with IMCs. Acid solutions to process these concretions were found, and optimal ("soft") processing conditions were formulated. A solution of composition HCl + H$_2$O$_2$ can be used for extracting Mn from the concretions and also for enriching the insoluble residue by light helium isotope ^3He. This offers scope for concentration of the cosmic material and, thus, for its detailed analysis.

REFERENCES

1. S. I. Andreev, *Metallogeny of Iron–Manganese Concretions in the Pacific Ocean* (Nedra, St. Petersburg, 1994) [in Russian].
2. *Morphological Types of Iron–Manganese Concretions in the World Ocean: An Atlas*, Ed. by B. Kh. Egiazarov and V. Zyk (Brno, 1990) [in Russian].
3. I. I. Volkov, *Chemistry of the Ocean*, Vol. 2: *Geochemistry of Bottom Deposits* (Nauka, Moscow, 1979) [in Russian].
4. *Isotopic Geochemistry*, Ed. by É. V. Sobotovich, E. N. Bertenskiĭ, O. B. Ts'on', and L. V. Kononenko (Énergoizdat, Moscow, 1982) [in Russian].
5. G. S. Anufriev, B. S. Boltenkov, I. I. Volkov, and I. N. Kapitonov, Litologiya i Poleznye Iskopaemye, No. 1, 3 (1996).
6. G. S. Anufriev and B. S. Boltenkov, *Russian Science: Creation at the Border of Centuries* (Nauchnyi Mir, Moscow, 2000), pp. 310–321.
7. G. S. Anufriev, A. Ya. Krylov, V. P. Pavlov, and T. I. Mazina, Dokl. Akad. Nauk SSSR **237**, 284 (1977).
8. G. S. Anufriev, Dokl. Akad. Nauk SSSR **249**, 1202 (1979).
9. G. S. Anufriev, B. S. Boltenkov, L. V. Usacheva, and I. N. Kapitonov, Izv. Akad. Nauk SSSR, Ser. Fiz. **47**, 1830 (1983).
10. L. N. Blinov, M. Arif, and G. S. Anufriev, in *Proceedings of the 8th All-Russia Conference "Fundamental Research in Technical Universities"* (SPbGPU Izd., St. Petersburg, 2004), pp. 224–225.
11. M. Arif and L. N. Blinov, Fiz. Khim. Stekla **30**, 488 (2004).
12. G. S. Anufriev, G. I. Afonina, G. A. Mamyrin, *et al.*, Prib. Tekh. Éksp., No. 3, 244 (1979).
13. G. S. Anufriev, V. N. Gartmanov, G. A. Mamyrin, and V. P. Pavlov, Prib. Tekh. Éksp., No. 1, 248 (1977).
14. J. Emsley, *The Elements* (Oxford Univ. Press, New York, 1989; Mir, Moscow, 1993).
15. S. Hlawatsch, C. D. Garde-Schonberg, F. Lechtenberg, *et al.*, Chem. Geol. **182**, 697 (2002).
16. S. Hlawatsch, T. Neumann, C. M. G. van den Berg, *et al.*, Mar. Geol. **182**, 373 (2002).
17. G. S. Anufriev and B. S. Boltenkov, in *Proceedings of the 16th Symposium on Geochemistry of Isotopes* (Moscow, 2001), pp. 13–14.
18. S. B. Kanungo, Hydrometallurgy **52**, 313 (1999).
19. S. B. Kanungo, Hydrometallurgy **52**, 331 (1999).

Translated by V. Isaakyan

SHORT COMMUNICATIONS

Dynamics of Al_2O_3 Particles Accelerated in the Field of a Submillisecond Pulse from an YAG : Er Laser

A. V. Belikov

St Petersburg State University of Information Technologies, Mechanics, and Optics,
St. Petersburg, 197101 Russia

e-mail: meddv@grv.ifmo.ru

Abstract—The effect of the laser energy density on the velocity of sapphire particles accelerated in the field of a submillisecond pulse from an YAG : Er laser is reported for the first time. An original velocity-measuring system is described. It is shown that the acceleration of sapphire particles immersed in water (suspension) requires an energy that is nearly half as great as that needed to accelerate the particles containing the water adsorbed by the surface (powder). It is found that the particle velocity is nonuniformly distributed over the period of the laser pulse action. © *2005 Pleiades Publishing, Inc.*

