Semiconductors, Vol. 35, No. 9, 2001, pp. 1006–1017. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1050–1062.

SEMICONDUCTOR STRUCTURES, INTERFACES, AND SURFACES

Oxidation of Semiconductors and the Constitution of Interfaces

S. M. Repinskiĭ

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

e-mail: repinski@thermo.isp.nsc.ru

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—The effects of the self-organization of phase boundaries in semiconductors (separation surfaces and reaction zones) are considered for semiconductor oxidation, the most technologically important process in microelectronics. The analysis of kinetic data yields entropies of activation of these processes and demonstrates that the silicon–oxygen system shows the strongest self-organization of the reaction zone—the silicon–silicon dioxide interface. *© 2001 MAIK "Nauka/Interperiodica".*

In the 1960s, it was customary to infer the fundamental aspects of electronic processes occurring on the semiconductor surface from results obtained while studying either atomically clean or real surfaces. The specific features of the real surface were attributed to the presence of a more or less thick film of oxide compounds, formed during the oxidative-solvation etching of crystals [1]. Although surfaces of this kind possessed a much lower density of surface states than the atomically clean surfaces, only the prolonged efforts of researchers succeeded in revealing two types of virtually stable and surface state–free interfaces: silicon–⟨silicon dioxide⟩ and germanium (silicon)–electrolyte. The silicon–⟨silicon dioxide⟩ interface proved to be an important discovery of the century and thus predetermined the development of microelectronics and the successful fabrication of large-scale integrated circuits [2].

The aim of this paper is to demonstrate that this success was largely predetermined by the phenomenon of the self-organization of phase boundaries in the semiconductor in the course of oxidation, the process underlying the main technological procedures used in the fabrication of integrated circuits.

MAIN CONCEPTS: DISCONTINUITY SURFACE AND REACTION ZONE

In studying the self-organization phenomena, we rely upon the concepts developed in Prigogine's works [3] and, later, in [4, 5]. The basic idea consists in that, provided there is energy and substance exchange with the environment, the evolution of an open system may proceed in such a way that states appear which have lower entropy than the initial value of this parameter. Such a state, exceedingly improbable from the standpoint of equilibrium thermodynamics, may exist indeterminately under stationary conditions. Any system studied possesses a spatiotemporal structure determined by processes occurring in this system, and, commonly, threshold states can be distinguished, with transitions via these states leading to fundamental changes in the organization of this structure. The evolution of a system, regarded as a sequence of transitions from one quasi-stable state to another, with these states being characterized by certain organization parameters, is called self-organization. The measure of organization of a system is commonly defined in terms of changes in entropy.

The concept of the phase boundary needs refinement. Only two notions should be considered meaningful: the phase discontinuity surface and the reaction zone. According to Gibbs, the phase discontinuity surface is a heterogeneous region of finite thickness between adjoining phases, whose properties are parameters of state (temperature, pressure, chemical potential, etc.) [6, 7]. The properties of a phase discontinuity surface can be studied experimentally if a thermodynamic equilibrium is established in a gas–solid or vacuum–solid system (e.g., adsorption processes occur), with equilibrium attained between the bulk of the solid and the discontinuity surface. The concept of the reaction zone has emerged in investigations into the catalysis of topochemical reactions. The reaction zone is a solid layer of finite thickness, with the chemical composition and structure of this layer being determined by the action of the reaction mixture and characterizing the organization of the system for a certain stationary state.

The self-organization of the reaction zone was discovered comparatively long ago as a result of studying heterogeneous catalysis. In 1980, when analyzing the experimental data on the surface structure of the metal and oxide catalysts available at that time, G.K. Boreskov came to the conclusion that a solid catalyst surface interacting with reactants cannot be characterized by preset invariable properties. The catalyst

1063-7826/01/3509-1006$21.00

surface acts as a labile component of the reaction system, reflecting changes in composition and in other process parameters. Reconstruction during the chemisorption of oxygen onto the (110) and (100) nickel faces and auto-oscillations of the rate of heterogeneous oxidation can serve as indicative examples of this situation [8].

The effect of the self-organization of phase boundaries in semiconductor crystals was considered in my previous paper [2]. Several types of phase boundaries can be distinguished. The simplest system is a semiconductor crystal in contact with a vacuum; data on superstructures and two-dimensional (2D) phase transitions at relatively high temperatures are available for this system. Evaporation and crystal growth are observed in the system ⟨semiconductor crystal⟩–⟨gas phase containing atoms or molecules of the starting substance⟩. In the crystal–⟨active gas⟩ system, adsorption and semiconductor dissolution (etching) and oxidation occur. Finally, crystal dissolution occurs in the system semiconductor–⟨condensed phase⟩, which may be, in particular, an electrolyte.

SEMICONDUCTOR–VACUUM AND SEMICONDUCTOR–VAPOR PHASE DISCONTINUITY SURFACES

The simplest case of surface self-organization in silicon and germanium crystals involves surface reconstruction and the formation of 2D phases [9–11]. These states are, as a rule, associated with the establishment of a crystal–surface thermodynamic equilibrium and can be described in terms of the Gibbs approach as crystal–vacuum phase discontinuity surfaces. Let us briefly characterize this system. Fewer nearest neighbors to the surface atoms leads to changes in their chemical bonding. This manifests itself in the first place in a lower share of *s*-orbitals of the surface atoms in chemical bonding [12]. A study of the surface of covalent crystals by low-energy electron diffraction demonstrated that the structure of surface layers is, as a rule, different from the corresponding planes in the bulk of the crystal. For most of surface structures, two kinds of 2D phase transitions are observed with changing temperature: order–disorder and order–order. An order–disorder transition can serve as an example of the disappearance of the 7×7 Si(111) structure at temperatures exceeding 880°C. Reversible order–order transitions are observed on the (110) faces and vicinal faces of silicon, Si(211), on which a regular system of one-atom steps is transformed into a system of steps two atoms high [13, 14]. The actual structure of the surface unit cells is known exactly only in a few cases. The 2×1 Si(100) structure is relatively simply described as an asymmetric dimer. A specific feature of this structure is that the surface atoms of neighboring rows are shifted toward one another to form a chemical bond. This shift leads to the relaxation of their nearest neighbors.

The self-organization phenomena in the crystal–vapor system can be conveniently considered for the example of the evaporation of silicon and germanium crystals [15–17]. In the vicinity of the melting point, the gas phase composition is mainly determined by monatomic particles. The content of polyatomic clusters is no higher than fractions of a percent. The temperature dependence of the evaporation rate shows that its activation energy is equal to the heat of sublimation: 346.9 ± 16.7 kJ/mol for germanium and 438.9 kJ/mol for silicon, with a pre-exponential factor on the order of 10^{28} at. cm^{-2} s^{-1}. The rates of evaporation from the (111), (110), (100), and (211) faces are the same within experimental error. Planar etching pits with a steplike structure are formed on the (111) face, and the velocity at which these steps shift in the course of evaporation allowed an evaluation suggesting that only less than 1% of the total loss of silicon is associated with the changing dimensions of the etching pits [16].

Thus, evaporation proceeds uniformly over the entire crystal surface. This means that the conventional analysis of the process in terms of the terrace–kink–step mechanism is inexact. The factors noted suggest that the constitution of the reaction zone in evaporation is the same for crystals with different surface orientations.

The evaporation of silicon and germanium proceeds by the mechanism of zero-barrier reactions for which the activation energy is equal to the enthalpy of sublimation of the crystal [2]. The rate-determining stage of the process is that of atom detachment, corresponding to a configuration "position at half-crystal," i.e., to a structure with a doubly coordinated atom. In quantitative respect, the organization of the semiconductor–vacuum and semiconductor–⟨vapor consisting of starting crystal atoms⟩ discontinuity surfaces is much the same, being characterized by a minor increase in entropy.

REACTION ZONE IN SEMICONDUCTOR–GAS AND SEMICONDUCTOR–LIQUID SYSTEMS

The dissolution and oxidation of semiconductors are based on reactions with halogens, oxygen, and other oxidants. A vast body of experimental data has found reflection in a number of reviews [18–21]. Let us discuss, relying upon these papers, the results obtained in terms of the problem in question.

Table 1 summarizes the kinetic characteristics of the main types of chemical reactions occurring at the surface of silicon and germanium crystals: the adsorption and dissolution of crystals in liquids and gas (chemical evaporation).

The rate constant k is found using the formula

$$k = \nu \exp(\Delta S/R) \exp(-\Delta H/RT). \qquad (1)$$

Here, T is temperature; R, the gas constant; ΔS, the entropy and ΔH, the enthalpy of activation of the reaction; and ν, the frequency factor.

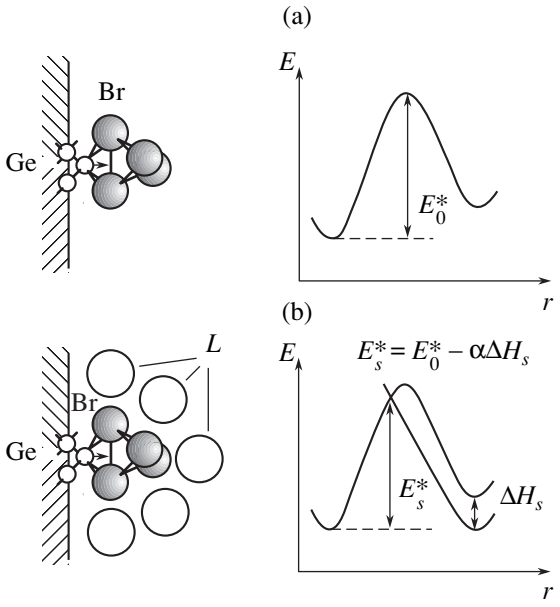


Fig. 1. Constitution of the reaction zone in the reactions in the germanium–bromine system. (a) Reaction in the gas phase and (b) reaction in nonpolar solvent. *L* is the solvent molecule. The right-hand part shows how the energy of the system changes in moving along the reaction coordinate. E_0^* is the activation energy of the reaction in the gas phase, E_s^* is the same for the nonpolar solvent, and ΔH_s is the heat of solvation of the reaction product.

It can be seen from Table 1 that all the reactions between the active gas phase and semiconductor crystal, yielding gaseous products, are characterized by a minor increase in entropy, whereas adsorption and crystal dissolution in a liquid are accompanied by a decrease in the entropy of activation by 20–80 J/(K mol). In the last two cases, the effect of self-organization of the reaction zone is clearly pronounced. The constitution of the reaction zone is, as a rule, not identical to that of a monomolecular adsorption layer. The formation of a chemisorption layer is commonly accompanied by the appearance of a reconstructed layer of surface atoms. For some reaction types, the formation of super-monolayer coatings was observed, which can be conventionally termed a layer of physically adsorbed molecules [21, 22].

The silicon (germanium)–halogen systems are characterized by the formation of di- and tetrahalide compounds. The nature of the processes occurring in the systems under consideration is illustrated in Fig. 1. From the adsorption of halogens onto an atomically clean surface, a preadsorption state is formed, with translational degrees of freedom lost and rotational degrees of freedom partly retained. Tetrahalides are formed via supermonolayer adsorption. Such a structure is stable and is preserved at the surface in passing from the gas phase into a nonspecific solvent, e.g., in the system Ge–(Br_2 + $GeBr_4$) [23, 24].

Let us consider specific features of reactions for the example of the reaction of germanium with bromine and sulfur:

$$Ge + 2Br_2 \longrightarrow GeBr_4,$$

$$Ge + (1/2)S_2 \longrightarrow GeS.$$

Figures 2 and 3 show how the reaction rate *w* depends on the partial pressures of the reagents. It can be seen qualitatively that the dependences have the form of curves tending to limiting values, different for different temperatures of the experiment. The data for the reaction with bromine are described by an equation of the type [22]

$$w = k_2 b p_{Br_2}(1 + b p_{Br_2})^{-1}, \quad (2)$$

in which

$$k_2[\text{at. cm}^{-2}\ \text{s}^{-1}] = 10^{28}\exp(-137.9/RT),$$

$$b[\text{Torr}^{-1}] = 1.75 \times 10^{-3}\exp(37.6/RT).$$

The data for the sulfidizing reaction are described by an equation of the type [25, 26]

$$w = k_2 \beta p_{S_2}(1 + \beta p_{S_2})^{-1}, \quad (3)$$

in which

$$k_2[\text{at. cm}^{-2}\ \text{s}^{-1}] = 1.2 \times 10^{29}\exp(-182.6/RT),$$

$$\beta[\text{Torr}^{-1}] = 4.9 \times 10^{-10}\exp(146.3/RT).$$

An analysis of the dependences presented can be made using a general model of reactions of the given type. Let us denote a gas species by *A*, an adsorption center on the solid surface by *S*, a reaction product by *B*, and adsorbed species A and B by *AS* and *BS*. Then the reaction scheme is described by a three-stage process:

Table 1. Kinetic characteristics of chemical reactions occurring at the surface of silicon and germanium crystals

Type of reaction	Temperature range, °C	ΔH, kJ/mol	ΔS, J/(K mol)
Si, Ge evaporation	600–1000	250–340	0–80
Si + O_2, Ge + O_2 (thermal oxidation)			
Reactions of Ge and Si with Cl_2, Br_2, I_2, tetrahalides	200–500	125–170	0–40
Dissolution of Si, Ge in liquids	25–100	60–80	–20–(–80)
Adsorption of O_2, S_2, Br_2	200–600	0 ± 6.0	–80

(1) reagent adsorption

$$A + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} AS;$$

(2) transformations in the adsorption layer

$$AS \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} BS;$$

(3) desorption of the product and regeneration of the surface center

$$BS \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} B + S.$$

Commonly, the rate of a process can be measured under steady-state conditions, when the concentrations of intermediates *AS* and *BS* are time-independent.

If the process occurs in such a way that, in the sequence of stages, the rate of conversion in the adsorption layer is slow, i.e., $k_{-2} < k_2$, and each of the constants k_1, k_{-1}, $k_3 > k_2$, then we have

$$w = k_2 b p_A (1 + b p_A)^{-1}, \quad (4)$$

$$b = k_1/k_{-1} = b_0 \exp(q/RT). \quad (5)$$

Here, b is the adsorption equilibrium constant. At high pressures $bp_A > 1$, $w = k_2$, and the reaction rate becomes independent of pressure p_A, whereas at low pressures $bp_A < 1$, $w = k_2 b p_A$, and the reaction rate linearly depends on pressure, with the activation energy of the process in this range lower than the activation energy in the region of zero reaction order by the heat of reagent adsorption, q.

If, by contrast, the reaction in the adsorption layer is so fast that it can be considered instantaneous, then the overall scheme of the process is described by two stages

$$A + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} AS,$$

$$AS \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} B + S$$

and the process rate is given by

$$w = k_2 \Theta_{AS} = k_2 (k_1 p_A + k_{-2} p_B)/(k_1 p_A + k_{-1} + k_2). \quad (6)$$

This expression is simplified under the conditions of a flow-through reactor, $p_B \ll p_A$, and, in the case of irreversible chemisorption of A, $k_{-1} = 0$:

$$w = k_2 \beta p_A (1 + \beta p_A)^{-1}, \quad (7)$$

where

$$\beta = k_1/k_2 = (k_{01}/k_{02}) \exp[-(E_1 - E_2)/RT], \quad (8)$$

and E_1 and E_2 are the activation energies of the adsorption and desorption stages, respectively. At low pressures of A, $\beta p_A < 1$, the process rate is expressed as $w = k_1 p_A$, which means that adsorption is the rate-determin-

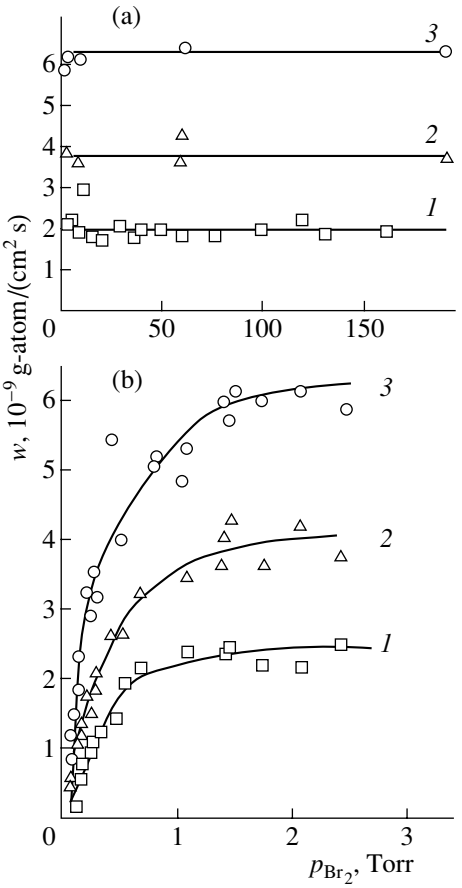


Fig. 2. Rate of reaction between germanium and bromine vs. the bromine pressure. Reaction temperature: (*1*) 270, (*2*) 280, and (*3*) 290°C. The pressure ranges in (a) and (b) are different.

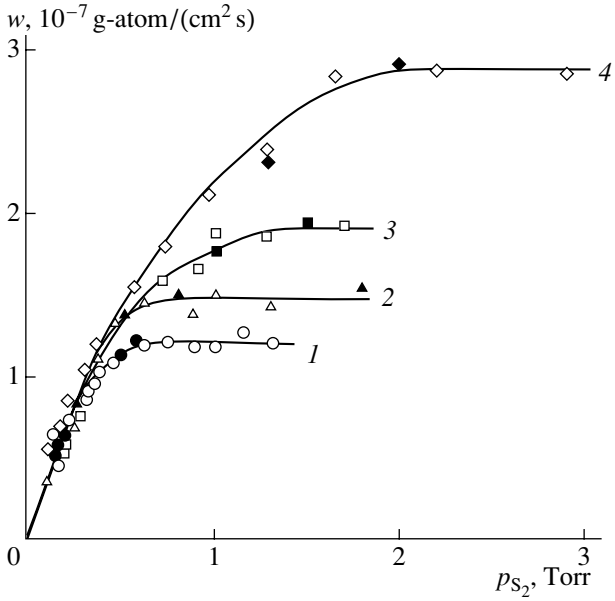


Fig. 3. Rate of reaction between germanium and sulfur vapor vs. sulfur pressure at different crystal orientations and temperatures T: (*1*) 500, (*2*) 510, (*3*) 520, and (*4*) 530°C.

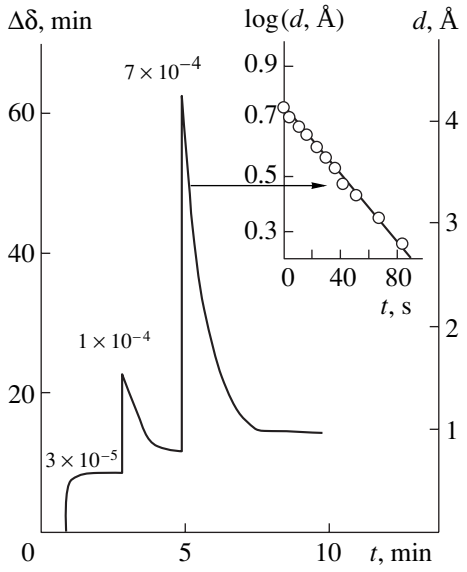


Fig. 4. Variation of the adsorbed bromine layer thickness at different germanium sample exposures (pressures in Torr are shown in the figure). Insert: calculated rate of germanium tetrabromide formation. Temperature 190°C.

ing stage of the process. At $\beta p_A > 1$, the process rate $w = k_2$ is pressure-independent and the process is controlled by the desorption stage.

Formally having the same form, Eqs. (4) and (7) describe physically different situations. Equation (4) characterizes the rate of a process whose rate is determined by the stage of chemical transformation in the adsorption layer. In this case, the preceding and succeeding stages of, respectively, reagent adsorption and desorption of the reaction products are regarded as equilibrium stages. Equation (7) characterizes the process course with comparable rates of reagent adsorption and product desorption. For widely varying reagent pressures, the nature of the rate-determining stage changes in such a way that the adsorption stage is slowest at low pressures, and the product desorption stage, at high pressures.

The heat of bromine adsorption onto a brominated germanium surface is, apparently, much lower than the heat of adsorption onto an atomically clean surface of the same material. In this case, the adsorption becomes reversible, and the slow stage is the reaction in the adsorption layer

$$GeBe_2 \cdot Br_{2\ ads} \longrightarrow GeBe_{4\ ads}.$$

Since there is no reason to assume that the heats of $GeBr_4$ adsorption are high, we consider the stage of germanium tetrabromide desorption to be fast. Thus, the process rate will be described in this case by an equation of type (4), corresponding to empirical equation (2), with the correspondence being more than merely qualitative. It is noteworthy that $b_0 = 1.75 \times 10^{-3}$ Torr^{-1} agrees with the characteristic value of the adsorption equilibrium constant in the case of the formation of a labile adsorption layer.

The problem of the reaction mechanism is central to chemical physics in general and to the chemical physics of surfaces in particular; since, of course, no scheme based solely on logical premises can be considered proven. The direct experimental detection of intermediates predicted by the given scheme would be of decisive importance. Although it is not always possible to perform such an experiment, it could be done in this case. Automated ellipsometry was applied to determine the coverage of the germanium surface by bromine molecules directly in the course of the chemical reaction. Figure 4 presents the results obtained in this experiment. It can be seen that at small exposures of an atomically clean germanium surface irreversible chemisorption of a monolayer of bromine atoms to bromine vapor occurs (several minutes at a pressure of 3×10^{-5} Torr), with layer thickness of 1.14 Å. When the pressure increases to 10^{-4} Torr and more, the coating thickness grows rapidly to 6 Å and then decays exponentially with time to values of about 1.0–1.5 Å. The rate constant of the exponential decay, found from these dependences, is equal, with high precision, to the rate constant of germanium etching at low bromine pressures.