The idea of light pressure dates back to Kepler, who postulated in 1619 that, because of light pressure, the tail of a comet is always directed away from the Sun. The corpuscular theory of light of Newton made the idea of light pressure more plausible and stimulated many attempts to measure it experimentally. In the 18th and 19th centuries, all attempts to discover the light pressure or find a force that cannot be assigned to convection in air failed. Crookes in 1873 believed that he had discovered light pressure in a partially evacuated chamber, although actually he had invented a radiometer. Finally, the existence of light pressure free of disturbing thermal effects was proved by Lebedev in Russia and Nuchols and Hull in the United States [1]. Advances in the laser technology have caused a rebirth of interest in light pressure. As early as in 1962, Askar'yan [2] showed that an intense light beam may have a strong influence on charged and polarizable particles and that the associated forces may change sign in going through the resonant frequency of polarization. Application of such an effect for preventing the particle concentration difference, for particle transport, and for producing rarefied or condensed areas in different media was noted. At the same time, Askar'yan and Moroz [3] demonstrated that light absorption by a particle irradiated may generate three types of forces related to (i) heating and motion of the medium itself, convective drag; (ii) heating of the medium by the absorbing surface of the particle, radiometric pressure; and (iii) pressure due to the evaporation of the particle itself, light-reactive pressure. All the effects listed may exceed the light pressure by many times and show up in experiments with a high probability. It is noteworthy

that the reactive pressure may be as high as 10^{12} atm. Such a pressure may be used, in particular, for accelerating microparticles [4] to velocities on the order of 10^6–10^8 cm/s to produce micrometeors, particles that release a high local energy when colliding with a target or with each other. These particles may be applied for processing composites (including biological composites) when the potential of conventional and laser-assisted technologies is exhausted. For example, a flux of laser-accelerated Al_2O_3 particles may be applied for increasing the rate of processing the enamel of human teeth [5]. In this case, the water surrounding the particles (suspension) or adsorbed by them (powder) effectively absorbs radiation at a wavelength of 2940 nm and vaporizes, favoring the light-reactive motion of the particle. In this work, we for the first time experimentally investigate the effect of laser energy density on the velocity of hard sapphire particles accelerated in the field of a submillisecond pulse from an YAG : Er laser. The energy sufficient to accelerate the particles of different size is estimated, and the particle velocity variation with time after the laser action is determined.

In the experiments, we used a prototype of an YAG : Er free-running 100-mm-long laser of diameter 6.3 mm operating at 2940 nm. The pulse FWHM was 250 μs. The radiation passes through a CaF_2 lens and is focused into a spot of diameter 600 ± 50 μm. In this way, an energy density of 250 J/cm^2 was achieved. As hard particles, we used Swam-Blast ultrapure Al_2O_3 powder. The rated density of the particle material was 3.97×10^3 kg/m^3. The diameters of the particles were 12, 27, 40, and 160 μm. Both the suspension and pow-
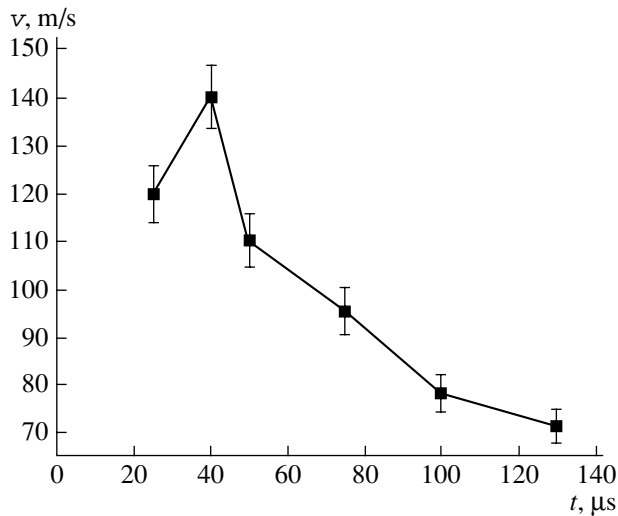
Fig. 1. Variation of the particle mean velocity with time after the YAG : Er laser pulse has been switched on (the Al$_2$O$_3$ particle diameter is 12 μm, the energy density is 135 J/cm^2).
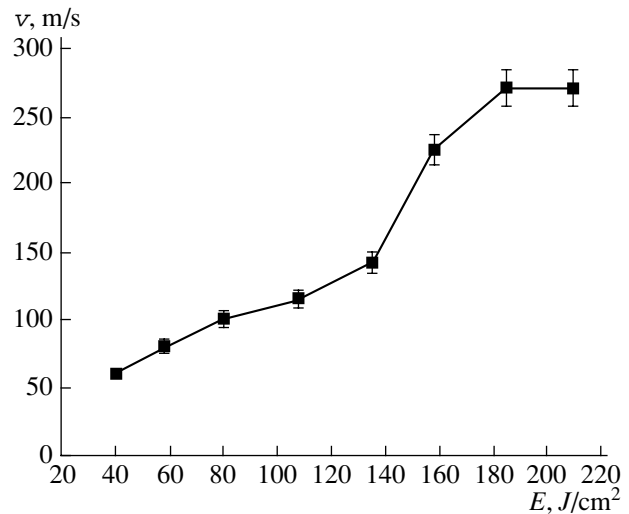
Fig. 2. Variation of the particle mean velocity with YAG : Er laser energy density (the Al$_2$O$_3$ particle diameter is 12 μm). The velocity was measured 40 μs after the laser pulse had been switched on.

der were studied. The suspension was prepared in proportion 300 mg of the particles per 0.2 ml of water. The particles (suspension or powder) were placed in a glass cell with a sapphire bottom. The thickness of the particle bed was 200 ± 50 μm. The laser radiation was focused on the contact plane between the particles and cell bottom. The radiation-accelerated particles left the space of interaction and were photographed with a MINTRON high-sensitivity controlled-shutter video camera interfaced to a computer and a synchronizer. The optical axis of the objective of the video camera was directed normally to the direction of laser beam propagation and passed through the space of particle–radiation interaction. The synchronizer generated a lasing-starting sync pulse and a frame sync pulse. Interval Δt_1 between the time of the frame pulse (t_0) and the time of the laser pulse (t_1) may be varied. The exposure time, i.e., the time over which the shutter is open, may be varied from 100 μs to 2 ms (in the experiments, it equaled 500 μs). We could also vary interval Δt_s between the time of the frame pulse (t_0) and the time of shutter opening (t_s). Thus, over a time interval of 500 μs, we could record events taking place within time Δt_s. Cutting Δt_s made the image of particles flying apart disappear from the monitor, while a small increase in t_s returned the image to the monitor. This time instant (t_1) was taken to be the beginning of particle motion. Estimating the difference between t_1 and t_1, we determined the delay between the beginning of the laser pulse and the beginning of particle motion. Then, increasing t_s, we could record the accumulation of particles flying apart within time Δt that is the difference between t_s and t_1. The computer controlled the synchronizer, as well as photo

acquisition, storage, and processing. The images were used to measure the length of the most intense tracks of pulse-field-heated particles flying almost normally to the bed. In this way, we minimized the error in determining the track length due to the angular divergence of the particles. Dividing the track length by Δt, one can easily find the mean velocity of the particles over this time period.