Thus, the experimental results confirm the assumption that germanium oxidation giving $GeBr_4$ proceeds only in the case of a supermonolayer coverage of the surface by bromine molecules.

As frequently happens, the practically interesting system is too complicated to easily make a detailed analysis of the physicochemical relationships, and it is necessary to find systems with model properties. Of interest in this regard are the reactions of silicon and germanium with halogens. In particular, there is good experimental material for $GeBr_4$ formation in the reaction of germanium with bromine. It is also important that kinetic relationships for the same reaction occurring at the semiconductor–gas and semiconductor–solution interfaces can be compared in this case [23, 24].

Germanium tetrabromide boils at 186°C under normal pressure. At lower temperatures, this compound is a low-polarity liquid, in which bromine vapor is highly soluble, and it is possible to monitor the reaction of germanium with a solution of bromine in germanium tetrabromide. It seems natural to expect that the activation energy of dissolution will be lower according to the heat of product dissolution, i.e., according to the heat of $GeBr_4$ evaporation, in the given case.

The rate of dissolution of germanium crystals at the (111) and (110) faces depends on the concentration of dissolved bromine, c:

$$w = w_0 bc(1 + bc)^{-1}, \qquad (9)$$

where w_0 is the reaction rate in the zero-order region and $b = b_0 \exp(q/RT)$. The (100) faces of germanium crystals dissolve in such a way that a linear law $w = kc$

is obeyed up to a concentration of 1.3 M, and at higher concentrations the rate becomes concentration-independent. The temperature dependence of the dissolution rate in the zero-order range is not affected by the crystal surface orientation and has an activation energy of 104.5 kJ/mol. For the reaction in the gas phase

$$Ge_{sol} + 2Br_{2\ gas} = GeBr_{4\ gas}$$

the activation energy is 137.9 kJ/mol. The difference between these two activation energies is equal, with good precision, to the heat of $GeBr_4$ evaporation (37.6 kJ/mol).

Thus, it is clear that the exponential factor in the expression for the process rate in the zero-order range has the form

$$\exp[(E_{gas} - \alpha q_{GeBr_4})/RT],$$

where E_{gas} is the activation energy of the reaction in the gas phase, and q_{GeBr_4} is the heat of $GeBr_4$ evaporation. The given expression formally conforms to a linear relation between the activation energy and the enthalpy of the process, with $\alpha = 1$ in the given case, which means that the activated complex and reaction product species differ only slightly. Account should be taken of the fact that the state of bromine molecules in a germanium tetrabromide solution corresponds to a decrease in the adsorption energy by a value equal to the heat of dissolution, q_{ds}. A bromine molecule can be localized at the germanium surface with effective heat of adsorption q_{ads}. If reckoned from the standard state, the heat of adsorption is given by

$$q^0_{ads} = q_{ads} + q_{ds}.$$

This means that the effective heat of adsorption q_{ads} is determined by two main factors: the structure of the solution and the structure of the crystal surface in contact with the solution. If the enthalpy of mixing grows substantially on changing the solution concentration, this lowers the heat of adsorption of the active component of the solution.

The entire process of germanium dissolution can be represented as a multistage reaction. The first stage corresponds to the fast irreversible chemisorption of bromine on a germanium atom on the crystal surface, appearing upon completion of the preceding dissolution events:

$$Ge_{cr} + Br_2 \longrightarrow Ge(Br_2)_{chem}.$$

Reconstruction of the surface layer of germanium occurs at this stage, and the extent of its completion determines the energetics of the subsequent reversible interaction of bromine molecules:

$$Br_{2\ liq} \rightleftharpoons Br_{2\ ads}.$$

The orientation dependence of the crystal dissolution rate arises as a consequence of the specific features of the adsorption interaction in this stage. In particular, reconstruction presumably proceeds easily during the chemisorption of bromine on the Ge(100) surface and, in the case of monolayer coverage, subsequent interaction with bromine molecules is only possible via forces arising from the intermolecular interaction.

For the Ge(111) and (110) surfaces, the rearrangement of the surface region is not complete in the stage of premonolayer formation, and effective interaction between bromine molecules and atoms of the subsurface layer becomes possible. Qualitatively, this means that

$$b^{(100)} > b^{(111)} > b^{(110)} \quad \text{and} \quad q^{(100)} < q^{(110)} < q^{(111)}.$$

The b_0 value can be estimated as a ratio of the complete statistical sum for bromine molecules in the near-surface layer and in the solution bulk. Thus, it would appear reasonable that the molecules in the near-surface layer lose one translational degree of freedom (normal to the crystal surface):

$$b_0 = h/(2\pi MRT)^{1/2}d, \qquad (10)$$

where h is the Planck constant, M the molecular weight, and d the length of the diffusion displacement of bromine molecules. In the case in question, $M = 160$, and $d \approx 5 \times 10^{-8}$ cm. Estimates based on this expression give a value corresponding to within 50% to the pre-exponential factor of the kinetic equation for the rate of dissolution of the (100) face of the germanium crystal. In other words, in the case of Ge(100) the dissolution proceeds by the mechanism of direct collisions of Br_2 molecules with the chemisorption complex $Ge(Br_2)_{chem}$. The rate-determining stage of the process is the stage of transformation in the adsorption layer

$$Br_{2\ ads} + Ge(Br_2)_{chem} \longrightarrow GeBr_{4\ ads}.$$

In this case, the final product, identical to the solvent molecule, is readily desorbed

$$GeBr_{4\ ads} \rightleftharpoons GeBr_{4\ liq}.$$

The system $GeBr_2$–Br_2 can be classed with regular solutions for which no changes are observed in the structure of a solution on varying its concentration. This means, for example, that it is possible to pass gradually from solutions of bromine in germanium tetrabromide to solutions of germanium tetrabromide in bromine. With increasing bromine concentration in the solution, it becomes possible that bromine molecules not only occupy voids surrounded by germanium tetrabromide molecules, but come into direct contact with one another. This occurs beginning at a bromine concentration of 3M, when there are less than four $GeBr_2$ molecules per molecule of bromine. This fact accounts for the zero order of the reaction of Ge(100) dissolution in solutions with bromine concentrations exceeding 2M.

A new type of surface structure appears in reactions in polar solvents: Ge + (Br_2 + CCl_4 + CH_3OH). This is indicated by the magnitude of the negative entropy of

activation of the reaction. The effect of self-organization is sensitive to the solvent structure. For example, the involvement of CH_3OH monomer molecules can be explicitly observed in the $CH_3OH + CCl_4$ solvent [11].

Processes occurring at the semiconductor–electrolyte interface should be analyzed taking account of the potential profile in the Helmholtz layer and the semiconductor, i.e., the space charge region should be included into the constitution of the reaction zone [11, 27]. The electric double layer on metals commonly consists of a dense part of adsorbed cations with a solvate shell and a diffuse part. The dense part, named the Helmholtz layer, is 4–5 Å thick. As for the thickness of the diffuse part, it depends on the electrolyte concentration, varying between several micrometers and several angstroms. The formation of the electric double layer establishes a dynamic equilibrium between two concurrent processes—the transfer of metal ions into the solution in the form of solvated cations, facilitated by interaction with solvent molecules, and a reverse process in which cations are discharged and incorporated into the metal lattice. In the case of an established equilibrium, the rates of these two processes are equal and their extent is characterized by the exchange current. The exchange currents can be determined from electrochemical measurements and independent experiments employing isotope techniques. When the equilibrium is shifted because of the electrode polarization, the equality of the rates is disturbed and, depending on the type of polarization, one of the following two processes prevails: an anodic (oxidation) process, with cations passing into solution, and a cathodic process in which ions are reduced (discharged) and incorporated into the crystal lattice. The structure of the electric double layer at the semiconductor–electrolyte interface is more complex. In addition to the Helmholtz and diffuse layers, formed on the electrolyte side of the interface, there is a layer in the surface part of the semiconductor, charged with respect to the semiconductor bulk and commonly called the space charge region. The space charge region is frequently several micrometers thick. This means that the potential drop across the interface is distributed in such a way that its major part is associated with the space charge region, and only a minor part, with the Helmholtz layer. In contrast to the metal electrode, the carrier concentration at the surface of the semiconductor varies with its polarization, and conditions may arise leading to the formation of enriched, depletion, and inversion layers with respect to the crystal bulk. This system should be described taking into account the fundamental aspects of the processes occurring in two mutually dependent subsystems: atomic-ionic and electronic-ionic. The exchange currents in the atomic-ionic subsystem are commonly small for typical semiconductors. For example, in the case of germanium dissolution this value is 10^{-10} A/cm^2. The exchange currents in the electronic-ionic subsystem are commonly several orders of magnitude higher, ranging from 10^{-6} to 10^{-2} A/cm^2. Consequently, the equilibrium at the semiconductor–electrolyte interface is established readily and rapidly as regards the electronic-ionic system, whereas the atomic-ionic system exhibits slow behavior. This is manifested experimentally in that the rate of self-dissolution of the semiconductor, found by extrapolating the anodic and cathodic portions to the point of intersection at the stationary potential, is 1–2 orders of magnitude lower than the rate determined by weight measurements [11].

It is known that silicon is relatively inert in an aqueous medium because of the passivation of its surface by a dense layer of silicon dioxide. The reaction rate is fast only in an alkaline medium, in which the solubility of silicon hydroxides is high [11]. With increasing alkali concentration, the reaction rate grows to a maximum value and, on passing into the range of concentrated solutions, becomes slower. Another important feature consists in that the rates of dissolution of crystals with different surface orientations are not the same. It is clear that the dissolution reaction proceeds as a chemical interaction of surface silicon atoms with water molecules. Moreover, this interaction is a multistage process, with the adsorption of water molecules on the crystal surface being its initial stage. The stages of hydrogen evolution and egress of the surface atom from the crystal lattice are, presumably, slow. In the final stages hydration of the forming silicon hydroxide occurs.

The initial stage of interaction involving surface silicon atoms is characterized by the fast chemisorption of water molecules to form an adsorption complex with the following presumed structure

In the same stage, rearrangement of the surface region in silicon gives rise to a dependence of the dissolution rate on the crystal surface orientation. Further, upon adsorption of water molecules on the adsorption complexes $=Si(OH)_2$, the oxidation of silicon is complete with the formation of H_2SiO_3 which dissolves depending on the participation of hydroxyl ions

$$H_2SiO_3 + OH^- \longrightarrow HSiO_3^-,$$

$$H_2SiO_3 + 2OH^- \longrightarrow SiO_3^{2-}$$

and the subsequent hydration of the reaction products in the solution. It becomes clear that water molecules show two kinds of behavior during the process: during the adsorption interaction stage they act as oxidant species, and at the final stages of interaction, as solvating agents. Consequently, a kinetic equation for the process of silicon dissolution in alkalis should be sought in the form

$$w = k\theta_{H_2O}c_{sat}^{\alpha}. \quad (11)$$

Here, θ_{H_2O} is the coverage of the silicon surface by water molecules in the course of dissolution, and the factor c_{sat} characterizes the solubility of reaction products and reflects the effect of solvation by water molecules. Qualitatively, this means that the dissolution rate grows with increasing solubility, i.e., with the alkali concentration, while there is a sufficient number of free water molecules in the concentrated solutions. If, however, water molecules are only present in hydrate shells of the cation and hydroxyl, then the coverage of the silicon surface by water molecules, θ_{H_2O}, will fall dramatically, leading to a decrease in the silicon dissolution rate. Hence follows that the dissolution rate depends on the thermodynamic activity of water.

Figure 5 shows how the rate of silicon dissolution varies in ethylenediamine solutions [28]. Ethylenediamine $NH_2CH_2CH_2NH_2$ behaves in aqueous solutions as a base, being ionized by the scheme

$$NH_2CH_2CH_2NH_2 + H_2O$$

$$\rightleftharpoons NH_2CH_2CH_2NH_3^+ + OH^-,$$

$$NH_2CH_2CH_2NH_3^+ + H_2O$$

$$\rightleftharpoons NH_3^+CH_2CH_2NH_3^+ + OH^-.$$

This means that the dissolution of silicon in solutions of this kind must proceed similarly to dissolution in alkalis. However, Fig. 5 clearly demonstrates a specific feature consisting in that, as the ethylenediamine concentration varies, the dissolution rate passes through two maxima: the first at a concentration of 0.04–0.05 mol fraction, and the second at 0.5 mol fraction of ethylenediamine. It may be thought that the peak at the 0.5 mol fraction of ethylenediamine is similar in nature to the peak in the dependence of the silicon dissolution rate in alkalis. At the same time, the peak at low concentrations is apparently associated with the structure of the aqueous solution. It should be noted that liquid water is not an isotropic medium, being characterized by an icelike labile skeleton of water molecules bound together by hydrogen bonds. This skeleton has a rather open structure and water molecules may penetrate into its voids. Water molecules situated in these voids are comparatively mobile and their thermodynamic characteristics are close to the values for gaseous water. Having considered these data, we assume that at low concentrations of ethylenediamine its hydration does not cause any significant changes in the structure of water, and only consists in the displacement of water molecules from the voids of the icelike skeleton.

These two practically important cases, describing the kinetics of silicon dissolution in aqueous solution, allowed us to make the following two conclusions. First, we made certain that the solution cannot be regarded as an isotropic liquid medium by considering the real molecular structure of the solution. Second, it became clear that the effects observed in stationary dissolution are governed by adsorption processes occurring at the semiconductor–solution interface.

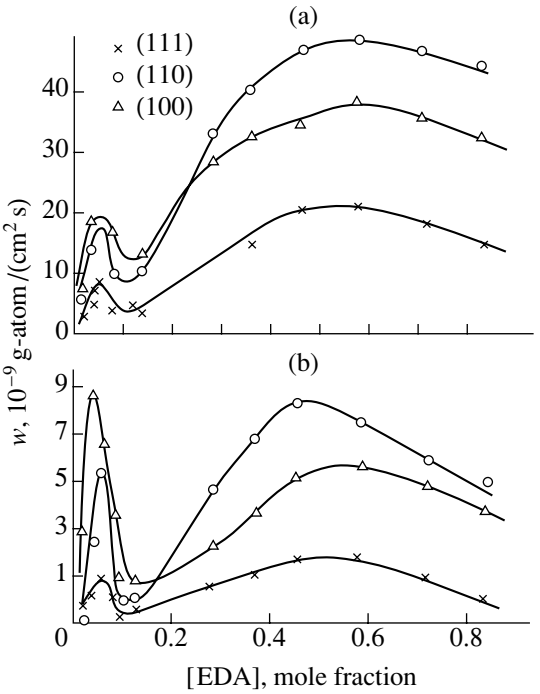


Fig. 5. Dissolution rate of (111), (110), and (100) silicon crystals vs. solution composition. The ethylenediamine concentration in solution is expressed in mole fractions. Dependences are presented for (a) 30 and (b) 70°C.

To conclude this section, it should be noted that the oxidation process should be described as a multistage chemical reaction. The rate-determining stage is a stage of chemical transformation in the adsorption layer usually formed from the oxidant and polar solvent species. The free energy of activation of this stage, $\Delta G^{\neq}$, decreases by a certain fraction α of the dissolution energy of the forming reaction product, ΔG_{ds}. Quantitatively, the dissolution rate w can be represented as

$$w = n_s \nu \theta_1 \theta_2 \exp[-(\Delta G^{\neq} - \alpha D G_{ds})/RT], \qquad (12)$$

where n_s is the density of surface atoms; ν is the thermal frequency of their oscillations; and θ_1 and θ_2 are the extents of surface coverage by adsorbed species of, respectively, the oxidant and polar solvent.

CRITICAL CONDITIONS FOR PROCESS TRANSITION FROM THE ETCHING TO OXIDATION MODE

It is known that the reaction of silicon and germanium with oxygen can produce gaseous monoxides and condensed dioxides of these elements:

$$Si + (1/2)O_2 = SiO, \quad Ge + (1/2)O_2 = GeO,$$

$$Si + O_2 = SiO_2, \quad Ge + O_2 = GeO_2.$$

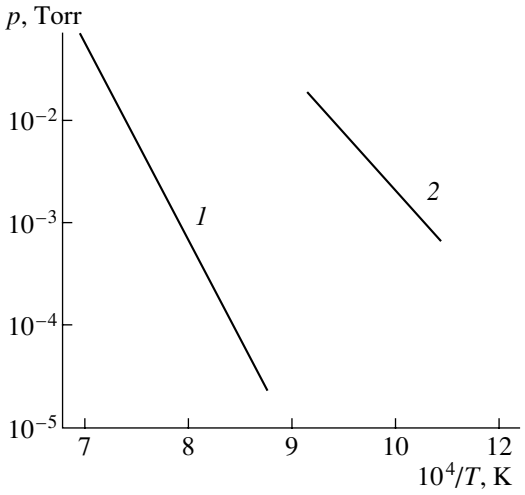
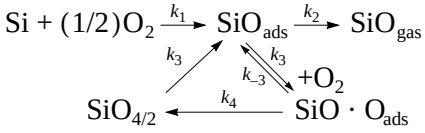


Fig. 6. Dependence of oxidant pressure on inverse temperature, determining the critical conditions for etching–oxidation of silicon by (*1*) oxygen and (*2*) nitrous oxide.

Similarly, the reaction of germanium with sulfur yields

$$Ge + (1/2)S_2 = GeS, \quad Ge + S_2 = GeS_2.$$

Whether a volatile SiO or protective SiO_2 layer is formed depends on temperature and partial pressure in such a way that the occurrence of etching is preferable at an elevated temperature and a low oxidant pressure. At high partial pressures of oxygen and low temperatures, the growth of a dielectric layer predominates. The critical conditions for the reaction to pass from one mode into the other have been found for the reaction of silicon with oxygen and nitrous oxide (Fig. 6) [29]. The reaction of oxygen with silicon proceeds as a multistage two-pathway process whose basic features are described by the scheme



Here, account is taken of the fact that, in thin oxide layers, SiO_2 reduction is possible, characterized by the rate constant k_5

$$SiO_{4/2} + Si \longrightarrow (\rangle SiO),$$

leading to the regeneration of a clean silicon surface. The activation energy of this stage is 403.8 kJ/mol. The above scheme disregards the reverse reaction of disproportionation:

$$2SiO_{gas} \longrightarrow Si + SiO_2.$$

Estimates show that the activation energy of such a reaction is approximately 146 kJ/mol. This reaction may lead to the formation of stacking faults at the $Si–SiO_2$ interface—a phenomenon invariably observed in the thermal oxidation of silicon.

Qualitatively, three basic cases should be distinguished. First, at low oxygen pressures and high temperatures, the main reaction pathway proceeds via stages 1 and 2. This corresponds to the mode of gas etching of silicon by oxygen, with silicon monoxide formed as the reaction product. The second pathway involves SiO_2 formation and proceeds via stages 1, 3, and 4. Finally, the third process mode is characterized by comparable contributions of the above two pathways. Such modes are related by critical conditions in terms of pressures and temperatures, which can be expressed as

$$\begin{aligned} p_{cr} &= p_0 \exp[(-E_2 - E_4 + q_3)/RT], \\ p_0 &= b_{03}^{-1} k_{02}/k_{04} \simeq b_{03}^{-1}, \end{aligned} \tag{13}$$

since $k_{02} \simeq k_{04}$. Here E_2 and E_4 are, respectively, the activation energies of stages 2 and 4; and q_3 is the heat of oxidant adsorption in stage 3.

A fundamental aspect of the scheme presented is the reversibility of stage 3, and we succeeded in demonstrating by means of ellipsometry the reversibility of the adsorption of O_2 and N_2O molecules on the oxidized silicon surface and in obtaining the isotherm of adsorption of these gases.

Occasionally, critical conditions were determined in terms of the thermodynamic model on the basis of data on the equilibrium pressure of the SiO molecule vapor. Without making a detailed analysis of the fallacy of this approach, we only mention that the critical conditions would be expected to be independent of the kind of oxidant molecules in terms of this model. However, it was established that the exponent in the expression for N_2O is 204 kJ/mol, being 376 kJ/mol for O_2. Apparently, this difference is due to the different heats of the reversible supermonolayer adsorption of O_2 and N_2O molecules. In particular, this means that the heat of adsorption on the oxidized silicon surface is higher for O_2 molecules, compared with N_2O.

The structure and characteristics of the activated complex can be rigorously established by calculating the potential surface profile for a system of interacting atoms. For the reactions in question, a simple analysis is possible, leading to important qualitative conclusions.

The specific features of the potential energy curves follow from a comparison of the thermodynamic and kinetic characteristics. The overall change ΔH in the enthalpy of the reaction

$$(1/2)A_2 + E = EA,$$

is given by

$$\Delta H = (1/2)\Delta H_1 + \Delta H_2,$$

where ΔH_1 is the enthalpy of A_2 atomization and ΔH_2 is the enthalpy of EA formation. For exothermic reactions, $\Delta H < 0$. An analysis of the experimental data for

reactions yielding products with bivalent germanium and silicon shows that

$$|\Delta H_1 + \Delta H_2| \approx E_2.$$

This means that the reaction and formation of bivalent compounds proceed without a barrier. The zero-barrier course of the reaction corresponds to a case when the activated complex gradually changes its structure from that corresponding to the minimum potential energy to the structure of the final product. The occurrence of the process with a barrier predetermines the formation of a certain specific complex which, undergoing a rearrangement, yields the final product. In some cases, the lifetime of this complex is longer than 10^{-12} s, which means that an intermediate is formed.

The reaction pathway leading to the formation of a dioxide layer on the semiconductor surface corresponds to modes lying to the right of, and above the line of, critical conditions in Fig. 6. We assume that the oxidation reactions proceed via a single identical elementary stage of rupture or transformation of valence bonds between the ⟩SiO center and lattice atoms irrespective of whether the oxidation yields a gaseous or a condensed product, i.e., whether the reaction occurs by pathway 1 or 2.

When the reaction of silicon with oxygen is performed in an open system, with the oxygen flow passing over the silicon crystal surface, there are two steady states corresponding to different pathways:

$$1)\ \mathrm{Si} + (1/2)\mathrm{O_2} \longrightarrow \mathrm{SiO},$$

$$2)\ \mathrm{Si} + \mathrm{O_2} \longrightarrow \mathrm{SiO_2}.$$

A certain set of oxygen pressures and temperatures corresponds to the occurrence of the reaction by each of these pathways. When the reaction proceeds by the first pathway, we have two stages:

$$\mathrm{O_{2\ gas}} + \mathrm{Si} \longrightarrow \mathrm{O_{2\ ads}},$$

$$\mathrm{O_{2\ ads}} \longrightarrow 2(\rangle\mathrm{SiO}),$$

$$(\rangle\mathrm{SiO}) \longrightarrow \mathrm{SiO_{gas}}.$$

The first stage proceeds rapidly via an intermediate preadsorption state to give a ⟩SiO complex. A vast body of experimental data and the results of molecular calculations indicate that the preadsorption stage corresponds to a single-center form of adsorption when an oxygen molecule loses translational degrees of freedom, with a single rotational degree of freedom retained. This corresponds to a decrease in the entropy of gas molecules by 80–160 J/(K mol). Whatever the silicon face on which the reaction proceeds, the final state corresponds to the ⟩Si=O form, which means that surface reconstruction occurs in the surface region of the crystal, being more pronounced on the (111) face. Desorption of a SiO molecule into the gas phase occurs as a synchronous rupture of bonds in a doubly coordinated silanone group.

Thus, we come to the conclusion that the steady-state oxidation of silicon by oxygen involves two structures: undissociated O_2 molecules in the preadsorption state and a doubly coordinated silanone group. This conclusion is of a rather general nature, since processes of this kind are also observed in Ge + O_2 and Ge + S_2 systems. By changing the oxygen pressure or the temperature of the system, we can pass to another steady state in which the reaction proceeds by pathway 2. These critical conditions correspond to a nearly monolayer coverage of the surface by ⟩SiO complexes and to the formation of a supermonolayer filling with oxygen. This third form of organized structures on the surface is characterized by a state of reversible supermonolayer adsorption of oxygen molecules possessing a certain degree of 2D mobility. The decrease in entropy is 40 J/(K mol).

REACTION ZONE IN SEMICONDUCTOR OXIDATION

Vast experimental material has been accumulated for the reaction of the thermal oxidation of silicon. The whole set of the established relationships makes it possible to represent the processes occurring in the silicon–oxygen system schematically in Fig. 7. The oxidation process can be described by the following sequence of stages:

$$(1)\ \mathrm{O_2^{oxide}} \rightleftharpoons \mathrm{O_2^{react.zone}},$$

$$(1a)\ \mathrm{Si} + (1/2)\mathrm{O_2^{react.zone}} \longrightarrow \rangle\mathrm{SiO}\cdot\mathrm{O_{ads}}$$

(instantaneously)

$$(2)\ \rangle\mathrm{SiO}\cdot\mathrm{O_{ads}} \longrightarrow \rangle\mathrm{SiO_{2/2}}\cdot\mathrm{O_{ads}},$$

$$(3)\ \rangle\mathrm{SiO_{2/2}}\cdot\mathrm{O_{ads}} \longrightarrow \mathrm{SiO_{4/2}}.$$

In the first stage, oxygen dissolved in silicon passes into the reaction zone and then the dissociative chemisorption of the oxygen molecule occurs instantaneously. The second stage corresponds to the transfer of a silicon atom from the crystal to the SiO_2 matrix, and the third, completing the process, to the incorporation of the formed structural $SiO_{2/2}O_{ads}$ unit into the $SiO_{4/2}$ matrix. The thermal oxidation of semiconductors, analyzed as a reaction at the ⟨covalent crystal⟩-dioxide interface, basically reduces to the mechanism of semiconductor dissolution in the condensed phase. The meaning of this approach consists in that the rate-determining stages of formation of both the gaseous silicon monoxide and a structural unit of silicon dioxide are similar in nature, being associated with the zero-barrier rupture or transformation of chemical bonds of the doubly coordinated silanone group SiO with the lattice atoms.

A specific feature of the oxidation process consists in that the activation energy of this stage is lower than

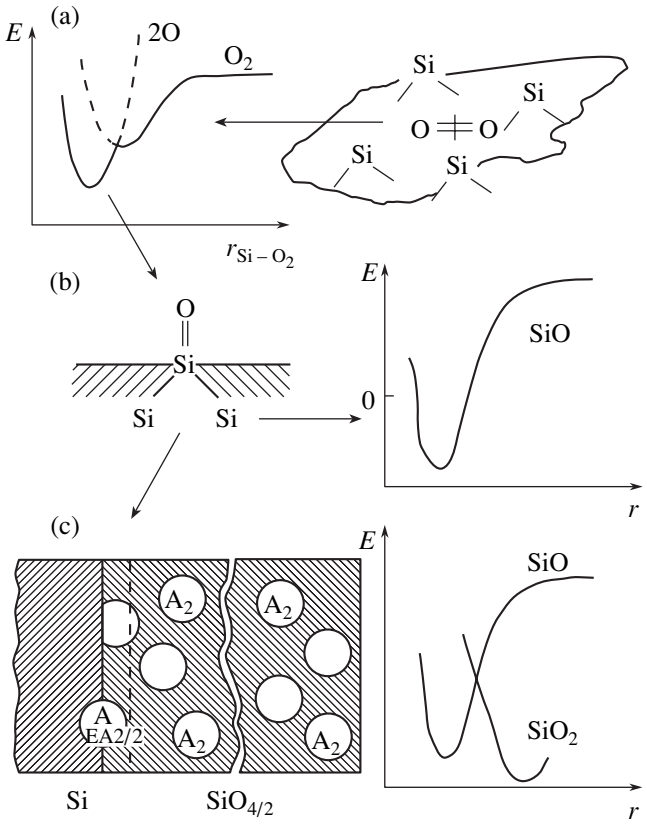


Fig. 7. Evolution of the reaction zone constitution and energy profiles along the reaction coordinate in the silicon–oxygen system: (a) preadsorption state and dissociative adsorption of O_2; (b) stage of SiO desorption; (c) thermal oxidation of silicon; A, oxygen atom; E, silicon atom.

the activation energy of gaseous product formation by the energy of rearrangement of the silicon–oxygen network of silicon dioxide.

Zero-barrier reactions are characterized by the value $\alpha = 1$. By definition, the reorganization energy corresponds to energy released upon the transfer of a particle of the given kind from a vacuum into the medium of interest. When the reaction product has the same chemical composition as the medium, the quantity is equivalent to the heat of evaporation or sublimation. In the cases of polar and molecular solutions, the times of molecular relaxation of the medium are so short that the activated complex or the intermediate formed in the course of dissolution can be considered completely solvated. However, in the thermal oxidation of semiconductors, the rates of structural relaxation are comparable with the rate of chemical reaction at the phase boundary.

Table 2. Enthalpy and entropy of self-organization of the reaction zone in thermal oxidation of silicon and germanium

System	$-\Delta H$, kJ/mol	$-\Delta S$, J/(K mol)
Si/O_2	134	111
Ge/O_2	58	65
Ge/S_2	30	42

In the oxide bulk, silicon atoms have fourfold coordination, whereas the intermediate formed in the rate-determining stage of the process is only threefold coordinated. This feature is common to all of the considered reactions of the thermal oxidation of semiconductors. Relying upon this factor, the nearest neighbor approximation is quite exact for systems with covalent bonding; we can calculate the energies of reorganization of the intermediates formed in the reactions under consideration. It can be readily seen that the quantity of interest is given by

$$\Delta H_{gas} = \Delta H_{E=O} - 2\Delta H_{E-O},$$

where $\Delta H_{E=O}$ is the energy of the double bond in the E=O molecule, e.g., in SiO, and ΔH_{E-O} is the energy of the single bond E–O, e.g., Si–O. Using data for the germanium–sulfur, germanium–oxygen, and silicon–oxygen systems, we obtain ΔH values presented in Table 2. These quantities can be readily evaluated using chemical bond strengths. In absolute value, the reorganization energy ΔH is the highest for the silicon–oxygen system. This means that this system is characterized by the maximum efficiency of incorporation of semiconductor atoms into the network of the glassy dielectric layer. Using the ΔH value and the temperature corresponding to the medium value of the interval in which oxidation reactions occur in the systems in question, the changes in entropy, ΔS, associated with self-organization of the reaction zone under the steady-state oxidation conditions were determined.

CONCLUSION

We considered the effects of the self-organization of phase boundaries (phase discontinuity surfaces and reaction zones) for the systems underlying the basic technological processes of microelectronics in typical semiconductors. The analysis we performed demonstrated that it is the silicon–oxygen system that exhibits the strongest self-organization of the reaction zone, commonly called the silicon–⟨silicon dioxide⟩ interface.

It is significant that, previously, certain hopes for obtaining electrically perfect semiconductor surfaces were placed on the high-vacuum technology, and that results obtained in attempts to create a virtually surface state–free semiconductor–electrolyte interface have not been comprehended until recently. It has now become clear that it is the effects of self-organization of the semiconductor–electrolyte and semiconductor–insulator interfaces that underlie the technology yielding electrically perfect semiconductor surfaces. The absence of surface electronic states, including recombination centers, in the silicon–⟨silicon dioxide⟩ system is

a phenomenon of fundamental importance, which predetermined the present-day level of the technology of integrated circuits. The analysis we made indicates that the degree of self-organization is highest, among the numerous possible systems, for the silicon–⟨silicon dioxide⟩ system. That is why, of the large number of chemical compounds and elements possessing semiconducting properties, only silicon has been successfully used in the industry [2, 30].

REFERENCES

1. A. V. Rzhanov, *Electronic Processes on Semiconductor Surface* (Nauka, Moscow, 1971).
2. S. M. Repinskiĭ, Poverkhnost, Nos. 7–8, 12 (1995).
3. I. Prigogine, *From Being to Becoming: Time and Complexity in the Physical Sciences* (Freeman, San Francisco, 1980; Nauka, Moscow, 1985).
4. M. Eigen, Naturwissenschaften **58**, 465 (1971); *Selforganization of Matter and the Evolution of Biological Macromolecules* (Mir, Moscow, 1973).
5. Yu. L. Klimontovich, Usp. Fiz. Nauk **158** (1), 59 (1989) [Sov. Phys. Usp. **32**, 416 (1989)].
6. J. W. Gibbs, in *The Collected Works of J. Willard Gibbs* (Yale Univ. Press, New Haven, 1948; Nauka, Moscow, 1982), Vols. 1, 2.
7. A. V. Rusanov, *Thermodynamics of Surface Phenomena* (Leningr. Gos. Univ., Leningrad, 1960).
8. G. K. Boreskov, *Heterogeneous Catalysis* (Nauka, Moscow, 1988).
9. B. A. Nesterenko and O. V. Snitko, *Physical Properties of Atomically Clean Semiconductor Surface* (Naukova Dumka, Kiev, 1983).
10. B. A. Nesterenko and V. G. Lyapin, *Phase Transitions on Free Faces and Interphase Boundaries in Semiconductors* (Naukova Dumka, Kiev, 1990).
11. V. G. Lifshits, *Electron Spectroscopy and Atomic Processes on Silicon Surface* (Nauka, Moscow, 1985).
12. S. M. Repinskiĭ, *An Introduction to the Chemical Physics of Solid Surface* (Nauka, Novosibirsk, 1993).
13. B. Z. Olshanetsky and A. A. Shklyaev, Surf. Sci. **67**, 681 (1977).
14. B. Z. Olshanetsky, A. A. Shklyaev, and S. M. Repinsky, Surf. Sci. **69**, 205 (1977).
15. M. Wautlet, Surf. Sci. **134**, 803 (1983).
16. A. V. Kozhukhov, B. Z. Kanter, S. I. Stenin, and Z. Sh. Yanovitskaya, Poverkhnost, No. 7, 54 (1988).
17. G. J. Russel and D. J. Haneman, J. Electrochem. Soc. **114** (4), 398 (1967).
18. A. V. Rzhanov and S. M. Repinskiĭ, Zh. Fiz. Khim. **52** (12), 3044 (1978).
19. S. M. Repinskiĭ, Usp. Khim. **52** (6), 922 (1983).
20. M. R. Baklanov and S. M. Repinskiĭ, Poverkhnost, No. 3, 79 (1984).
21. M. R. Baklanov, S. M. Repinskiĭ, and A. V. Rzhanov, Kinet. Katal. **31** (2), 292 (1990).
22. M. R. Baklanov and S. M. Repinsky, Surf. Sci. **88**, 427 (1979).
23. S. M. Repinskiĭ, V. M. Eliseev, and L. L. Sveshnikova, Zh. Fiz. Khim. **55** (7), 1737 (1981).
24. V. M. Eliseev, S. M. Repinskiĭ, and L. L. Sveshnikova, Neorg. Mater. **17** (9), 1529 (1981).
25. S. M. Repinsky and O. I. Semenova, Thin Solid Films **75**, 391 (1981).
26. S. M. Repinskiĭ and O. I. Semenova, Neorg. Mater. **14** (5), 831 (1978).
27. V. A. Myamlin and Yu. V. Pleskov, *Electrochemistry of Semiconductors* (Nauka, Moscow, 1965).
28. M. R. Baklanov and S. M. Repinskiĭ, Izv. Sib. Otd. Akad. Nauk SSSR, Ser. Khim., No. 14, 89 (1981).
29. M. R. Baklanov, V. N. Kruchinin, S. M. Repinsky, and A. A. Schklyaev, React. Solids **7**, 1 (1989).
30. S. M. Repinskiĭ, Priroda, No. 5, 8 (1997).

Translated by M. Tagirdzhanov

Semiconductors, Vol. 35, No. 9, 2001, pp. 1018–1021. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1063–1066.

SEMICONDUCTOR STRUCTURES, INTERFACES, AND SURFACES

Diffusion of Cu over a Clean Si(111) Surface

A. E. Dolbak, R. A. Zhachuk, and B. Z. Olshanetsky*

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: olshan@isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 14, 2001

Abstract—Auger electron spectroscopy and low-energy electron diffraction were used to study the diffusion of Cu over an atomically clean Si(111) surface. It is found that the diffusion gives rise to concentration distributions of copper with a sharp boundary and to the formation of the Si(111)–"5×5"–Cu surface phase. It is shown that transport of copper over the Si(111) surface occurs via the solid-state spreading process. The temperature dependence of the coefficient of Cu diffusion over the Si(111) surface was determined as given by $D_{Cu} = 10^4 \exp(-1.9/kT)$ cm^2/s. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Studies of surface diffusion are important for surface physics and for its application in technology. It is well known that surface diffusion may depend on surface orientation, surface structure, the concentration of atomic steps, the chemical nature and concentration of adsorbed atoms, external fields, and on other factors (see, for example, [1–4]). Thus, Gavrilyuk and Lifshits [5] used low-energy electron diffraction (LEED) and Auger electron spectroscopy (AES) to show that the coefficient of diffusion of Au over the Si surface depends on the surface structure induced by Au. In contrast, Dolbak *et al.* have shown [6] that the transport of Ni over the Si surface is accomplished via the diffusion of Ni atoms through the Si bulk with the subsequent segregation of these atoms at the surface due to the reduction of Ni solubility in the bulk during cooling of the sample rather than via the surface diffusion itself. Similar results have been obtained in studies of Co diffusion over the Si surface [7, 8].

Although the Cu/Si(111) system is comparatively well understood [9–16], information about Cu surface diffusion is lacking. Most publications are concerned with studying the incommensurable quasi-(5×5) structure formed after annealing at temperatures of 300–600°C [9]. This structure is often designated as Si(111)–"5×5"–Cu. The formation of the "5×5" structure is completed when the surface is covered by one monolayer (ML). Further deposition of Cu results in the formation of Cu_3Si islands that coexist with the "5×5" surface structure. It is well known that Cu has a high diffusivity in the Si bulk [17]. The objective of this study was to determine the mechanism of copper diffusion over the silicon surface and to measure the diffusion coefficients.

2. EXPERIMENTAL

In our experiments, we used the *p*-Si(111) samples with a resistivity of 5–10 Ω cm, a surface area of 22×5 mm^2, and a thickness of 0.3 mm. The surface was cleaned in a vacuum chamber with a residual pressure of $(1–2) \times 10^{-10}$ Torr by annealing for 1–2 min at a temperature of $T = 1200$°C. The samples were heated by passing an alternating current through them. The sample temperature was measured using an optical pyrometer.

In order to study the copper diffusion over the silicon surface, we sputter-deposited a Cu strip with a sharp boundary onto an atomically clean Si(111) surface in a chamber designated for the preliminary treatment of the samples. The coating thickness at the strip was about 40 ML; i.e., it was sufficiently large to ensure that the Cu source strength remained unchanged during the experiment. The copper fragments welded to a tantalum strip were used as the sputtering source. The content of impurities in the Cu was no higher than 0.001 at. %. The Ta strip was heated by an electric current. We calibrated the Cu atom flux using the MNN (60 eV) and LMM (920 eV) Cu Auger peaks and a Si Auger peak (92 eV). The elemental-sensitivity factors were taken from the relevant handbook [18]. In order to determine the coverage of the surface by Cu, we compared our data with those reported by Dallaporta [13], in which case a correspondence was set up between a certain Cu coverage measured with a quartz microbalance and the intensity of the Cu and Si Auger signals. After the sample was heated, the LEED patterns obtained from the Cu strip surface corresponded to the Si(111)–"5×5"–Cu surface structure, which is probably caused by the formation of Cu_3Si islands [9–12] on the surface; these

islands are separated by surface areas with the aforementioned structure. The electron-beam diameter in the Auger spectrometer was about 30 μm, and the electron-beam diameter in the LEED system was about 0.8 mm. The Cu deposition rate was 0.2 ML/min at a pressure of $p \approx 5 \times 10^{-10}$ Torr.

3. RESULTS AND DISCUSSION

The experiments were conducted in the temperature range of 550–650°C. Studies at temperatures below 550°C were not performed because, at these temperatures, long annealings were required to obtain the diffusion profiles that could be resolved by the AES method. If the temperatures were higher than 650°C, the evaporation of copper from the silicon surface became appreciable, which was determined from the disappearance of the signal corresponding to the Si(111)–"5 × 5"–Cu structure in the LEED pattern and from a decrease in the Cu Auger peak heights. After the sample was annealed for a time t at a certain temperature, it was cooled to room temperature with a rate no lower than 50 K/s. Further, we used the AES to measure the copper-concentration distribution $C_{Cu}(x)$ over the surface (x is the distance from copper strip edge) and studied the surface structure using the LEED patterns.

The experiments showed that the Auger signals from the silicon surface can be recorded during the sample annealing, i.e., during diffusion. The Cu concentration measured at a certain point at the surface at a distance x from the copper strip edge is equal to the concentration measured at the same point after the sample had been rapidly cooled to room temperature. This indicates that the Cu transport over the Si surface occurs via surface diffusion itself.

A typical copper distribution $C_{Cu}(x)$ after an annealing for 16 h at a temperature of 600°C is shown in Fig. 1. According to the LEED data, the aforementioned Si(111)–"5 × 5"–Cu surface structure is formed at the silicon surface with adsorbed copper. This distribution has a sharp boundary. The shape of this distribution suggests that the copper diffusion proceeds by the mechanism of the solid-phase spreading [2]. This mechanism is effective when the diffusion coefficient of atoms of the substance that spreads over the clean substrate surface is much smaller than the coefficient of diffusion over the surface phase induced by this substance at the substrate surface. In this situation, the atoms of transported substance included in the surface phase remain stationary. Propagation of the diffusion front is accomplished by diffusion of the spreading-substance atoms over the surface phase. When the diffusion front reaches these atoms, they react with the clean surface and form the Si(111)–"5 × 5"–Cu surface phase, thus increasing the surface area occupied by this phase.

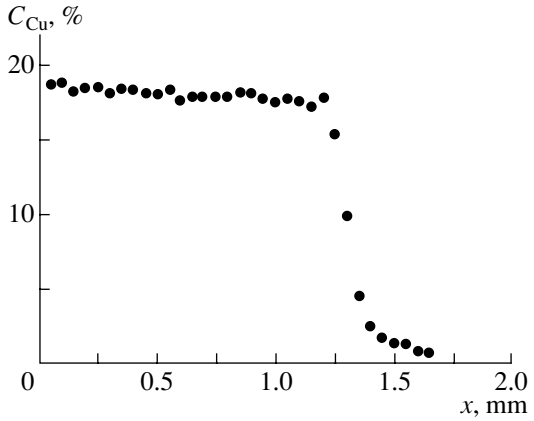


Fig. 1. Typical distribution of copper concentration $C_{Cu}(x)$ on the atomically clean Si(111) surface after annealing for 16 h at a temperature of 600°C.