In the experiments, we determined the threshold value of the laser energy at which particles of different diameters start moving. It turned out that the energy density 0.3 J/cm^2 suffices to accelerate particles of diameter 12 μm from the suspension. To accelerate these particles from the powder, 0.7 J/cm^2 is required. To accelerate particles of diameter 27 μm from the powder, it is necessary that the energy density be 1.5 J/cm^2; for those of diameter 40 μm, 2.2 J/cm^2; and for those of diameter 160 μm, 6.0 J/cm^2. That is, the energy density required for sapphire particles to be accelerated from water (suspension) is roughly half as high as that needed for the particles containing water adsorbed by the surface (powder) to be accelerated. It was established experimentally that the particle velocity varies within the time of laser action (Fig. 1), increasing with laser energy density (Fig. 2). Over the time of laser action, the velocity of 12-μm Al$_2$O$_3$ particles first sharply grows, reaches a maximum (by the 40th microsecond), and then declines gradually. Moreover, the particles go on moving after the pulse is switched off. It should be noted that such behavior was also observed for particles of the other diameters. In the field of a submillimeter pulse from an YAG : Er laser, a nonabsorbing Al$_2$O$_3$ particle surrounded by surface-adsorbed

water (powder) may be accelerated to a velocity of 275 m/s, which corresponds to a kinetic energy of about 0.27 μJ for the 12-μm particles. This energy suffices to destroy a very hard material, such as tooth enamel [5].

REFERENCES

1. E. F. Nuchols and G. F. Hull, Phys. Rev. **17**, 26 (1903).

2. A. G. Askar'yan, Pis'ma Zh. Éksp. Teor. Fiz. **42**, 1567 (1962).

3. A. G. Askar'yan and E. M. Moroz, Pis'ma Zh. Éksp. Teor. Fiz. **43**, 2319 (1962).

4. A. G. Askar'yan, M. S. Rabinovich, M. M. Savchenko, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **5**, 258 (1967) [JETP Lett. **5**, 208 (1967)].

5. G. B. Altshuler, A. V. Belikov, and Y. A. Sinelnik, Lasers Surg. Med. **28**, 435 (2001).

Translated by V. Isaakyan

SHORT COMMUNICATIONS

Morphology, Texture, and Properties of Products Obtained by Annealing of Porous Silver–Polyacrylate Nanocomposites

T. I. Izaak, O. V. Babkina, and G. M. Mokrousov

Tomsk State University, Tomsk, 634050 Russia

e-mail: taina@mail.tomsknet.ru

Received September 27, 2004

Abstract—The methods of thermal analysis and scanning electron microscopy are used to study phenomena taking place at thermal decomposition of porous polyacrylate composites filled with silver nanoparticles that are synthesized *in situ* by UV reduction. Self-organization processes showing up in the redistribution of the silver nanoparticles in the polymer melt followed by their sintering into agglomerates are revealed. The morphology of the agglomerates is found to depend on the structure of the porous polymeric matrix. Using X-ray diffraction, it is shown that annealing produces silver crystallites 70–90 nm across, which are partially textured in "sheetlike" {110} planes and are stable against dissolution in nitric acid. © 2005 Pleiades Publishing, Inc.

Processes taking place when nanoheterogeneous materials consisting of metal nanoparticles immobilized in a porous matrix are subjected to high temperatures are of interest both for predicting their behavior in real catalytic and other high-temperature operations and for synthesis of novel materials with specially tailored properties. Earlier, we studied the thermal decomposition of porous polyacrylates that are filled with silver nanoparticles synthesized *in situ* and revealed characteristic features of this process that are attributed to inhomogeneities inherent in the polymer matrix and to the presence of the nanoparticles, which may be redistributed (self-organize) in the melt under the action of surface tension forces [1]. However, the composition and structure of annealing products are not yet understood. In this work, we show that the self-organization of nanoparticles during the decomposition of the porous nanocomposite influences the structure of the final annealing product and describe its composition and properties.

Polymeric matrices were produced by the copolymerization of methyl methacrylate and potassium methacrylate in the presence of a porogene (polyethylene glycol) and a polymerization initiator (benzoyl peroxide) at 65°C. The block samples obtained were 2 mm thick.

The polymer porous structure formed by extraction of the ethylene glycol by deionized water represents spherical macropores connected via reagent-permeable intercommunicating open micropores. The spherical macropores contain the carboxylate groups of potassium methacrylate, owing to which the polymer is capable of forming complexes with metal ions. Small aspherical inhomogeneities serve as through micro-channels connecting the macropores and allowing the ions to diffuse into the volume. The size and amount of the macropores depend on the molar ratio between the diluent and methyl methacrylate in the mixture to be polymerized and, thus, can be easily controlled [2]: an increase in the polyethylene glycol concentration in the system increases the size of the macropores in the polymeric matrix and diminishes their number. During drying of the polymeric matrix, some of the micropores collapse and the size of the macropores decreases.