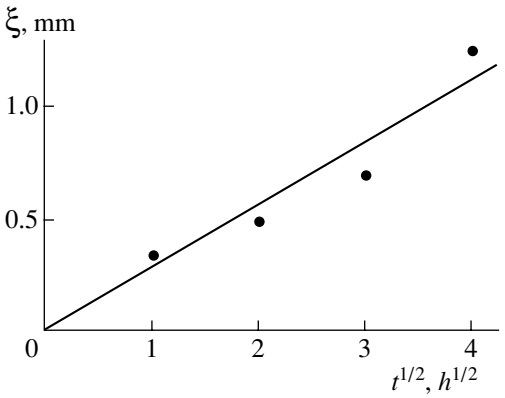


Fig. 2. Dependence of the Cu diffusion-front position on duration t of heat treatments of the sample at a temperature of 600°C.

In order to check the assumption about the diffusion mechanism, we prepared the copper-diffusion source in the shape of a 1-ML-thick strip. For such a coverage with copper, the formation of the Si(111)–"5 × 5"–Cu surface phase at the silicon surface is completed. The annealing of a sample with such a strip for 8 h at a temperature of 600°C did not result in the emergence of the copper concentration distribution over the surface. This experiment provided support for the conclusion that copper atoms incorporated into the Si(111)–"5 × 5"–Cu structure are not involved in diffusion.

The problem of diffusion from a constant-strength source according to the mechanism of solid-phase spreading has been solved [19, 20] for the limiting case when the coefficient of diffusion over the clean surface is equal to zero. The diffusion-related distribution of concentration after a lapse of time t is given by

$$C(x, t) = C_0\left(1 - \frac{\mathrm{erf}[x/2(Dt)^{1/2}]}{\mathrm{erf}(q^{1/2})}\right) \text{ for } 0 < x < \xi, \quad (1)$$

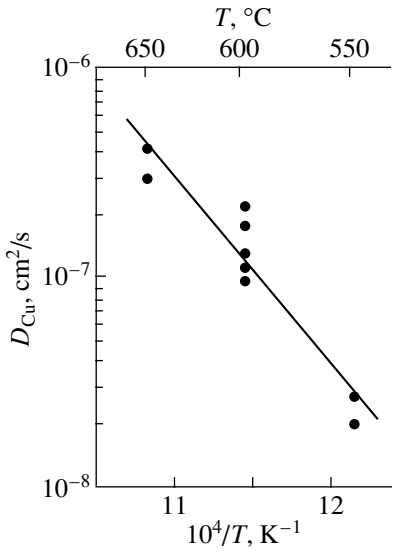


Fig. 3. Temperature dependence of coefficient of Cu diffusion over the Si(111)–"5 × 5"–Cu surface.

$$\xi = 2(qDt)^{1/2} \text{ for } t > 0. \quad (2)$$

Here, $C_0 = C(0, t)$ is the copper concentration in the source; x is the distance from the source boundary; D is the diffusion coefficient of Cu atoms; t is the diffusion duration; ξ is the diffusion-front position; and q is a constant determined by solving the equation

$$C_0 C_{5\times5} = e^q (q\pi)^{1/2} \text{erf}(q^{1/2}), \quad (3)$$

where $C_{5\times5}$ is the surface copper-atom concentration equal to the concentration in the Si(111)–"5 × 5"–Cu surface structure. The value of C_0 is close to $C_{5\times5}$ if Cu diffuses over the S(111) surface. Apparently, this is related to the fact that the escape of copper atoms from the Cu_3Si copper silicide islands is limited. We did not observe the temperature dependence of C_0 in the range of 550–650°C. Assuming that $C_0 = C_{5\times5}$ and solving numerically Eq. (3), we obtain $q \approx 0.384$. Thus, using the relation $D = \xi^2/1.536t$, we can estimate the coefficient of copper diffusion over the Si(111)–"5 × 5"–Cu structure. The experimental dependence of ξ on $\sqrt{t}$ is shown in Fig. 2. It can be seen that experimental points fit adequately the straight line $\xi/\sqrt{t}$ = const. This indicates that the time required for the formation of the Si(111)–"5 × 5"–Cu structure is not a factor limiting the copper transport over the surface.

The temperature dependence of the coefficient of diffusion of the Si(111)–"5 × 5"–Cu surface is shown in Fig. 3. An insignificant deviation of experimental points from the straight line is probably caused by the onset of desorption of Cu at a temperature of 650°C [21]. In the case under consideration, the export of copper atoms to the sample bulk may be disregarded, because the Cu solubility-limited concentration in Si does not exceed about 10^{15} cm^{-3} [22] in the temperature range of 550–650°C. Thus, the temperature dependence of the coefficient of the copper diffusion over the Si(111) surface may be represented as

$$D_{Cu} = 10^4 \exp(-1.9/kT) \text{ cm}^2/\text{s}.$$

4. CONCLUSION

It is found that the transport of copper over the Si(111) surface occurs by intrinsic surface diffusion according to the "solid-phase spreading" mechanism and is accompanied by the formation of the Si(111)–"5 × 5"–Cu surface structure. We determined the temperature dependence of the coefficient of Cu diffusion over the Si(111) surface.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, Federal Program "Research and Development in Topical Fields of Civil Science and Technology" of the Ministry of Industry, Science, and Technology of the Russian Federation, and the Interstate Russian–Ukrainian Program "Nanophysics and Nanoelectronics."

REFERENCES

1. A. G. Naumovets and Yu. S. Vedula, Surf. Sci. Rep. **4**, 365 (1984).
2. Ya. E. Geguzin, in *Surface Diffusion and Spreading*, Ed. by Ya. E. Geguzin (Nauka, Moscow, 1969), p. 11.
3. Yu. S. Vedula, A. T. Loburets, and A. G. Naumovets, Zh. Éksp. Teor. Fiz. **77**, 773 (1979) [Sov. Phys. JETP **50**, 391 (1979)].
4. Yu. L. Gavrilyuk and V. G. Lifshits, Poverkhnost, No. 4, 143 (1983).
5. Yu. L. Gavrilyuk and V. G. Lifshits, Poverkhnost, No. 4, 82 (1983).
6. A. E. Dolbak, B. Z. Olshanetsky, S. I. Stenin, and S. A. Teys, Surf. Sci. **218**, 37 (1989).
7. A. E. Dolbak, B. Z. Ol'shanetskii, and S. A. Tiis, Poverkhnost, No. 11, 29 (1996).
8. M. Y. Lee and P. A. Bennett, Phys. Rev. Lett. **75**, 4460 (1995).
9. E. Daugy, P. Mathiez, F. Salvan, and J. M. Layet, Surf. Sci. **154**, 267 (1985).
10. M. Mundschau, E. Bauer, W. Telieps, and W. Swiech, J. Appl. Phys. **65**, 4747 (1989).
11. S. A. Chambers and J. H. Weaver, J. Vac. Sci. Technol. A **3**, 1929 (1985).
12. L. Calliari, F. Marchetti, and M. Sancrotti, Phys. Rev. B **34**, 521 (1986).
13. H. Dallaporta and A. Cross, Surf. Sci. **178**, 64 (1986).

14. R. J. Wilson, S. Chiang, and F. Salvan, Phys. Rev. B **38**, 12696 (1988).

15. D. D. Chamblis and T. N. Rhodin, Phys. Rev. B **42**, 1674 (1990).

16. J. Nichols, F. Salvan, and B. Reihl, Phys. Rev. B **34**, 2945 (1986).

17. A. A. Istratov, C. Flink, H. Hieslmair, *et al.*, Phys. Rev. Lett. **81**, 1243 (1998).

18. P. W. Palmberg, G. E. Riach, R. E. Weber, and N. C. Mac-Donnald, *Handbook of Auger Electron Spectroscopy* (Physical Electronics Industrials Inc., Eden Prairie, 1972).

19. H. S. Carslaw and J. C. Jaeger, in *Conduction of Heat in Solids* (Clarendon, Oxford, 1947), p. 71.

20. H. Fujita, J. Chem. Phys. **21**, 700 (1953).

21. T. Ikeda, Y. Kawashima, H. Itoh, and T. Ichinokawa, Surf. Sci. **336**, 76 (1995).

22. Eicke R. Weber, Appl. Phys. A: Solids Surf. **A30**, 1 (1983).

Translated by A. Spitsyn

Semiconductors, Vol. 35, No. 9, 2001, pp. 1022–1029. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1067–1074.

SEMICONDUCTOR STRUCTURES, INTERFACES, AND SURFACES

The Simulation of Epitaxy, Sublimation, and Annealing Processes in a 3D Silicon Surface Layer

A. V. Zverev, I. G. Neizvestnyĭ, N. L. Shvarts*, and Z. Sh. Yanovitskaya

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: natasha@spy.isp.nsc.ru*

Received February 14, 2001; accepted for publication February 15, 2001

Abstract—The Monte Carlo model for simulating epitaxy, sublimation, and annealing of the (111) surface of a diamond-like crystal is developed. The model accounts for the so-called overhanging structures and allows one to consider a 3D surface layer with a complex structure and volume comprising as many as 10^7 atom sites. The initial stage of epitaxial growth on smooth as well as on porous surfaces and the annealing of both step-containing and porous surfaces are simulated. The effect of Schwoebel barriers on the formation of compact 3D islands during epitaxy is suggested as a possible kinetic mechanism for quantum-dot formation. The sublimation behavior of monatomic steps on the (111) surfaces of a diamond-like crystal is also studied. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Molecular beam epitaxy (MBE) is a well-known technique that is conventionally used for the production of atomic-level solid-state structures [1, 2]. An ideal substrate for MBE would be a perfectly clean, atomic-smooth crystal surface; at the same time, the actual surfaces used at best include some isolated atomic-smooth terraces separated by monatomic and multiatomic steps. The initial surface relief as well as its evolution during growth markedly affect the electrical properties of nanoelectronic devices. This has motivated many researchers to perform a comprehensive investigation of the surface morphology at the atomic level with the use of various techniques providing atomic resolution [3–5]. Since it is difficult to combine the technological growth of crystals with the surface atomic structure observations, computer simulations have received wide recognition. At the present time, along with the molecular-dynamic methods, the numerical static Monte Carlo (MC) methods [6] permitting a real-time simulation of atomic processes are used extensively.

An inadequate understanding of the atomic processes on the surface of a diamond-like crystal and the requirement for high operational speed force one to use simplified 3D models. Most simulations consider the (001) surface of a cubic crystal and employ a solid-on-solid (SoS) concept, which forbids the existence of overhanging structures in the surface layer. So far, 3D simulations in the surface layer of a diamond-like crystal have been performed in a few studies only: the papers [7–11] represent an almost exhaustive list. The development of computer models which would adequately account for the properties of semiconductor crystal surfaces lags far behind the current requirements. Thus, it turns out that investigating surfaces with a complex initial relief including mesa structures, faceted surfaces, or porous surface layers calls for 3D models, which account for the actual crystal structure.

In this study, we develop a 3D lattice Monte Carlo model of epitaxial growth on the (111) surface of a diamond-like crystal without resorting to the SoS [7] concept. Taking into account the position of every atom in the 3D surface layer, the model allows for voids and any other overhanging structures. A new high-speed algorithm enables us to simulate epitaxy along with the other high-temperature processes including sublimation and annealing.

To compare the results of simulations with experimental data, the actual surface with specific features whose physical nature is difficult to be incorporated within a model must sometimes be considered. The representation of a crystal as a set of rigidly fixed lattice sites rules out the lattice transformations that are typical of most semiconductor surfaces. In particular, the (7×7) structure is always present on the atomic-smooth Si(111) surface at temperatures below 830°C [12]. In this context, the closest correlation between the results of simulations for silicon and the experimental data can be expected at high temperatures. At low temperatures, it is the Si(111) surface stabilized with boron that is not reconstructed [13]. Due to high roughness, the porous silicon surface also shows no signs of structural rearrangements [14]. Growth processes occurring on a reconstructed surface can be readily described within our model in terms of effective coefficients of surface diffusion. To account for the special effects stemming from the diffusion of atoms across the surface steps, additional energy barriers were introduced into the model. The physical origin of the barriers may

1063-7826/01/3509-1022$21.00

be related to surface defects and foreign atoms, lattice transformations, charged atoms or steps, lattice mismatch at heteroepitaxy, and so on. The model developed is employed for simulating growth, sublimation, and annealing on a smooth, stepped, and porous surface.

2. PHYSICAL AND MATHEMATICAL BACKGROUND

We consider the diffusion hop when an atom can move to a free site in the first, second, or third coordination shells as a key event. For the different coordination shells, different probabilities of the atom incorporation into a given site are allowed. An atom with dangling bonds may either occupy, with corresponding probability, a randomly chosen neighboring site or overcome the desorption barrier and be desorbed. Adsorption is regarded as the process of an atom residing at regular intervals of time at random sites on the surface.

The surface layer (SL) to be simulated is represented as a stack of monolayers (ML) in the (111) plane. A monolayer consists of atoms located in the same geometrical plane. Two neighboring monolayers form a bilayer (BL). The atoms in the upper monolayer of a bilayer have a single free bond directed normally to the surface. In the SL model, the configuration of every monolayer can be specified arbitrarily, thus allowing for the desired initial relief to be formed. With Pentium III-type computers, the consideration of surface layers with a volume incorporating about 10^7 atomic sites is possible (the corresponding surface area may include up to 10^6 atomic sites and the surface layer thickness may amount to 200 atomic layers). For the Si(111) surface, this means that the SL may have an area of 1.3×10^5 nm^2 and a thickness of 63 nm.

Let us refer to the atoms connected by a single bond with the nearest neighbor on the smooth (111) surface as free atoms or adatoms. These atoms have one nearest and three next nearest neighbors, i.e., one-quarter of the volumetric bonds. For an atom surrounded by n nearest and k next nearest neighbors ($0 < n < 4$ and $0 < k < 12$), the detachment energy E_{hop} is defined as $E_{\text{hop}} = nE_1 + kE_2$, where E_1 and E_2 are the energies of interaction with the nearest and next nearest neighbors, respectively. The parameters to be specified are the energies E_1 and E_2 and also the energies E_{1n}, E_{2n}, and E_{3n}, which are the respective energy barriers to incorporation into the sites in the first, second, and third coordination shells. The energy of adatom desorption from a smooth terrace E_d should be specified independently.

The conditions of epitaxial growth are specified by the substrate temperature T (K) and the deposition rate V (BL/s). According to [15], these two parameters of epitaxial growth are not independent. Both of them determine the rate of the two-dimensional nucleus initiation; a decrease in the deposition rate is equivalent to an increase in the substrate temperature. This relationship justifies the introduction of the model parameter n_{dif} equal to the number of diffusion hops of each adatom per adsorption event.

From the viewpoint of overcoming the diffusion barrier, the number of diffusion hops between two subsequent attachment events is expressed as

$$\nu_{\text{dif}} = \nu \exp(-E_{\text{dif}}/k_{\text{B}}T)/V, \tag{1}$$

where E_{dif} is the surface-diffusion activation energy (eV), V is the deposition rate (BL/s), ν is the thermal vibration frequency of an atom (1/s), k_{B} is the Boltzmann constant, and T is the substrate temperature.

Moreover, being in the molecular flux of intensity F, the coverage θ grows with time t as $\theta = Ft$. The minimum change in the coverage on the simulated surface with an area S corresponds to the addition of a single atom, $\Delta\theta_{\text{min}} = 1/S$. The time it takes to effect the minimum change in $\Delta\theta_{\text{min}} = 1/S$ can be found as a product of the diffusion hop duration

$$\tau_{\text{dif}} = [\nu \exp(-E_{\text{dif}}/k_{\text{B}}T)]^{-1}$$

and of the given number of diffusion hops n_{dif} each surface atom makes on average; i.e., $\Delta t = \tau_{\text{dif}} n_{\text{dif}}$. The product of the flux intensity divided by the surface area yields the deposition rate $FS = V$. Therefore, by writing

$$\frac{d\theta}{dt} = F = \frac{V}{S}, \tag{2a}$$

$$\frac{\Delta\theta_{\text{min}}}{\Delta t} = \frac{1}{Sn_{\text{dif}}\tau_{\text{dif}}}, \tag{2b}$$

and equating the right-hand sides of the expressions, we obtain formula (1).

Within the model, the epitaxial growth is completed after a preset number of monolayers have been deposited; the duration of the simulated sublimation and annealing is defined by the number of atoms being desorbed.

The configurations of the 3D surface layer are expressed in terms of the prescribed number of adsorption or desorption events. Special utility programs permit the visualization of the surface relief evolution in the form of a computer-generated movie, thus providing an opportunity to analyze the relief special features after the simulation is over. At any growth stage, the analysis yields the surface's distinctive features: the number of free bonds on the surface characterizing the surface roughness, the size distribution of the atomic or vacancy islands, the thickness of a continuous layer overgrown above pores, etc. At sections made through the SL by two perpendicular planes with specified coordinates, the transformations of buried structures are observable.

The model described above was mathematically implemented as follows. Each lattice site was assigned to an escape class defined as the number of the nearest

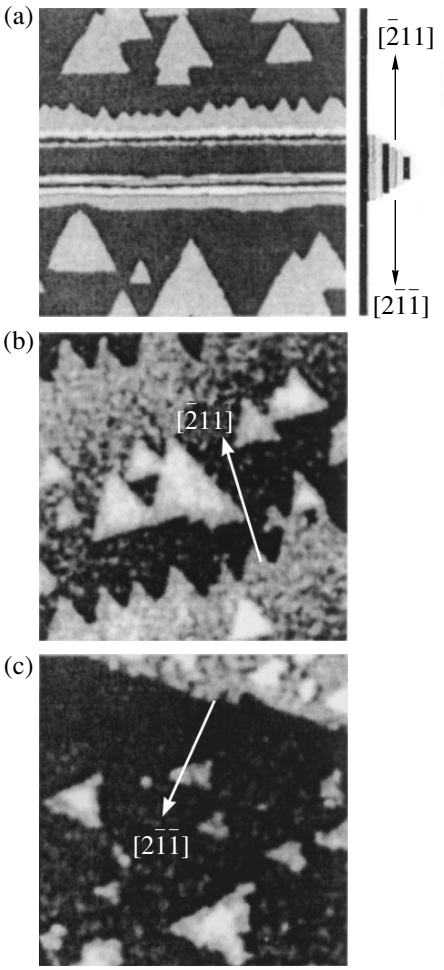


Fig. 1. Images of the Si(111) surface at the initial stages of the first-bilayer growth. Every pattern includes 300 × 300 atomic sites (a. s.), T = 690 K. (a) Simulated surface after deposition of the 0.25 bilayer (BL); an initial profile of the terraced surface is shown at the right; STM-images of Si(111) surface (the data from [25]) (b) with a step of ($\bar{2}$11) type after deposition of 0.25 BL and (c) with a step of (2$\bar{1}\bar{1}$) type after deposition of 0.18 BL.

and next nearest neighboring atoms. In a diamond-like lattice, there are 65 such classes. For a preset initial SL configuration, the array of effective probabilities for an atom to leave its site was calculated. The effective probability incorporates the probability (defined by the escape class) of an elementary detachment event and the number of atoms in the class. An element from the array of probabilities calculated was chosen randomly; then, an arbitrary monolayer containing an atom of the given class was selected. The more atoms there were belonging to the class specified in a monolayer, the more likely it became that this monolayer would be chosen. In this monolayer, the atom to be displaced was randomly chosen making use of a fast processor search. A free neighboring site was chosen in the first, second, and third coordination shells according to the incorporation probabilities. After the chosen atom had been transferred to the new site, the array of the effective probabilities was recalculated. The calculation scheme is based on the algorithm [16], which excludes "empty" events that leave the atom positions unchanged.

3. SIMULATION RESULTS

The software developed was applied to simulate the epitaxial growth and the high-temperature annealing accompanied by sublimation on smooth, vicinal, and porous Si(111) surfaces.

3.1. Epitaxial Growth on Smooth and Stepped Surfaces

In what follows, we will consider the initiation of 2D islands on a smooth Si(111) surface. Ample evidence for the formation of 2D islands on the Si(111) surface during epitaxy has been gained using scanning tunneling microscopy (STM) [17–21]. However, simulations of initial stages of the island initiation on Si(111) surfaces are almost absent. This is mostly due to the lack of models accounting for the (7 × 7) reconstruction characteristic of the growth temperatures. An experimental observation of the 2D islands initiated on a nonreconstructed surface was reported in [13]. According to [13], the activation energy of the surface diffusion on the (1 × 1) silicon surface is only 0.3 eV higher than that on the (7 × 7) surface. Furthermore, the islands initiated on the nonreconstructed surface were found to be triangular in shape, which is also typical of the (7 × 7) surface.

At the specified temperature and growth rate, the density of the islands is defined by the activation energy of a diffusion hop E_{dif} [22] (disregarding reconstruction, only an effective E_{dif} value can be suggested). After the temperature and the growth rate have been found experimentally, the E_{dif} can be determined by comparing the simulated value of the 2D nucleus density with the experiment. By using experimental data [17], we estimated the effective energy on the Si(111) surface as E_{dif} = (1.75 ± 0.15) eV [7], which is within the range obtained previously [23, 24].

With the E_{dif} derived as described above, we managed to obtain a realistic simulation for the motion of islands and steps in the [$\bar{2}$11] and [2$\bar{1}\bar{1}$] directions during the growth on the vicinal Si(111) surfaces. In particular, the step moving in the [2$\bar{1}\bar{1}$] direction remained smooth, whereas the step moving in the opposite direction [$\bar{2}$11] became hackly, which was not due to the remains of adjoining islands (Fig. 1). This behavior is in agreement with experiment [25]. The shape and the size of 1-bilayer-high islands were close to the experimental values obtained for the same coverage. The islands grow and merge together to form a continuous layer. The number of free bonds varied periodi-

cally with growth, with the period being equal to the time it takes for one bilayer to be completed. The number of free bonds oscillated much in the same manner as the intensity of an electron beam in the reflection high-energy electron diffraction (RHEED) [26]. The oscillation amplitude depended on the temperature and growth rate. With time, as the surface roughness increased, the oscillations were damped out.