The sorption of the silver ions into the volume of the porous matrix was accomplished from a 0.3 M solution of silver nitrate. Silver nanoparticles were synthesized *in situ* by the method of photoreduction ($\lambda = 350$ nm). According to X-ray diffraction data, the mean size of the nanoparticles was 7.3 nm and the silver concentration amounted to 12–15 wt%. The nanocomposites thus prepared were dried in air and subjected to heat treatment.

The thermal decomposition of the porous nanocomposites was studied in air with a Paulic–Paulic–Erdey 1500*Q* derivatograph. The heating rate was 10°C/min in the interval 25–700°C. The weighed sample was varied from 40 to 45 mg. The morphology of the decomposition products was examined in a Philips SEM 515 scanning electron microscope.

X-ray diffraction data were taken with a Bruker D8 GADDS diffractometer (CuK_α radiation, graphite monochromator, angular range $2\Theta = 5°–75°$). Crystallite size D was calculated by the Debye–Scherrer formula [3], and the intensity of reflections from the annealed silver was calculated using the POWDER-SELL 2.0 program.
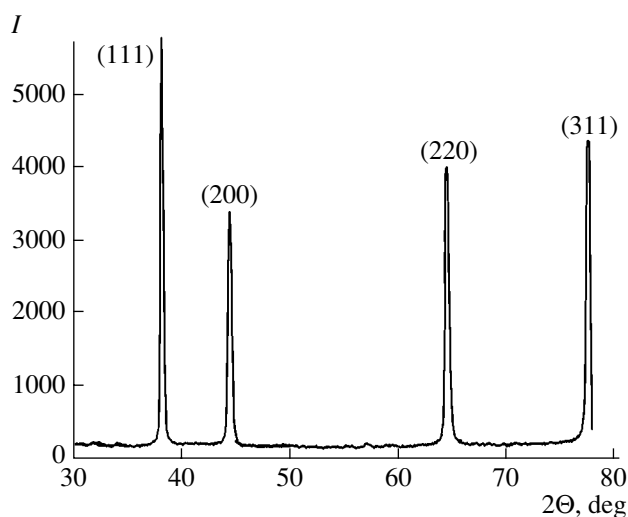
Fig. 1. Diffraction pattern taken from the products of annealing of the porous silver–polyacrylate nanocomposite.
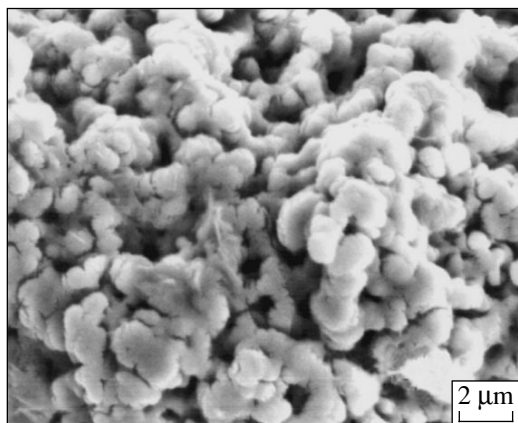

Fig. 2. SEM micrograph taken of the silver agglomerates (the final products of annealing of the porous silver–poly-acrylate nanocomposite).

As the temperature of the porous polyacrylate nanocomposites rises to 130–170°C, the polymer matrix starts melting. The macropores transform into cavities, which are gradually filled with gaseous products of polymer decomposition. Initially, the decomposition of the polymer is similar to the decomposition of pure PMMA, which starts at 160°C [4]. Foaming of the melt accompanied by a several-fold increase in the sample volume is observed visually. During foaming, the silver nanoparticles are redistributed, concentrating at the melt–gas interface on the boundaries of resulting bubbles. On further thermal oxidation, the polymer consolidates and the bubbles collapse. At temperatures of 415–420°C, the metal nanoparticles are sintered, producing spheroidal agglomerates. Carboniferous spacings between the agglomerates prevent the silver nanoparticles from being sintered into a single continuous block. At temperatures of 440–630°C, these spacings