3.2. Epitaxial Growth of 3D Compact Islands

We also considered epitaxial growth with additional energy barriers (the Schwoebel barriers) where an atom can cross the surface step. In several studies [27, 28], these barriers are presumed to cause the formation of 3D compact islands during heteroepitaxy; these islands are known as quantum dots (QD). Epitaxial joining of two materials with mismatched lattice constants gives rise to internal elastic stresses. This may increase the probability of the atom transition from an island edge onto the upper terrace, since this process decreases the island's size and, therefore, relieves the stress. An enhanced rate of the second-layer island initiation was attributed in [29] to the asymmetry of the Schwoebel barriers in the SiGe system.

In order to study the effect of the Schwoebel barriers on the formation of the surface relief during epitaxy, additional energy barriers E_{st} were introduced for the atoms that cross monatomic steps. The probability of a diffusion hop when crossing the step changes by a factor of $P = \exp(-E_{st}/kT)$ compared to the probability of a hop with no change in atomic level. The probabilities of atom transfer to the upper and lower layers are governed by P_{up} and P_{down} factors, respectively. The phase diagram in the P_{up}–P_{down} plane has been drawn [30] with the area of the Stranski–Krastanov growth separating the region of 2D growth from the region of 3D nucleation. The condition $P_{up}/P_{down} > 2$–5 should be met for the formation of compact 3D islands without a wetting layer.

Figure 2 shows the simulated surface after the deposition of 2 ML, where the values of P_{up} and P_{down} typical of 3D growth were chosen. By introducing the dependence of these parameters on the average size of an island, we managed to account to some extent for a change in the Schwoebel barrier with an increase in the integrated flux being deposited. The results presented in Fig. 2 were obtained for the following integrated-flux dependence of the Schwoebel barrier:

$$P(\theta_i) = 1 + (P_0 - 1)\theta_i/\theta_{cr} \quad \text{at} \quad \theta_i < \theta_{cr}, \qquad (3a)$$

$$P(\theta_i) = P_0 \quad \text{for} \quad \theta_i > \theta_{cr}, \qquad (3b)$$

where $P_0 = \exp(-E_{st}/kT)$, θ_i is the ith bilayer coverage, and θ_{cr} is a critical coverage expressed in fractions of a monolayer. The average height of an island as a function of the critical coating is plotted in Fig. 2c.

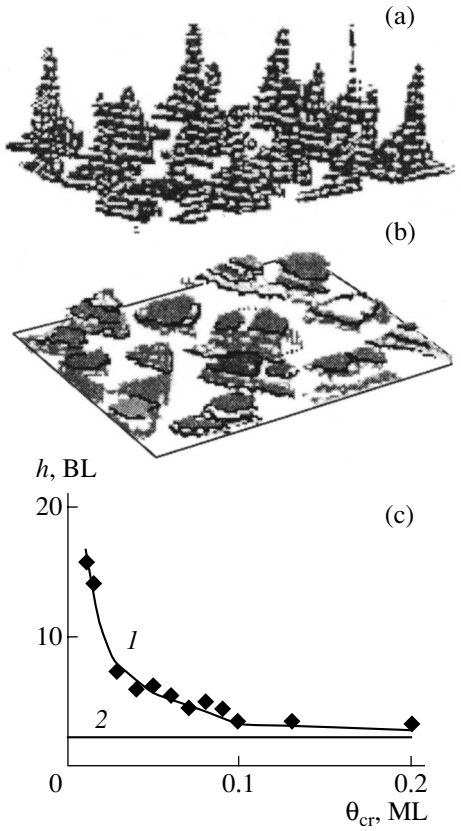


Fig. 2. Characterization of the simulated surface after deposition of 2 ML at $T = 725$ K, $P_{up} = 3$, $P_{down} = 0.3$, and critical coverage θ_{cr} = (a) 0.01 and (b) 0.1 ML; (c): (curve *1*) 3D island average height versus θ_{cr} and (curve *2*) average thickness of the deposited layer. BL stands for bilayer.

3.3. Epitaxial Growth on a Porous Surface

The considerable current interest in epitaxial growth on porous silicon (*por*-Si) surfaces is due to the new possibilities presented by *por*-Si pliable substrates [31–33] for the reduction of mechanical stress stemming from the lattice mismatch of the structures during heteroepitaxy. Furthermore, *por*-Si seems to hold much promise for developing the new silicon-on-insulator (SOI) technology [34].

Key parameters for the silicon heteroepitaxial growth to be initiated are the deposition dose that ensures the complete coverage of pores and the dose required for the surface smoothing. The only means for *in situ* quality control of the growing surface is presented by the RHEED technique. The reflections from the surface superstructures are sometimes attributed to a continuous film formed on the porous substrate [14]; however, they are only indicative of the appearance of perfectly structured surface fragments, which can easily coexist with areas containing open pores. Simulation proves to be a useful tool for estimating the minimum dose required for the pore coverage and the sur-

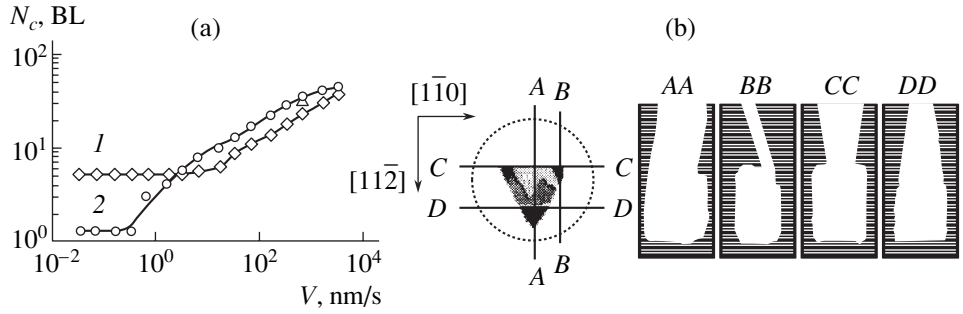


Fig. 3. (a) Critical integrated flux N_c (measured in bilayers (BL)) as a function of the deposition rate V at T = 1073 K; pore diameter d = 14 nm; porosity P_s = (*1*) 13% and (*2*) 50%. (b) Filling of a 100-BL-deep pore: at the left, top view of the pore opening, where the relative sizes of the pore opening before epitaxy and after the deposition of 7.8 BL with V = 0.1 BL/s are depicted with the dashed circle and the internal triangle, respectively; at the right, pore sections with the planes perpendicular to (111) surface.

face smoothing as a function of the growth conditions and the surface parameters.

In what follows, we consider the simulation of the pore coverage during Si/Si(111) epitaxy. We are concerned with the critical dose N_c, which is the minimum number of the deposited BLs required for the formation of a continuous surface layer. Cylindrical pores of the same size were assumed to be normal to the (111) surface and uniformly distributed over the initial surface; the deposited atom flux was also assumed to fall normally on the surface. For most of the simulations, the substrate temperature was set equal to 800°C. We varied the parameters of the porous substrate according to the experiments [14, 32] in the following ranges: porosity 13% ≤ P_s ≤ 64% and the pore diameter d = 3.5–15 nm.

The critical dose N_c versus the deposition rate is presented in Fig. 3a. It can be seen that, at small deposition rates, N_c is dose-independent, since all of the atoms incident on the surface are adsorbed by pores and N_c is limited by the deposition dose. An increase in N_c with porosity is due to a reduction in the area within which adatoms are collected and to an enhancement of the pore volume to be filled. A decrease in the adatom diffusion length with deposition rate leads to a reduction in the area from which the material is fed to each pore, thus resulting in the N_c growth.

For high deposition rates, we observed a decrease in the critical dose N_c with porosity. This is connected with an increase in the total feed area for the pores with a simultaneous lowering of density of islands initiating between pores. Both the surface porosity and the pore diameter affect the flux density of the particles incident on a pore side wall from the surface. Inside the pores that are deeper than the adatom diffusion length along the wall, a "bottleneck" is formed near the surface and these pores are filled more slowly than shallower pores of the same diameter, where the adatoms from the surface could reach the bottom. Figure 3b shows the faceted bottleneck in a deep pore. The bottleneck turns out to be faceted by three pairs of parallel planes of (015) type.

4. SIMULATION OF ANNEALING AND SUBLIMATION PROCESSES

4.1. Porous Surface

High-temperature annealing is an indispensable preliminary stage for epitaxial growth on nearly every type of surface. A porous surface is no exception. The annealing of a porous Si(111) surface at T = 1400 K was simulated with the result that, at the initial stage of annealing, small pores are etched and become triangular, which is typical of the (111) surface; then, the surface smoothens owing to the diffusion and sublimation effects. Deep pores that annealed at the same temperature become covered by a continuous layer as atoms diffuse from the upper layers along the pore side walls. After that, the surface smoothens quite rapidly leaving the pores buried in the bulk [35]. Prolonged annealing results in a deep pore splitting into separate fragments, which tend to the equilibrium shape (Fig. 4). Experimental observations of the fragmentation were reported in [36].

The results presented in this section give an example of the vast possibilities that simulation offers for studying the processes of epitaxy and annealing on a porous surface layer.

4.2. Vicinal Surface

Recent interest has been focused on the behavior of silicon surface steps during sublimation and the role of vacancies during high-temperature annealing [37–44]. However, the number of reported relevant simulations is only moderate compared to the number of the related experimental studies. In what follows, we consider the results of simulating high-temperature annealing on a Si(111) vicinal surface.

The Si(111) vicinal surfaces most often involve two types of steps perpendicular to the ⟨112⟩ direction. The

steps where the surface atoms have one or two dangling bonds belong to the I and II types, respectively. The step structure accounts for the asymmetry between the atomic fluxes coming to the upper and the lower terraces and results in the surfaces with steps of different type behaving differently.

We simulated the behavior exhibited by steps of both types during sublimation and compared the results with the Schwoebel theory [45]. Although Schwoebel considered the motion of steps in relation to the asymmetry of atomic fluxes coming to the step from the upper and lower terraces for the case of epitaxial growth, the analytical expressions obtained in [45] are also valid for sublimation.

Let the initial width distribution of terraces be given in the form

$$S_k(t = 0) = L(1 + A\cos Kk), \tag{4}$$

where S_k is the width of the terrace adjacent to the kth step from below, L is a terrace average width, $A = \Delta L/L$ is the maximum relative deviation of a terrace width from the average width, and $K = 2\pi/n$ is defined by the number of terraces per period n. Then, the time dependence of the terrace widths $S_k(t)$ can be expressed as

$$\begin{aligned} S_k(t) = L\{1 + A\exp[Ct(\gamma_- - \gamma_+)(1 - \cos K)] \\ \times \cos[Kk + Ct(\gamma_+ + \gamma_-)\sin K]\}, \end{aligned} \tag{5}$$

where γ_+ and γ_- are the probabilities of the atom transfer to the upper and lower terraces and $C\gamma_+$ and $C\gamma_-$ denote the atom fluxes onto the upper and lower terraces, respectively.

According to (5), the terrace size oscillates in time with the period defined by the total atomic flux from the step. The amplitude of oscillations increases when the atomic flux to the lower terrace prevails and decreases when atoms arrive predominantly at the upper terrace.

The sublimation process simulated on the surface containing steps of type II is illustrated in Fig. 5a. The initial surface profile contains two terraces of three steps each; cyclic boundary conditions make this distribution periodic. The time dependences of the terrace widths $S_3(t)$ and $S_4(t)$ are completely described by expression (5). From the curve oscillation period, the total atomic flux from a step to both the adjacent terraces can be derived, while a decrease in the oscillation amplitude unambiguously indicates that the arrival of atoms at the upper terrace is dominant. Logarithmic decrement specifies the difference of the fluxes to the upper and lower terraces $C(\gamma_+ + \gamma_-)$. For the profile under consideration, our simulation yields

$$C(\gamma_+ + \gamma_-) = 4 \times 10^{-2} t_d,$$

$$C(\gamma_+ - \gamma_-) = 1.5 \times 10^{-3} t_d,$$

where t_d is the interval of time between two desorption events.

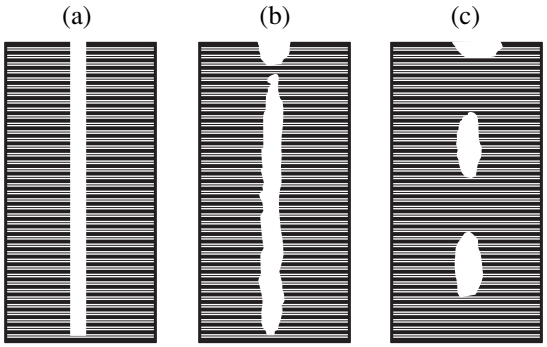


Fig. 4. Kinetics of a deep-pore annealing at $T = 1573$ K, $d = 3$ nm, $h_0 = 105$ BL: (a) the initial pore; the same pore after annealing for (b) 1300 and (c) 7000 t_d (t_d is the time between two events of desorption from the simulated surface).

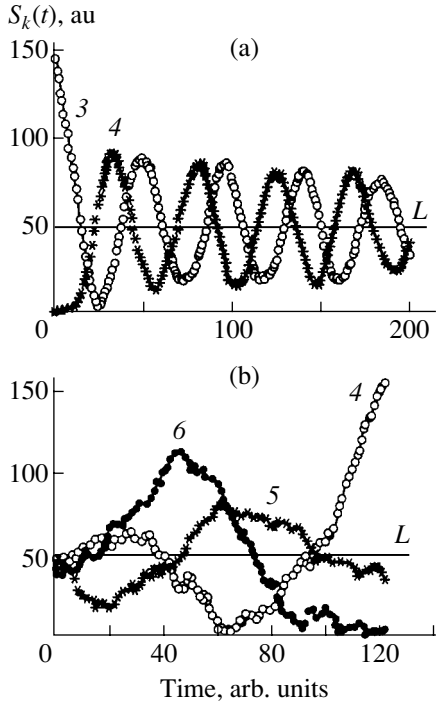


Fig. 5. Terrace width versus the duration of sublimation (curve numbers correspond to those of the terraces k); here, the substrate temperature $T = 1400$ K and the model parameters of the crystal are $E_1 = 1.3$ eV, $E_2 = 0$ eV, $E_d = 1.1$ eV, and $E_{1n} = E_{2n} = E_{3n} = 0$; (a) steps of type II: the width exhibits periodic oscillations with $A = 0.96$ and (b) steps of type I: the width changes nonperiodically with $A = 0.2$. S_k is measured in atomic units (a.u.), $L = 50$ a.s. (atomic sites).

For surfaces with steps of type I, an increase in the maximum deviation of the step width from the average value is plotted in Fig. 5b. It can be seen that the fourth terrace grows considerably while the sixth one vanishes. This behavior of the terrace sizes points to the increasing role of the exponential term in (5) and, therefore, to the prevailing transfer of atoms from the step to

the lower terrace. Total flux to the terraces was estimated also from the characteristic period of changes in the width of a terrace with steps of type I: $C(\gamma_+ + \gamma_-) = 2 \times 10^{-2} t_d$, which is half that for the steps of type II. These values are in good agreement with those we obtained for steps of both types by measuring the relative velocities of their motion during sublimation [46].

To emphasize the main idea of this section, our simulation has demonstrated that the asymmetry of atom fluxes from a step to the neighboring terraces is related to the diamond-like lattice structure.

5. CONCLUSION

A 3D model of a diamond-like crystal was developed and successfully employed in studying the epitaxial growth and annealing on a single-crystal Si(111) surface with a complex relief. From the comparison of the simulated density of islands with experimental data, the effective activation energy required for a diffusion hop of a single atom on a smooth Si(111) surface was estimated as $E_{\text{dif}} = 1.75 \pm 0.15$ eV.

The effect of the Schwoebel barriers on the formation of compact islands was studied. The islands are found to form if the probability of an atom leaving the step for the upper layer mostly exceeds the probability of a downwards transition. This relation between the barriers results in a decrease in the island's lateral size, which can lead, during heteroepitaxy on a real surface, to the elastic stress relaxation through the formation of 3D islands.

Filling of pores on the Si(111) surface is studied thoroughly for different conditions of homoepitaxial growth and various initial parameters of a porous substrate. A weak dependence of the critical dose N_c required for a continuous layer to be formed on the surface porosity was found for the typical experimental temperatures and deposition rates. The buried-pore shape is investigated in relation to the porous surface properties and the growth conditions. At low deposition rates, faceting of the pore walls with (015) planes was observed.

Simulating annealing on a porous surface, we found that the shallow and deep pores behave differently: in particular, shallow pores become wider as a result of etching and the deep pores become covered with a continuous surface layer. Prolonged high-temperature annealing resulted in the fragmentation of the deep buried pores.

We also studied the behavior of monatomic steps during sublimation on Si(111) surfaces. The difference in the fluxes to the upper and lower terraces was attributed to the diamond-like lattice geometry.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 99-02-167742), the Federal Special Scientific and Technological Program "Research and Development in the Most Important Fields of Civil Science and Technology", and Subprograms "Promising Technologies and Devices in Microelectronics and Nanoelectronics" and "Surface Atomic Structures".

REFERENCES

1. N. D. Zakharov, P. Werner, U. Gosele, *et al.*, Appl. Phys. Lett. **76**, 2677 (2000).
2. K. Zhang, Ch. Heyn, W. Hansen, *et al.*, Appl. Phys. Lett. **76**, 2229 (2000).
3. T. T. Tsong, Prog. Surf. Sci. **64**, 199 (2000).
4. A. V. Latyshev, A. B. Krasilnikov, and A. L. Aseev, Thin Solid Films **306**, 205 (1997).
5. A. A. Shklyaev, M. Shibata, and M. Ichikawa, J. Appl. Phys. **88**, 1397 (2000).
6. *Monte Carlo Methods in Statistical Physics*, Ed. by K. Binder (Springer-Verlag, Berlin, 1979; Nauka, Moscow, 1981).
7. A. V. Zverev, I. G. Neizvestnyĭ, N. L. Shvarts, and Z. Sh. Yanovitskaya, Mikroélektronika **28**, 377 (1999).
8. J. Dalla Torre and M. D. Rouhani, J. Appl. Phys. **84**, 5487 (1998).
9. P. L. Novikov, L. N. Aleksandrov, A. V. Dvurechenskiĭ, and V. A. Zinov'ev, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 512 (1998) [JETP Lett. **67**, 539 (1998)].
10. D. L. Woodraska and J. A. Jaszczak, Surf. Sci. **374**, 319 (1997).
11. S. Kersulis and V. Mitin, Semicond. Sci. Technol. **10**, 653 (1995).
12. A. V. Latyshev, A. B. Krasilnikov, and A. L. Aseev, Appl. Surf. Sci. **60/61**, 397 (1992).
13. H. Grube and J. J. Boland, Surf. Sci. **407**, 152 (1998).
14. Y. Yasumatsu, T. Ito, H. Nishizawa, and A. Hiraki, Appl. Surf. Sci. **48/49**, 414 (1991).
15. C. Heyn, T. Franke, R. Anton, and M. Harsdorft, Phys. Rev. B **56**, 13483 (1997).
16. P. A. Maksym, Semicond. Sci. Technol. **3**, 594 (1988).
17. B. Voigtlander, A. Zinner, T. Weber, and H. P. Bonzel, Phys. Rev. B **51**, 7583 (1995).
18. B. Voigtlander, M. Kastner, and P. Smilauer, Phys. Rev. Lett. **81**, 858 (1998).
19. J. Myslivecek, T. Jarolimek, P. Smilauer, *et al.*, Phys. Rev. B **60**, 13869 (1999).
20. T. Sato, S. Kitamura, and M. Iwatsuki, Surf. Sci. **445**, 130 (2000).
21. I.-S. Hwang, Mon-Su Ho, and T. T. Tsong, Phys. Rev. Lett. **83**, 120 (1999).
22. J. A. Venables, G. D. T. Spiller, and M. Hanbucken, Rep. Prog. Phys. **47**, 399 (1984).
23. M. I. Larsson and G. V. Hansson, Surf. Sci. Lett. **321**, 1261 (1994).
24. A. V. Latyshev, A. B. Krasilnikov, and A. L. Aseev, Thin Solid Films **281–282**, 20 (1996).
25. M. Fehrenbacher, H. Rauscher, U. Memmert, and R. J. Behm, Surf. Sci. **398**, 123 (1997).

26. T. Shitara, D. D. Vedensky, M. R. Wilby, *et al.*, Phys. Rev. B **46**, 6815 (1992).
27. V. A. Shchukin and D. Bimberg, Rev. Mod. Phys. **71**, 1125 (1999).
28. N. N. Ledentsov, V. M. Ustinov, V. A. Shchukin, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 385 (1998) [Semiconductors **32**, 343 (1998)].
29. M. Kummer, B. Vogeli, T. Meyer, and H. Kanel, Phys. Rev. Lett. **84**, 107 (2000).
30. I. G. Neizvestny, L. N. Safronov, N. L. Shwartz, *et al.*, in *Proceedings of the 8th International Symposium "Nanostructures: Physics and Technology", St. Petersburg, 2000*, p. 129.
31. Y. Hayashi, Y. Agata, T. Soga, *et al.*, Jpn. J. Appl. Phys. **37**, L1354 (1998).
32. S. I. Romanov, V. I. Mashanov, L. V. Sokolov, *et al.*, Appl. Phys. Lett. **75**, 4118 (1999).
33. S. Jin, H. Bender, L. Stalmans, *et al.*, J. Cryst. Growth **212**, 119 (2000).
34. S. I. Romanov, A. V. Dvurechenskii, V. V. Kirienko, *et al.*, in *Proceedings of NATO Advanced Research Workshop "Perspectives, Science and Technologies for Novel Silicon on Insulator Devices"*, Ed. by P. L. F. Hemment *et al.* (Kluwer, Dordrecht, 2000), p. 29.
35. A. V. Zverev, I. G. Neizvestnyĭ, N. L. Shvarts, and Z. Sh. Yanovitskaya, Izv. Akad. Nauk, Ser. Fiz. **64**, 337 (2000).
36. T. Sato, K. Mitsutake, I. Mizushima, and Y. Tsunashima, Jpn. J. Appl. Phys., Part 1 **39**, 5033 (2000).
37. P. Finnie and Y. Homma, Phys. Rev. Lett. **82**, 2737 (1999).
38. K. Thurmer, D.-J. Liu, E. D. Williams, and J. D. Weeks, Phys. Rev. Lett. **83**, 5531 (1999).
39. C. Alfonso, J. C. Heyraud, and J. J. Metois, Surf. Sci. Lett. **291**, 745 (1993).
40. Y. Homma, H. Hibino, and T. Ogino, Phys. Rev. B **58**, 13146 (1998).
41. J. M. Bermond, J. J. Metois, J. C. Heyraud, and C. Alfonco, Surf. Sci. **331–333**, 855 (1995).
42. Y. Homma, H. Hibino, Y. Kunii, and T. Ogino, Surf. Sci. **445**, 327 (2000).
43. A. Pimpinelli and J. Villain, Physica A (Amsterdam) **204**, 521 (1994).
44. A. V. Latyshev, A. B. Krasilnikov, and A. L. Aseev, Surf. Sci. **311**, 395 (1994).
45. R. L. Schwoebel and E. J. Shipsey, J. Appl. Phys. **37**, 3682 (1966).
46. I. G. Neizvestny, N. L. Shwartz, Z. Sh. Yanovitskaya, and A. V. Zverev, Thin Solid Films (in press).

Translated by A. Sidorova-Biryukova

Semiconductors, Vol. 35, No. 9, 2001, pp. 1030–1037. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1075–1083.

SEMICONDUCTOR STRUCTURES, INTERFACES, AND SURFACES

Properties of Silicon-on-Insulator Structures and Devices

V. P. Popov*, A. I. Antonova, A. A. Frantsuzov, L. N. Safronov, G. N. Feofanov, O. V. Naumova, and D. V. Kilanov

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: popov@isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—The physical grounds for making SOI structures by the DeleCut (ion irradiated deleted oxide cut) method are considered. This method is a modification of the commonly known Smart Cut® technique and aims at eliminating the disadvantages of the basic method [1]. The proposed method makes it possible to considerably lower the annealing temperature and the content of radiation defects in SOI structures. It allows the thickness of a split-off Si layer and a transition layer between the SOI layer and a buried oxide to be reduced. The method also reduces the nonuniformity in the thickness of the SOI layer and the insulator to several nanometers. By using DeleCut, new SOI structures were formed on wafers with diameters as large as 150 mm; the structures included dislocation-free SOI layers of 0.003–1.7 μm in thickness and a buried thermal SiO_2 oxide (0.05–0.5 μm). These structures have good electrical characteristics, which is supported by fabricating the submicrometer (0.2–0.5 μm) SOI-based CMOS transistors and test integrated circuits. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is known that silicon-on-insulator (SOI) structures offer noticeable advantages in the production of a wide variety of semiconductor devices and microcircuits over silicon wafers, including those with epilayers. Compared to the bulk materials, they consume less power and have a higher speed. From the beginning of the 1970s, progress in microelectronics has strictly obeyed Moore's law: the efficiency of microprocessors doubles every 18 months. The size of elements is rapidly approaching 50 nm, and the switched charge is being reduced to a mere 1000 electrons or less. In spite of the fact that some companies (Intel, NEC, Lecent Technology) have succeeded in reducing the size of the elements (to less than 50 nm) [2], the physical limits (quantum effects and uncertainty in the behavior of low currents) and the technological limits (the limited dissipated power, complexity in topology, and tunnel currents) can sufficiently hamper element miniaturization if the conventional CMOS technology based on bulk silicon is used. The technological problems, together with the rapidly growing cost of the development of the conventional technology of CMOS, favors the introduction of alternative concepts based on the use of new materials.

The use of SOI reduces the channel length of a MOS transistor to 15 nm [2], and the dual-gate design provides the possibility of reducing the channel length to 5–6 nm [2]. At present, two basic methods are used worldwide for obtaining the SOI wafers. The first method consists in the implantation of oxygen ions into a crystal (Silicon Implanted with Oxygen, SIMOX) with the subsequent synthesis of a buried oxide by annealing. The second method employs the hydrogen-based transfer of silicon with oxide (Smart Cut® SOI). It consists in the direct bonding of an oxidized hydrogen-implanted *n*-Si wafer with a supporting substrate and the subsequent, almost complete, removal of the donor wafer by shearing it off with the implanted hydrogen. The SOI wafers prepared by these two methods are commercially used despite being five–eight times more costly than silicon wafers of the same diameter. However, only these two methods satisfy the requirements of real submicrometer technology for the thickness of a split-off SOI layer less than 50 nm.

The Smart Cut® method proposed by Bruel [3] is closest to the method of obtaining SOI discussed here. The procedures in Bruel's method, after the transfer of the silicon layer with the oxide on a supporting substrate, involve heat treatment at 1100–1200°C, which improves the structural parameters due to the annealing of radiation defects in silicon and in buried SiO_2, removes the residual hydrogen, and strengthens the bonds at the interface. The final procedure consists in the precise polishing of a SOI structure which removes 0.1–0.2 μm from the rough and still damaged outer silicon layer.

The disadvantage of the method employed in [3] consists in the fact that it is time-consuming; in addition, with this method, nonuniformity in the thickness of the split-off layer is great due to the final procedure, and the thermal stability of defects in the ion-irradiated insulator is higher than in Si.

1063-7826/01/3509-1030$21.00

The use of SOI structures with a thickness of 30–50 nm for ultralarge-scale integrated circuits (ULSICs) and single-electron devices imposes stringent requirements on the perfection of the ⟨split-off Si layer⟩–⟨buried SiO_2⟩ interface from the standpoint of structural and electrical properties. It is known that the thickness of a transition layer at the Si-(thermal SiO_2) boundary can be as large as several nanometers [4]. The improvement in the properties of the Si/SiO_2 transition layer and the reduction in the thickness of this interface are of great importance for SOI structures.

Our study was aimed at developing a method for producing SOI wafers with properties of a split-off Si layer and of a buried insulator competitive with the properties of epitaxial or bulk material and of a thermal oxide. Additionally, we shall describe the technology for the fabrication of submicrometer (0.5–0.2 μm) SOI CMOS transistors for investigations of the features of the technological processes and electrical parameters of SOI.

The results were obtained by using a technique, like Smart Cut®, in which the SOI *n*-Si wafer was oxidized prior to irradiation [3]. The protective oxide layer prevented the escape of implanted hydrogen and the anneal of implantation defects; it also hindered the development of microcracks in the bulk and implantation relief on the surface in the course of irradiation. However, in contrast to the method described in [3], the layer was etched off prior to wafer bonding [1] (Fig. 1). The buried insulator was formed by thermal oxidation on a supporting wafer and was not irradiated, in contrast to the Smart Cut® method.

2. THE MECHANISMS OF BONDING, DELAMINATION, AND TRANSFER

The process of wafer bonding mainly consists of two stages:

(1) Joining the wafers with internal planar surfaces at room temperature; bonding relies on the Van der Waals forces and the dipole interaction of the radicals adsorbed on the surfaces.

(2) The heating of the joined pair at an elevated temperature (800–1200°C) in order to transform weak intermolecular bonds into strong covalent bonds.

The most important problem in bonding is the local absence of contact (macrovoids and microvoids forming due to the microparticles and surface irregularities) which reduce the bonding force between the silicon wafer and the yield of the structures [4]. The second requirement to the process of bonding is the cleanness of wafer's surfaces.

There are many methods of chemically cleaning Si wafers, the most commonly used is the RCA cleaning [4]; i.e., treatment in a peroxide–ammonium solution, the removal of natural oxide by diluted hydrofluoric acid, and treatment in a peroxide–acidic solution. We also used the RCA cleaning. After each stage of cleaning, the wafers were carefully rinsed in deionized water. Bonding occurred between the hydrophilic surfaces obtained by the treatment in peroxide–ammonium solutions of different concentration with the wetting angle for silicon and SiO_2 ranging from 0° to 10°. After hydrophilization, the wafers were dried in a centrifuge and joined in pairs in a bonding setup. The quality of bonding was checked by examining the images in near IR transmitted light using a CCD camera.

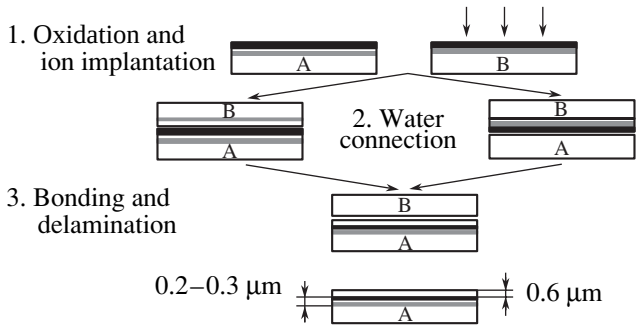


Fig. 1. Two variants of silicon-on-insulator structure fabrication: (on the left) DeleCut and (on the right) Smart Cut®.

The technology of hydrogen transfer [1, 3] is based on the effect of the formation of microcracks as large as 100 nm at a depth corresponding to the projected range R_p of hydrogen ions. The microcracks accumulate gaseous hydrogen during the subsequent heat treatment. This process results in the increase of pressure in the microcracks, in their growth, and the final separation of a silicon film from the crystal. After annealing at 400–600°C, large microcracks are formed (~1 μm) in an ion-damaged silicon layer. The characteristic feature of hydrogen is its active interaction with impurities and defects. As a result of such activity, hydrogen is captured by the stretched and weakened Si–Si bonds at the sites of the microcracks. The size of the microcracks increases in the course of hydrogen accumulation, and finally, the complete breakage of bonds and the formation of blisters with dimensions of 10–50 μm occur on the surface.

The blisters are formed during the annealing of the ion-implanted crystal if a certain hydrogen dose is exceeded. A similar surface relief can also be formed during implantation if the dose exceeds another critical value. Thus, there is a certain dose range within which no damage during irradiation occurs to a surface, and at the same time, the amount of hydrogen is sufficient for the detachment of the silicon layer during annealing. This range of doses depends on the temperature of irradiation, the energy and intensity of the ion beam, on the presence of additional coating on the silicon surface (for example, SiO_2 film), and on the starting material.

If the surface of irradiated crystal is coated with a mechanically stable film or joined to another wafer, the formation of protruding blisters will be suppressed, and

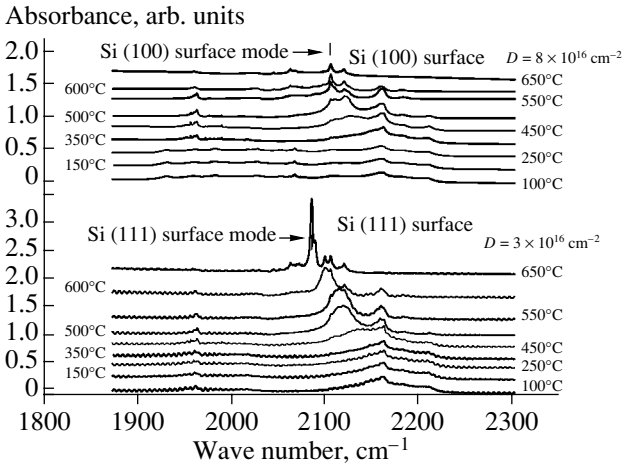


Fig. 2. The effect of annealing on the absorption spectra of silicon implanted with doses D of hydrogen equal to 3×10^{16} cm^{-2} (at the bottom) and 8×10^{16} cm^{-2} (at the top).

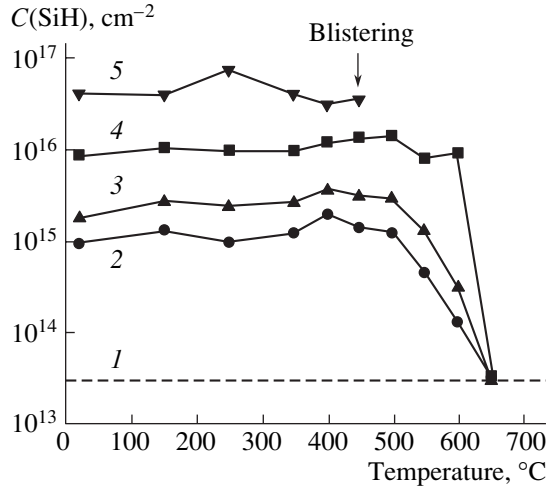


Fig. 3. Concentration of Si–H bonds in Si implanted with H as a function of annealing temperature: (*1*) represents the lowest level; other curves correspond to the implantation dose of (*2*) 3×10^{15}, (*3*) 10^{16}, (*4*) 3×10^{16}, (*5*) 8×10^{16} cm^{-2}.

the cracks will develop predominantly along the direction parallel to the surface, which will cause of the film to completely detach from the remaining crystal [3].

Let us consider the kinetics of blister formation and film detachment during subsequent annealing according to the data of IR spectroscopy. The IR absorption by Si–H bonds in silicon irradiated with hydrogen H_2^+ ions was studied. The 400-μm-thick *p*-Si(100) wafers (B concentration 10^{14} cm^{-3}) were irradiated with hydrogen ions with an energy of 112 keV (projected range R_p = 0.6 μm) and doses within the range (0.3–8) $\times 10^{16}$ cm^{-2}.

To enhance the sensitivity, the IR absorption was measured in the total-internal-reflection mode. For this purpose, prisms 40×20 mm in size were made from the wafers; the short faces of the prisms were polished at an angle of 45° to the major face. The prisms were placed into the evacuated chamber of an IFS 66-V/S spectrometer. The optical beam was directed onto the major faces of the prisms with the aid of mirrors. The multiple reflections inside the wafer allowed the light beam to pass through the hydrogen-saturated silicon layer about 100 times. The absorption spectra (Fig. 2) were recorded at room temperature with a resolution of 1 cm^{-1}. The absorption spectra of irradiated silicon contain a large number of lines in the range of 1800–2300 cm^{-1}. These lines are related to absorption by the elastic modes of the Si–H bonds of hydrogen atoms which saturate the dangling bonds of various vacancy and interstitial complexes [5]. The changes in IR spectra on heating in the range of the absorption by Si–H bonds are shown in Fig. 2. As a result of annealing silicon irradiated with large doses of hydrogen ions, a broad band with a peak near 2110 cm^{-1} appears due to the absorption by hydrogen incorporated in laminar defects. Using a well-known relation $C(\mathrm{SiH}) = 1.1 \times 10^{20}$ [cm^{-1}] $\int_\omega \alpha(\omega)/\omega d\omega$ [6], where ω is the wave number, one can determine the concentration C of Si–H bonds and, consequently, the concentration of chemically bonded hydrogen. The integrated density of Si–H bonds after irradiation as a function of the temperature of isochronous 30-min annealing is shown in Fig. 3.

The results of IR measurements showed that the amount of chemically bonded hydrogen immediately after irradiation with hydrogen amounts to about 20–50% of the total amount of the hydrogen in a dose. The fraction of bonded hydrogen has a tendency to increase with increasing dose. The bonded hydrogen exists in the form of various complexes with defects. The formation of hydrogen-containing laminar defects occurs at irradiation doses above 10^{16} cm^{-2}. The heat treatment at 200–450°C increases the intensity of the absorption band associated with the laminar hydrogen precipitates (Fig. 2); accordingly, the fraction of chemically bonded hydrogen also increases. For doses of about 10^{17} cm^{-2}, almost all the implanted hydrogen becomes chemically bonded with silicon after annealing at 300°C.

Under the chosen conditions of irradiation and annealing, blistering only occurred on the silicon wafers which were subjected to irradiation with doses exceeding 4×10^{16} cm^{-2}. Within the range of doses $(2–4) \times 10^{16}$ cm^{-2}, the line attributed to monohydrides appeared in IR spectra after high-temperature heating (450–550°C). These monohydrides covered the surface of the cracks inside the silicon wafer. The comparison of the positions of lines in the spectrum of monohydrides in wafers with (111) and (100) orientations suggests that, for the doses of irradiation used, the cracks are oriented mainly along directions parallel to the surface.

For doses lower than 2×10^{16} cm^{-2}, most of the hydrogen-containing defects become annealed at temperatures lower than 250°C. The exceptions, probably, are the VH_3 (2159 cm^{-1}) and VH_4 (2228 cm^{-1}) defects (Figs. 2, 3). At a dose of 2×10^{16} cm^{-2} and above, the local clusters of hydrogen atoms in laminar hydrogen precipitates are formed (~2110 cm^{-1}) in the vicinity of R_p. The same defects were observed after the treatment of silicon in a hydrogen plasma; they had the form of plates lying in the (111) [7] or (100) planes [8]. However, these defects were stable only at temperatures below 300–400°C. Owing to a high concentration of hydrogen-containing interstitial and vacancy complexes in the vicinity of R_p after implantation, their joint clusterization results in the formation of more stable laminar defects lying in the planes corresponding to a minimum of free energy. Some of the hydrogen-containing complexes (VH_2 and SiH_2, for example) can decompose at the boundaries of clusters with the formation of gaseous H_2 molecules inside a microcavity. This must lead (instead of to annealing) to the abrupt increase of internal stresses, which is actually observed experimentally in the form of negative annealing of defects at temperatures of 300–400°C when these complexes can migrate [9].

3. SILICON-ON-INSULATOR STRUCTURES

The SOI structures were prepared by DeleCut. Hydrogen ions were implanted through a thin layer of SiO_2 which preserved the implanted hydrogen and prevented contamination and damage to the surface during implantation. Unlike in the Smart Cut® method, this layer was removed after implantation. As a result, the dose of hydrogen atoms necessary to detach a thin silicon film was reduced to $(3–5) \times 10^{16}$ cm^{-2}. If a dose exceeded 5×10^{16} cm^{-2}, the defects were formed on the surface of a wafer immediately after irradiation, which hindered a close bonding and homogeneous transfer of a layer onto the oxidized supporting substrate from occurring. The amount of implanted hydrogen obtained at a dose of 2×10^{16} cm^{-2} was insufficient for the homogeneous splitting off of a film from the entire surface of the irradiated wafer, which is in accordance with the data of IR absorption.

The transfer was accomplished by a technique analogous to Smart Cut®. The final heat treatment of the SOI structure at 1100°C was necessary for increasing the force for bonding the silicon wafers to a value equivalent to the breaking stress in bulk silicon, and also for annealing out residual radiation defects and removing hydrogen from the split-off layer. The heat treatment lasted 0.5–1 h. The short annealing times at relatively low (for the oxide) temperatures were possible because the buried insulator was not subjected to irradiation; thus, it was not necessary to anneal the defects for long in order to recover the insulator properties. The removal of the outer damaged layer of a split-off silicon film was accomplished via oxidation.

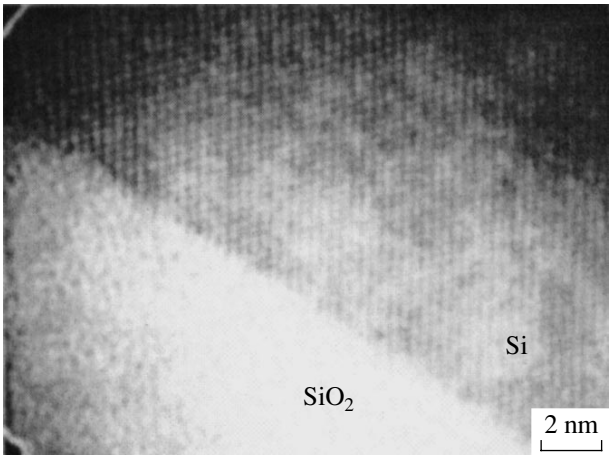


Fig. 4. Atomic-resolution image of a SOI section. The image shows the interface between a split-off silicon layer and a buried insulator. The substrate is below.

This method of forming the SOI structures substantially improves the Si/SiO_2 interface. The bonding boundary in structures formed by this method is located between the split-off Si and buried SiO_2 layers. According to the atomic-resolution microscopy data, the transition layer contains 1–2 atomic monolayers (0.3–0.6 nm). Atomically flat SOI structures (Fig. 4) were obtained by the standard techniques. According to the data obtained by transmission electron microscopy (TEM) and ellipsometry, the thickness of an interface between Si and thermal 280-nm-thick SiO_2 grown at 1000°C is equal to about 3 nm [10]. The ellipsometry yields the thickness of a transition layer at the bonding boundary Si/SiO_2 to be 0.3 nm. With the aid of layer-by-layer thermal oxidation and etching in an HF solution, single-crystal silicon films were obtained in the SOI/SiO_2 structure with a record thickness of 1–3 nm (Si) and 0.2 μm (SiO_2). The ultrathin layers are required for the realization of single-electron MOS transistors and memory elements operating at room temperature.