burn up, releasing a great amount of heat; however, the silver agglomerates already formed are still not sintered together, retaining a high specific surface area. SEM micrographs taken from the final products of thermal decomposition show mesostructures of size between 0.3 and 1.5 μm depending on the porosity of the as-prepared polymer matrix. If the number of pores is small and their size is large, the material decomposes into larger agglomerates. Thus, one can obtain annealing products with a desired morphology by varying the composition of the mixture to be polymerized. The influence of the polymer matrix structure on the silver agglomerate size described elsewhere [1] is due to the fact that the early stage of decomposition, which is accompanied by gas release, proceeds on the surface of structural irregularities present in the melt (these irregularities occupy the sites of former macropores). The temperature-induced processes of metal nanoparticle redistribution, sintering, and mesostructure formation, proceeding in the porous matrix, may be viewed as self-organization at the nano- and microlevels.

Figure 1 shows the diffraction pattern taken from the products of annealing of the silver–polyacrylate nanocomposite kept in air at 630°C for 30 min. The diffraction lines from the silver are distinctly seen. Reflections due to the carboniferous products of polymer decomposition are absent. It seems that the UV preirradiation of the polymer, which reduces its thermal stability, has a significant effect on the annealing product purity: according to X-ray diffraction data, silver agglomerates extracted from silver–polyacrylate composites by thermal reduction of silver salts in the volume of the same polyacrylate matrix [5] contain a much higher amount of carboniferous products.

The observed relative intensities of the (111), (200), and (220) reflections (Fig. 1) for the annealed silver differ drastically from those calculated for isotropic powder samples. The experimental intensity ratio may be explained by assuming the presence of sheetlike texture along {110} planes. The best coincidence between the experimental and calculated intensities is observed at a texture index of 0.35. The texture may be attributed to the pre-orientation of the silver particles under the action of surface tension forces at the gas–melt interface in the polymeric "foam." Another reason may be oriented heat removal during the recrystallization of the silver.

Ignoring internal elastic stresses and making allowance for a correction for line broadening, we find that the size of the coherent scattering region for the annealed silver calculated by the Debye–Scherrer formula equals 70–90 nm. This means that the agglomerates with an apparent grain size of 1–2 μm (Fig. 2) are actually composed of grains less than 100 nm across.

The silver agglomerates that are products of annealing of the silver–polyacrylate composite dissolve in concentrated nitric acid only after long-term heating, while silver produced by other techniques is readily

soluble in HNO_3. There are a number of publications noting that silver nanoparticles obtained by photolysis offer high stability against nitric acid. These particles ≈4 nm in size have the form of icosahedra [6] and a regular crystal structure. It seems likely that silver crystallites made up of nanoparticles obtained by photoreduction retain their regular structure during the thermal decomposition of the silver–polyacrylate composite and subsequent sintering. As a result, the products of nanocomposite annealing are hard to dissolve in nitric acid.

ACKNOWLEDGMENTS

REFERENCES

1. T. I. Izaak, O. V. Babkina, A. N. Salanov, *et al.*, Vysokomol. Soedin. **45**, 939 (2003).

2. T. I. Izaak, O. V. Babkina, T. N. Drebushchak, *et al.*, Zh. Prikl. Khim. (St. Petersburg) **76**, 1853 (2003).

3. L. M. Kovba and V. K. Trunov, *X-ray Phase Analysis* (MGU, Moscow, 1976) [in Russian].

4. B. M. Ginzburg, V. L. Ugolkov, L. A. Shibaev, *et al.*, Pis'ma Zh. Tekh. Fiz. **27** (19), 11 (2001) [Tech. Phys. Lett. **27**, 806 (2001)].

5. T. I. Izaak, O. V. Babkina, A. I. Boronin, *et al.*, Koloidn. Zh. **65**, 788 (2003).

6. S. C. Davis and K. J. Klabunde, Chem. Rev. **82**, 153 (1982).

Translated by V. Isaakyan