Thus, the main distinction of our method for obtaining MOS structures from the Smart Cut® method is the possibility of forming a buried insulator (thermal oxide) on an unirradiated substrate and to reduce the thermal and radiation impacts on the structure during the preparation of the SOI. The electron microscopy studies of the structures obtained by the proposed method show that no structural defects are observed in a split-off silicon layer. According to secondary-ion mass spectrometry (SIMS), annealing at 1100°C is sufficient to remove hydrogen from the split-off silicon layer.

The second most important advantage of DeleCut is the almost complete absence of stresses in the SOI structure (less than 10^{-5}) according to high-resolution X-ray diffraction [11].

4. ELECTRICAL PROPERTIES OF SILICON-ON-INSULATOR STRUCTURES

Since the bonding boundary in the method suggested for the preparation of SOI structures is located between the split-off silicon layer and a buried insulator, the purity of the bonded surfaces becomes very important. Fairly extensive contamination can occur in the course of bonding with resulting degradation of electrical parameters of the structure. In this section, the results of the measurements of the electrical parameters of the SOI structures are given. These parameters include: (1) the conductivity of a split-off silicon layer; (2) deep-level centers in a split-off silicon layer; (3) the states at the interfaces; and (4) the charge in the buried insulator.

The thickness of the split-off silicon layer was always equal to 0.5 μm. Some parameters of the starting materials used for forming the SOI structures are listed in Table 1. The two most important heat treatments (HTs) used in the process of forming the SOI structures are: (HT1) annealing at 450°C for the delamination of an *n*-Si wafer and strengthening the junction between the transferred layer and the substrate; (HT2) annealing at 1100°C, which is the final stage in obtaining the SOI. This treatment makes it possible to eliminate the radiation defects and hydrogen induced during the implantation and to strengthen the bonding at the boundary of the joint.

The measurements of the Hall voltage and capacitance–voltage (*C–V*) characteristics and also the capacitance- and charge-based deep-level transient spectroscopy (*C*- and *Q*-DLTS) were used. The high-frequency *C–V* characteristics of the SOI structures were measured using a mercury probe, the second contact was also located at the split-off silicon layer and had a radial symmetry relative to the probe; the same characteristics were also measured on mesa structures. In the latter case, Al-deposited contacts were used, with the second contact being deposited on the substrate. The area of the mesa structures was varied within the range of 0.5–1.5 mm^2.

It was found that a split-off silicon layer in the SOI structures has, as a rule, *n*-type conductivity, even if *p*-Si was used as the starting material [12, 13]. The data obtained by the *C–V* method and from the Hall effect measurements are listed in Table 2. When the starting material with *n*-type conductivity was used, the concentration of electrons in the split-off layer of SOI also increased (Table 2). It was only possible to study the spatial distribution of charge carriers in the split-off silicon layer in a heavily doped layer [13]. It was found that the concentration of charge carriers in a silicon film increases toward the surface and the Si/SiO_2 interface. The DLTS measurements showed that the donors forming in a split-off layer in the process of SOI preparation had shallow levels. The experimental results suggest that hydrogen is involved in the formation of donor centers stable up to high temperatures (1100°C). The high temperatures of donor annealing are apparently related to the participation of structural defects in the formation of donors (for example, these defects may be small

Table 1. The basic parameters of a starting material

Starting Si	Type of conductivity	Doping, cm^{-3}	Oxygen concentration, cm^{-3}	Mobility, $cm^2\ V^{-1}\ s^{-1}$	Orientation
Cz–Si-1	*n*-type	4×10^{14}	6×10^{17}	1100	(100)
Cz–Si-2	*p*-type	3×10^{14}	7×10^{17}	320	(100)
Cz–Si-3	*p*-type	2×10^{15}	7×10^{17}	350	(100)
Cz–Si-4	*p*-type	2×10^{18}	8×10^{17}	106	(111)

Table 2. Combination of wafers used for making SOI structures, duration of annealing at 450°C (HT1), and the conductivity of the silicon split-off layer after high-temperature annealing (HT2, 1100°C)

Combination of wafers	Designation	Duration of HT1, h	Type of conductivity after HT1	Conductivity after HT2
Cz–Si-3/4000A SiO_2/Cz–Si-3	SOI-1	0.5	*n*-type	*n*-type
Cz–Si-4/4000A SiO_2/Cz–Si-3	SOI-2	0.5	*n*-type	*n*-type, 2.5×10^{16} cm^{-3}
Cz–Si-1/4000A SiO_2/Cz–Si-1	SOI-3	1	*n*-type	*n*-type
Cz–Si-2/2800A SiO_2/Cz–Si-2	SOI-4	2	*n*-type	*n*-type
Cz–Si-3/2800A SiO_2/Cz–Si-3	SOI-5	2	*n*-type	*n*-type, 3×10^{15} cm^{-3}
Cz–Si-1/4000A SiO_2/Cz–Si-1	SOI-6	2	*n*-type	*n*-type, 5×10^{15} cm^{-3}
Cz–Si-3/4000A SiO_2/Cz–Si-3	SOI-7	3	*n*-type	*p*-type, $\sim 2 \times 10^{15}$ cm^{-3}
Cz–Si-4/4000A SiO_2/Cz–Si-3	SOI-8	3	*p*-type	*p*-type, 2×10^{18} cm^{-3}

dislocation loops or dangling bonds at the interface). The stability of a given center is most likely to be determined by the stability of structural defects.

It should be noted that the formation of donor centers in a split-off Si layer occurs not only in SOI structures obtained by DeleCut. Conversion of the conductivity from *p*- to *n*-type also takes place in the structures obtained by Smart Cut®. Furthermore, by varying the conditions of the SOI formation, a procedure can be chosen, which ensures that the conductivity in a split-off layer of SOI corresponds to the conductivity of the starting material [13].

According to the measurements of the Hall voltage, the charge carrier mobility in a split-off layer formed from a lightly doped starting material (SOI-1, 3–6) is 400–550 $cm^2 V^{-1} s^{-1}$. The mobility of charge carriers in the case of SOI-8 is 80 $cm^2 V^{-1} s^{-1}$.

The DLTS study of deep-level centers in a split-off silicon layer shows [14] that, in SOI structures, there exist deep levels with the energy $E_c - 0.39$ eV and the cross section $\sigma = 10^{-15}$ cm^2 and with $E_c - 0.58$ eV and $\sigma = 4 \times 10^{-14}$ cm^2. Both centers were found to be localized in a layer extending from the surface to a depth of 0.2 μm; their concentrations decreased with depth from the value of $(2–5) \times 10^{15}$ cm^{-3} at the surface [14]. A comparison of levels observed in the band gap of the SOI film in this study with the levels introduced by the extended defects showed that they were close to each other in their energy positions. It should be mentioned that, in SOI structures studied by DLTS, a split-off silicon layer was not thinned by polishing or oxidation. Such treatment is usually used to get rid of residual defects induced by hydrogen implantation and retained during annealing (most of the implantation-induced structural defects reside on the surface of SOI structures). In our opinion, the location of levels in the relatively thin surface layer of silicon SOI film and the descending profile of the distribution of deep-level centers suggest that the levels are related to residual post-implantation defects.

Deep-level centers are absent in the rest of the split-off layer ~0.3 μm in thickness. This unambiguously shows that the bonding of wafers at the ⟨split-off layer⟩/⟨buried insulator⟩ interface is not accompanied by the deterioration of the electrical properties compared to the Smart Cut® method. The removal of the surface layer about 0.2 μm in thickness ensures that the split-off layer does not contain any deep-level centers.

Another, no less important, feature of the SOI structure concerns the traps at the interface. The *Q*-DLTS which uses relatively low frequencies (1 kHz) has some advantages over the classical high-frequency DLTS, and allows interfaces in SOI structures to be studied [15]. The use of mesa structures with the same type of conductivity in a split-off layer and in the substrate provides the possibility of measuring independently the distribution of traps at the bonding boundary and at the bottom boundary between the substrate and the buried (thermal) oxide. The distribution of traps in the band gap at the substrate/⟨buried oxide⟩ and ⟨split-off layer⟩/⟨buried oxide⟩ interfaces is illustrated in Fig. 5 [16]. The distribution was obtained from studying the *n*-Si/SiO_2/*n*-Si and *p*-Si/SiO_2/*p*-Si structures, which made it possible to obtain information about the entire band gap in silicon. The results shown in Fig. 5 correspond to the *n*-Si/SiO_2/*n*-Si structures, for which only the upper half of the band gap of silicon was tested. It can be seen that the distribution of states over the band gap at the substrate/⟨thermal oxide⟩ interface in SOI structures is continuous and close to that occurring at the Si/⟨thermal SiO_2⟩ interface in conventional MIS structures [4]. The distribution of states in the upper half of the band gap for a bonding Si/SiO_2 interface is characterized by a relatively narrow band of states within $E_c = (0.17–0.36)$ eV. This change in the trap spectra at the bonded interface is probably not connected with the passivation by hydrogen because hydrogen is present at both boundaries almost in equal amounts during the preparation of the SOI [13]. From the curves in Fig. 5 it follows that the trap densities are actually the same. However, it is obvious that the inte-

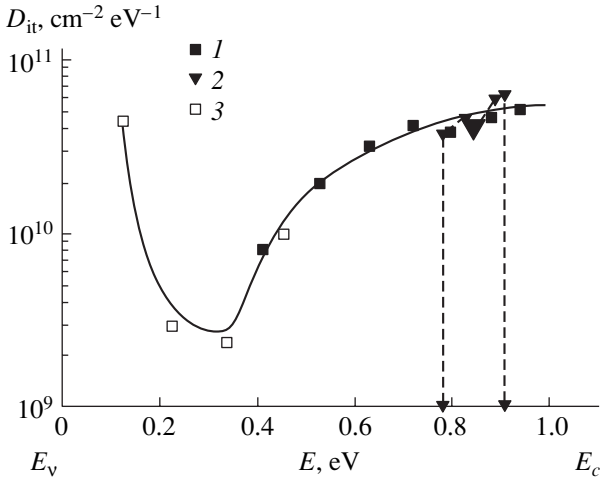


Fig. 5. Distribution of states at the substrate/⟨thermal oxide⟩ interface and at the bonded Si/SiO_2 interface: (*1*) SOI-5, substrate/SiO_2; (*2*) SOI-6, Si/SiO_2; and (*3*) SOI-8, substrate/SiO_2.

Table 3. The magnitude of charge in the buried insulator for SOI structures obtained by various methods

Type of SOI	Q_f, cm^{-2}	Q_{ss}, cm^{-2}
DeleCut	$(2–4) \times 10^{11}$	$(1.2–5) \times 10^{11}$
Smart Cut	$\sim 4 \times 10^{11}$*	3.5×10^{10}*
SIMOX	$(1–3) \times 10^{11}$	$(1–3) \times 10^{11}$

Notes: * The values are given according to data obtained for a single structure without statistics.

grated trap density at a bonding boundary is lower than at the substrate/⟨thermal oxide⟩ interface.

The values of a fixed charge Q_f and the surface density states Q_{ss} at the ⟨silicon split-off layer⟩/⟨buried insulator⟩ interface obtained from the *C–V* characteristics are given in Table 3. The same parameters are also given for SOI structures formed by Smart Cut® (SOITEC) and SIMOX methods. As can be seen from the comparison, the magnitude of the fixed charge is almost the same for all structures and amounts to $(1–3) \times 10^{11}$ cm^{-2}. A somewhat different situation exists for the density of states at the interfaces. The magnitude of Q_{ss} in DeleCut structures is close to that observed in SIMOX structures. The SOI structures manufactured by SOITEC have a lower surface density of states.

The breakdown voltages of the insulator in the DeleCut method are higher (≥100 V) than in the Smart Cut® method (~60 V, both magnitudes are given for SiO_2 of 200 nm in thickness). The higher breakdown voltages are explained by the fact that, in the method under consideration, the properties of SiO_2 were not affected during irradiation.

Table 4. Design and electrical parameters of test structures

Design parameters		Value
Thickness of silicon base layer	d_{Si}	100 nm
Thickness of front oxide	d_{ox1}	12 nm
Thickness of buried oxide layer	d_{ox2}	280 nm
Length of gate	L	20, 0.5, 0.3 μm
Width of gate	W	15–50 μm
Electrical parameters		**Value**
n-channel	Threshold V_{T1}	0.6 ± 0.05 V
	μ_n at L = 20 μm	350 cm^2 V^{-1} s^{-1}
	L = 0.5 μm	60 cm^2 V^{-1} s^{-1}
	L = 0.3 μm	45 cm^2 V^{-1} s^{-1}
	S	100 mV/decade
	Threshold of rear transistor V_{T2}	9 V
p-channel	Threshold V_{T1}	–1.3 V
	μ_p at L = 20 μm	125 cm^2 V^{-1} s^{-1}
	L = 0.5 μm	45 cm^2 V^{-1} s^{-1}
	L = 0.3 μm	35 cm^2 V^{-1} s^{-1}
	S	120 mV/decade
	Threshold V_{T2}	–12 V
Density of surface states Dit_2		$(1–3) \times 10^{11}$ cm^2 V^{-1} s^{-1}

5. SOI CMOS TRANSISTORS

The DeleCut wafers were tested by preparing the device structures. For this purpose a test-crystal topology was worked out. The test crystal consisted of 26 types of *n*- and *p*-channel transistors both with linear and annular geometry. Three variations of the channel length were chosen: 20, 0.5, and 0.3 μm. The implantation dose into the base layer was also varied.

For obtaining short-channel transistors, masks were made which ensured a channel length of about 0.5 and 0.3 μm after the photolithograpic procedure on polysilicon. The image of the polysilicon gate obtained in the scanning electron microscope is shown in Fig. 6. The gate has a T-shaped form. The region of the p^+ contact to the base region of an *n*-channel transistor is located to the left of the vertical bridge. The bridge width is 3 μm. In order to choose the doping level of the base layer, we numerically simulated the short-channel transistors to be designed. A program for solving the two-dimensional Poisson equation was written. The program finds the value of potential φ at all points of a split-off layer between the source and drain. Next, the concentration of free charge carriers and the conductivity of this layer in low-current approximation were calculated. Based on the results of calculations, the doping level of boron in the *n*-channel transistor and of phosphorus in the *p*-channel was chosen so that the conductivity of the layer did not exceed 2×10^{-11} S μm^{-1} in the absence of a voltage at the gate and at a channel length of 0.3 μm. This value corresponds to the leakage current not exceeding 4×10^{-10} A at a channel width of 20 μm, which satisfies well the requirements for the allowable leakage of a transistor in the "off" state.

For obtaining a split-off silicon layer with a width of 100 nm on the SOI wafer, a thermal oxide was repeatedly grown and was subsequently removed. The width of the remaining silicon layer was measured using an ellipsometer. The design parameters and the measured electrical characteristics of the transistors are given in Table 4. The following parameters of the front transistor were determined: the threshold voltage V_T, mobility μ, and the slope of subthreshold characteristic S. The density of surface states Dit_2 at the boundary of a split-off silicon layer (base layer) with a buried oxide was determined from the characteristics of a rear transistor.

As was expected, the short-channel effects are quite pronounced in the drain characteristics. The current does not saturate at $V_D > V_G - V_T$ and continues to increase superlinearly in this portion of the characteristic. The magnitude of the maximum current is 130 μA/μm for an *n*-channel transistor and 70 μA/μm for a *p*-channel transistor. The conductivity of the channel at small V_D is 0.5 S/cm for an *n*-channel transistor and 0.2 S/cm for a *p*-channel transistor. The smaller parameters for the *p*-channel transistor are explained by the lower mobility and a higher threshold voltage V_T.

The most unexpected result is the lowering of the effective mobility of an electron and holes with diminishing channel length. The decrease in μ_{eff} from 550 to

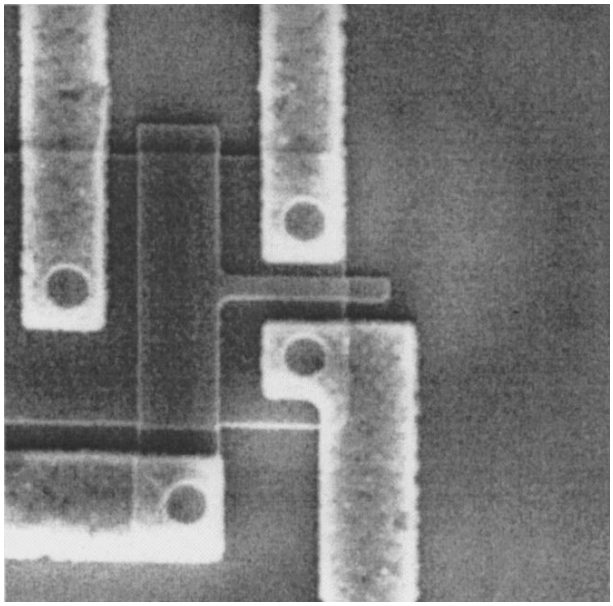


Fig. 6. SEM image of SOI MOS transistor. The image area is 20×20 μm^2.

170 cm^2 V^{-1} s^{-1} with a change in the channel length from 20 to 0.3 μm was also observed in a test wafer which was used for the preparation of bulk transistors. The mobility in bulk material decreases by three times, whereas in the SOI wafers, it decreases by six times. The distinction can be caused by different conditions for diffusion and annealing of defects in SOI and bulk Si since, for the SOI-layer thickness of 0.4 μm and the channel length of 20 μm, no difference in mobilities was observed. Some residual defects, which diffuse during annealing from the drain–source regions to the channel of a transistor, may cause the mobility to decrease.

6. CONCLUSION

The advantages of the DeleCut technology over Smart Cut® are due to the use of the thermal oxide on an unirradiated wafer as a buried insulator. This makes it possible to avoid radiation damage in the insulator. Moreover, the bonding Si/SiO_2 interface between the silicon split-off layer and the buried insulator has virtually no transition layer which enables silicon layers as thin as 3–5 nm to be obtained. The major advantage of DeleCut technology is the high electrical quality of the bonding ⟨transferred silicon layer⟩/⟨buried thermal SiO_2⟩ interface. In particular, the distribution of states in the upper half of the band gap for this interface is characterized by a relatively narrow band of states within the range of E_c – (0.17–0.36) eV and with the total concentration of states lower than 10^{11} cm^{-2}. The inner ⟨buried thermal SiO_2⟩/⟨Si substrate⟩ interface in a SOI structure is characterized by a continuous distribution of states. Two centers are observed in a silicon split-off layer with levels at E_c – 0.39 eV ($\sigma = 10^{-15}$ cm^2) and at E_c – 0.58 eV ($\sigma = 4 \times 10^{-14}$ cm^2) which are located within a subsurface layer 0.2 μm thick and are possibly caused by the residual defects. The removal of this layer by thermal oxidation after the preparation of the silicon-on-insulator structure allows an electrically perfect silicon layer with a thickness of 1–0.003 μm without deep-level centers to be obtained.

The fabricated submicrometer *n*- and *p*-channel MOS transistors with a channel length less than 0.5 μm in a silicon-on-insulator layer of 0.1 μm in thickness have the channel current 130 μA/μm and 70 μA/μm for *n*- and *p*-channel transistors, respectively, for $V_G = V_D = 3$ V. For small V_D, the channel conductivities are equal to 0.5 and 0.2 S/cm for the *n*- and *p*-channel transistors, respectively. The leakage currents do not exceed 10^{-9} A for a channel length of 20 μm. These data confirm the promising aspects for application of the DeleCut technology in the production of SOI-based ULSICs with low power consumption.

REFERENCES

1. V. P. Popov, A. I. Antonova, L. V. Mironova, and V. F. Stas', RF Patent No. 99120527/28(021735) (1999).
2. R. Chau, J. Kavalieros, R. Schenker, *et al.*, in *Abstracts of the International Electronic Devices Meeting, San Francisco, 2000.*
3. M. Bruel, Electron. Lett. **31**, 1201 (1995).
4. Q.-Y. Tong and U. Goesele, in *Wafer Bonding* (Wiley, New York, 1999), p. 52.
5. V. S. Vavilov, V. F. Kiselev, and B. N. Mukashev, in *Defects in the Bulk and at the Surface of Silicon* (Nauka, Moscow, 1990), p. 216.
6. L. S. Sidhu, T. Kosteski, and S. Zukotynski, J. Appl. Phys. **85**, 2574 (1999).
7. S. Muto, S. Takeda, and M. Hirata, Mater. Sci. Forum **143–147**, 897 (1994).
8. A. G. Ulyashin, A. I. Ivanov, R. Job, *et al.*, Mater. Sci. Eng., B **58**, 91 (1999).
9. V. P. Popov, A. K. Gutakovsky, I. V. Antonova, *et al.*, Mater. Res. Soc. Symp. Proc. **536**, 109 (1999).
10. V. P. Popov, I. V. Antonova, V. F. Stas, *et al.*, Mater. Sci. Eng., B **73**, 82 (2000).
11. V. P. Popov, I. V. Antonova, J. Bak-Misiuk, and J. Domagala, in *Proceedings of the Spring E-MRS Meeting, Strasbourg, 2000*, Mater. Sci. Eng. B (in press).
12. I. V. Antonova, V. P. Popov, V. F. Stas, *et al.*, Microelectron. Eng. **48**, 383 (1999).
13. I. V. Antonova, V. F. Stas', V. P. Popov, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 1095 (2000) [Semiconductors **34**, 1054 (2000)].
14. I. V. Antonova, J. Stano, D. V. Nikolaev, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **35** (8), 948 (2001) [Semiconductors **35**, 912 (2001)].
15. J. W. Farmer, C. D. Lamp, and J. M. Meese, Appl. Phys. Lett. **41**, 1063 (1982).
16. T. Katsube, K. Kakimoto, and T. Ikoma, J. Appl. Phys. **52**, 3504 (1981).

Translated by A. Zalesskiĭ

Semiconductors, Vol. 35, No. 9, 2001, pp. 1038–1044. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1084–1091.

SEMICONDUCTOR STRUCTURES, INTERFACES, AND SURFACES

In situ Study of the Interaction of Oxygen with the Si(111) Surface by Ultrahigh-Vacuum Reflection Electron Microscopy

S. S. Kosolobov, A. L. Aseev, and A. V. Latyshev*

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: latyshev@thermo.isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—Reactions of gases with an atomically clean silicon surface were examined by ultrahigh-vacuum reflection electron microscopy. The initial stages of interaction of oxygen with the Si(111) surface were studied in the temperature range from 500 to 900°C. Motion of monatomic steps to the upper terraces during etching by oxygen at high temperatures was visualized. The conditions for the formation of surface vacancies were determined. The dependence of the step velocity on the width of adjacent terraces was measured. Oscillations of the intensity of the electron beam reflected specularly from the Si surface were observed during the etching of silicon by molecular oxygen, which proceeded by a two-dimensional-island mechanism. The activation energy for the diffusion of surface vacancies, which are formed owing to the interaction of oxygen with silicon, was estimated to be 1.35 ± 0.15 eV. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Atomic processes on crystal surfaces can be described in the context of the behavior of such elementary defects as adsorbed atoms (adatoms) and surface vacancies, which are always present on a surface at temperatures above absolute zero. The migration and interaction of adatoms and vacancies with each other, monatomic steps, and superstructural domains determine structural processes on a real crystal surface. Structural investigations based on *in situ* observations of surface processes with a resolution on the order of one monolayer open up new possibilities for obtaining reliable information about the properties of the surface. Of interest is also the *in situ* study of atomic processes proceeding during the interaction of an atomically clean silicon surface with a gas atmosphere, in particular, the interaction of oxygen with a reconstructed silicon surface. Such investigations are of practical importance, because the oxidation of the silicon surface is a basic stage of the insulator-layer formation in present-day micro- and nanodevice technology.

The interaction of the silicon surface with oxygen was examined with diffraction- and laser-related ellipsometric and other methods. However, all these methods provide integral data on the surface structure, which are averaged over a large area. Silicon–(silicon dioxide) interfaces can be investigated by electron-microscopic methods, which possess a fairly high spatial resolution. However, the *in situ* electron-microscopy study is of limited usefulness because of the necessity to obtain and maintain an atomically clean Si surface during the experiment, which requires ultrahigh-vacuum (UHV) conditions in the column of an electron microscope. Interesting results were obtained in UHV electron-microscopy studies of transformations of a clean Si surface during its oxidation and the initial stage of etching by oxygen [1–4]. Scanning tunneling microscopy (STM) is an efficient method for examining the atomic structure of a surface interacting with a gas ambience [5, 6]. However, STM observations of the silicon surface in an oxygen atmosphere are hampered by the formation of a silicon-dioxide layer, which decreases the tunnel current. This study was devoted to the development of a controlled gas-inlet system for admitting a gas into the column of an ultrahigh-vacuum electron microscope and *in situ* investigations of the effect of oxygen adsorption on the structural transformations proceeding on the Si(111) surface at various temperatures and oxygen pressures. Much attention was concentrated on the analysis of transformations of the Si surface at rather low oxygen pressures when etching rather than oxidation of the Si surface in an oxygen atmosphere took place.

EXPERIMENTAL

Structural reconstructions of the silicon surface were studied by ultrahigh-vacuum reflection electron microscopy (UHV REM), which was described in detail elsewhere [7]. The unique design of the differential cryogetter pumping system provides a residual pressure of $\approx 10^{-9}$ Torr, which makes it possible to prepare and maintain an atomically clean silicon surface during the experiment [8]. The UHV REM allows for

1063-7826/01/3509-1038$21.00

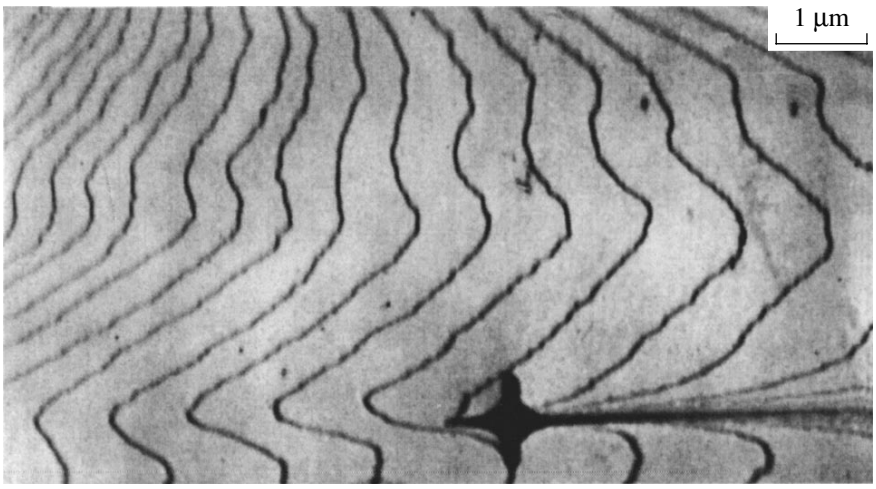


Fig. 1. REM image of the atomically clean Si(111) surface annealed at a high temperature. The surface contains a system of monatomic steps. An immobile particle acting as a step-pinning center (bottom of the photograph) serves as a reference. The terrace height increases from left to right.

the visualization of such elements of the surface relief as monatomic steps, which are 0.31 nm high on the Si(111) surface even at temperatures above 1200°C. After UHV REV studies, some samples were analyzed by atomic force microscopy (AFM) at atmospheric pressure (SOLVER P-47H NT-MDT).

Samples with dimensions of $8 \times 1 \times 0.3$ mm were cut from a standard Si(111) wafer misoriented by less than 1°. After conventional chemical cleaning, the sample was mounted on a holder with tantalum clamps. The sample was heated by a direct or alternating current. The holder with the attached sample was placed into the electron-microscope column. Then, the sample was heated to 1200°C in an ultrahigh vacuum by the alternating current. The sample temperature was calibrated against the current with the use of an optical pyrometer (at high temperatures) and a thermocouple (at low temperatures). The surface was considered to be atomically clean if its microdiffraction pattern contained no additional reflections, a reversible $(1 \times 1) \longrightarrow (7 \times 7)$ superstructural transition took place, and no centers pinning monatomic steps during sublimation were observed. REM images were taken on photographic plates or filmed with a special-purpose Gatan TV camera. The intensities of the diffraction reflections were measured with a semiconductor detector built into the microscope screen.

To analyze the reactions of gases with the crystal surface under study, we devised a gas-inlet system and connected it to the microscope column. This system consists of a high-pressure oxygen bottle connected to a ballast tank. The tank is linked to one of the channels of an electronic two-channel leak admitting the gas into a vacuum chamber of the differential cryogetter system through a built-in connecting pipe fitted with a diffuser made of a stainless-steel wire. Another ballast tank connected to an adsorption pump is coupled to the second channel of the leak. Closing the first channel of the two-channel leak while simultaneously opening the second channel during the gas admission, we minimized the additional exposure time of the sample to the gas atmosphere owing to the evacuation of the residual gas from connecting pipes through the second channel. Thus, the devised gas-inlet system equipped with the two-channel leak and the electronic control system allows us to let in the gas over a wide range of pressures with a high accuracy and a specified exposure time.

RESULTS AND DISCUSSION

Figure 1 shows a representative REM image of the Si(111) surface cleaned at a high temperature in the UHV chamber of the electron microscope. The twisting dark lines are images of monatomic steps 0.31 nm in height. The monatomic-step contrast is a superposition of diffraction and phase contrasts, which are caused by deformation fields near a monatomic step and a phase shift due to the reflection of the electron beam from terraces adjacent to the step. REM images have different magnification scales along and perpendicular to the direction of incidence of the electron beam due to the small angle of incidence of the electron beam on the surface studied. This results in a uniaxial contraction of the REM images. In this study, the images are contracted in the vertical direction by a factor of about 40. During sublimation at temperatures above 850°C, movement of the monatomic steps to the upper terraces was observed. This is attributed to the fact that, during sublimation, adatoms are removed from terraces and the steps act as sources of adatoms. The positions of the monatomic steps on the Si surface were related to positions of immobile particles, most probably, of silicon carbide or a refractory metal, which were not removed from the surface by thermal annealing. Such particles

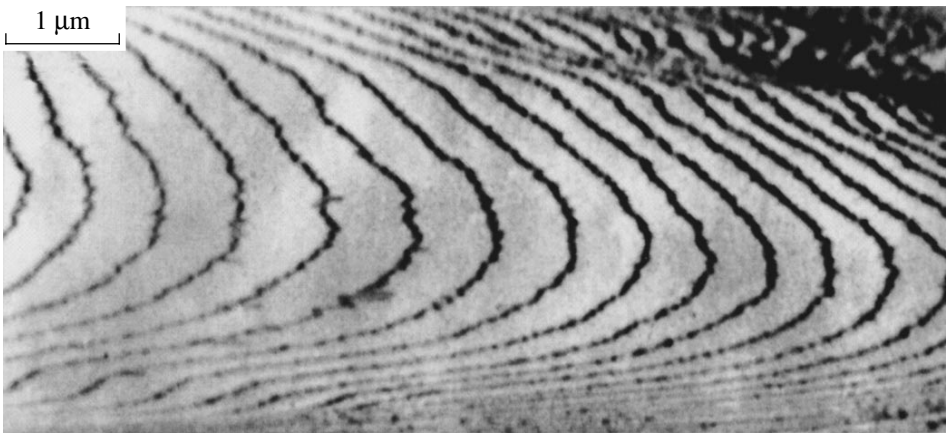


Fig. 2. Representative image of the stepped Si(111) surface exposed to the oxygen atmosphere at a pressure below P_{crit}. Monatomic steps moved from left to right.

are visible on the REM images owing to the shadow contrast (the dark spot at the bottom of Fig. 1). These particles do not migrate over the crystal surface at temperatures of up to the melting point. Their typical density is several particles per several square millimeters.

The step velocity depends on temperature and the separation between monatomic steps [9, 10]. A system of regular monatomic steps or a system of step bunches, which consists of bundles of monatomic steps and surface areas with a low step density, formed on the surface in certain temperature ranges, depending on the direction of the electric current heating the sample [11, 12]. Using the effect of a kinetic unstable system of monatomic steps induced by the electromigration of Si adatoms, we controlled the interstep distance on local surface areas from several nanometers to dozens of micrometers in size. The distribution of the steps or step bunches formed at sublimation temperatures was unchanged on fast cooling of the crystal to temperatures below 800°C and did not alter during the characteristic observation time (several hours).

After oxygen was admitted into the UHV chamber of the electron microscope, no appreciable change in surface morphology was observed at oxygen pressures below P_{crit} (Fig. 2). At rather low pressures, oxygen had no noticeable effect on intensity of main and superstructural reflections in diffraction patterns. The REM contrast did not change, but the monatomic steps moved to higher terraces. As the oxygen pressure increased (remaining below P_{crit}), intensities of the superstructural reflections decreased slightly.

On exposure to molecular oxygen, the monatomic steps move apparently due to the etching of the silicon surface by oxygen at high temperatures [13]. It is well known that, at sufficiently high temperatures, an oxygen molecule on the silicon surface dissociates into two atoms O(ads), which remain in the adsorbed state on the silicon surface [14]. In turn, adsorbed oxygen interacts with silicon to form volatile silicon monoxide (SiO):

$$\mathrm{Si} + \mathrm{O}(ads) \longrightarrow \mathrm{SiO}(gas)\uparrow.$$

This reaction describes etching by oxygen at high temperatures, resulting in the removal of Si atoms from the surface. This phenomenon may explain the directional motion of the monatomic steps observed with the reflection electron microscope.

Indeed, if atoms are sequentially removed from a step, the step will move. On REM images, this process will appear as the motion of the monatomic step to upper terraces. Let us consider several possible mechanisms of the displacement of the monatomic steps during etching by oxygen. First, oxygen atoms may be adsorbed onto the silicon surface, migrate over the surface toward a monatomic step, and react with Si atoms on the step to form silicon monoxide. According to another mechanism, atomic oxygen reacts with Si atoms adsorbed on terraces, thus decreasing their concentration. The decrease in concentration of Si adatoms is compensated for by the generation of adatoms by the monatomic steps, which also leads to the upward motion of the steps. Finally, surface vacancies may form owing to the removal of Si atoms from terraces (as a result of the reaction with the formation of silicon monoxide). The flow of the vacancies toward a monatomic step will lead to the removal of atoms from the step owing to the annihilation of the vacancies and the atoms. We should take into account that the recombination of surface vacancies and Si adatoms on terraces may also result in the movement of monatomic steps to upper terraces. Consequently, analyzing the interaction of oxygen with the Si surface, we should take into account several atomic-scale mechanisms for the step motion, depending on experimental conditions.

As demonstrated above, ensembles of point defects and vacancies, which migrate over the terrace surfaces and interact both with monatomic steps and with each

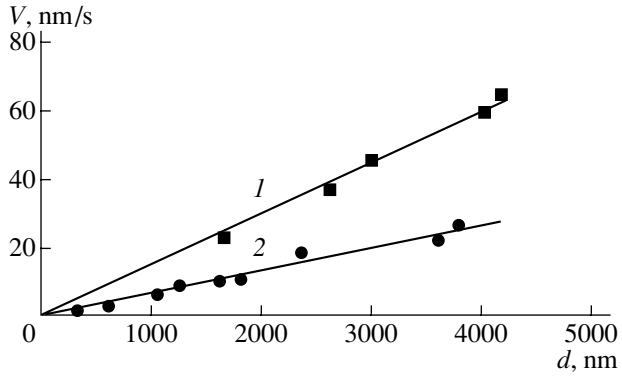


Fig. 3. Dependences of the velocity of a monatomic step on the distance to the neighboring step measured during exposure to the oxygen atmosphere at a pressure below P_{crit} and temperatures of (*1*) 870 and (*2*) 780°C.

other, may form during etching by oxygen at high temperatures. Let us consider specific features of the step-flow mechanism of etching of the Si(111) surface. This mechanism implies that there is no interaction between vacancies that result in their coalescence, i.e., in the formation of negative two-dimensional (2D) islands. According to the classical theory [15], the velocity of a monatomic step should depend on the width of adjacent terraces, all other factors being the same. This dependence should be linear if the diffusion length of surface vacancies exceeds the interstep spacing. If the terrace width is considerably greater than the vacancy-diffusion length, the velocity of the monatomic steps is independent of the width of adjacent terraces. Figure 3 shows the dependences of the velocity of a monatomic step on the width of adjacent terraces for etching by oxygen at two temperatures. The wider the adjacent terraces, the higher the velocity of the monatomic steps. The linear dependence of the step velocity on the terrace width indicates that, at a given temperature and oxygen pressure, the length of migration of surface vacancies is greater than or equal to the interstep distance. The step velocity increases as temperature rises, since the vacancy mobility increases with temperature.

At an oxygen pressure above P_{crit}, an additional REM contrast was observed on wide terraces between the monatomic steps. This contrast was identified as negative islands of monatomic depth (Fig. 4). This finding is consistent with the results from Shimuzu *et al.* [16]. As in homoepitaxial growth [17], no negative islands were formed on terrace areas adjacent to the steps. Moreover, the negative islands were absent on surface areas with a high density of monatomic steps. Further exposure of the Si surface to the oxygen atmosphere led to an increase in island size. After a certain time, the growing negative islands came into contact with the monatomic steps. As a result, the islands disappeared and the step shape changed. The absence of an REM contrast at points of contact between the islands and the steps suggests that the depth of the negative islands is equal to the height of the monatomic steps, i.e., one interplane distance. As the steps moved further, they gradually recovered their original shape. It should be noted that, as the oxygen pressure exceeded a certain value, oxidation of the Si surface began, which manifested itself in the appearance of homogeneous ripples on the REM images and the termination of the step motion.

The formation of the negative 2D islands on terraces suggests that the etching of the Si surface proceeds through the generation of surface vacancies, which coalesce to produce the negative 2D islands. The formation of similar negative islands on wide terraces was observed previously during high-temperature (>1200°C) sublimation [18] or a sharp change of the sublimation temperature, resulting in the decrease in diffusion length of adatoms [11]. However, in the case of etching the Si surface with oxygen at high temperatures, these islands form at lower temperatures (500–900°C).

Based on these results, we may conclude that the etching of the Si surface proceeds either by the step-

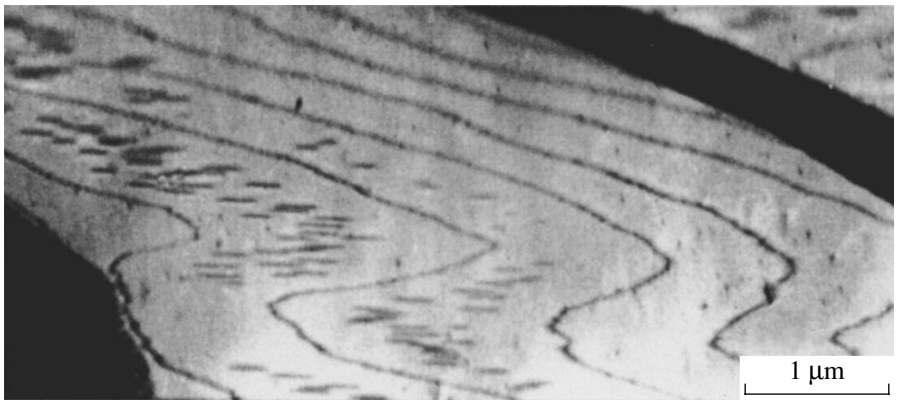


Fig. 4. Representative REM image of the Si surface containing negative 2D islands, which were formed during etching by oxygen at a pressure above P_{crit}.

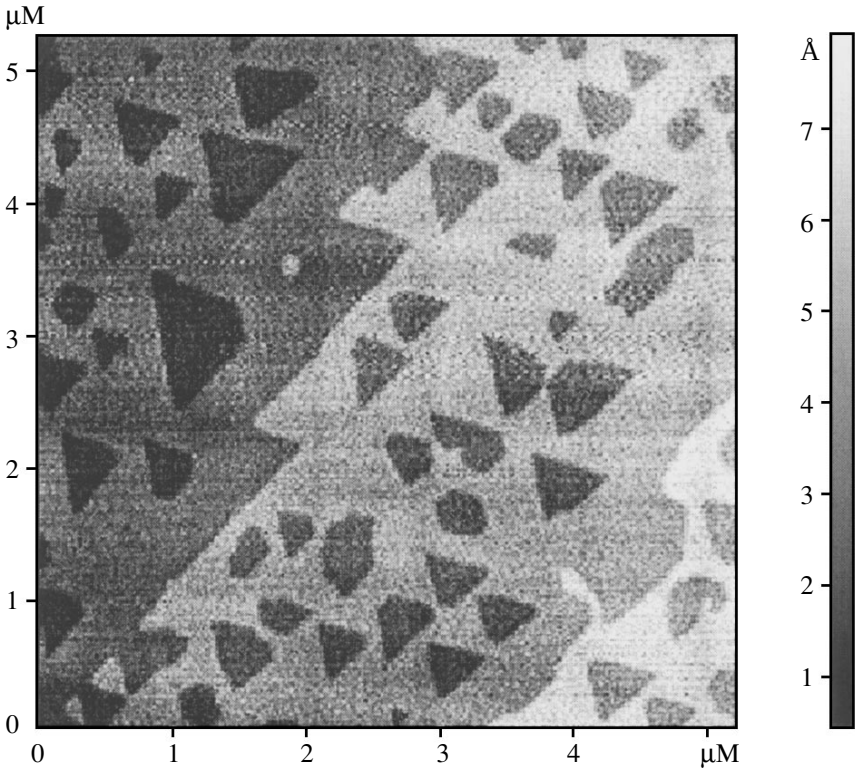


Fig. 5. An AFM image of the atomically clean Si(111) surface after exposure to the oxygen atmosphere. The oxygen pressure and the crystal temperature were chosen so that etching by oxygen proceeded via the formation of 2D islands.

flow mechanism or via nucleation of the 2D negative islands, depending on temperature and oxygen pressure. Displacement of the monatomic steps on exposure to the oxygen atmosphere is characteristic of the step-flow mechanism, which is effective at high temperatures or low oxygen pressures owing to the generation and migration of surface vacancies and their interaction with the steps. The 2D-island mechanism of etching operates at lower temperatures or higher oxygen pressures and is characterized by the formation of the 2D negative islands. Depending on the ratio of temperature to oxygen pressure, etching by the first or the second mechanism was observed, and sometimes these mechanisms operated concurrently. For example, both the formation of large islands and the step motion were observed at an oxygen pressure of $\approx 10^{-7}$ Torr and temperatures above 750°C, whereas only the step motion occurred at temperatures exceeding 860°C, and the origination of small 2D islands with high density mainly took place at temperatures below 640°C. It should be noted that, at a given temperature and oxygen pressure, etching of the sample surface proceeds by the step-flow mechanism or through nucleation of the 2D islands, depending on the step density, i.e., the terrace width.

It is noteworthy that, on exposure of the Si(111) surface to the oxygen atmosphere, the shape of the monatomic steps on a microscopic scale changed from smooth to staggered (Figs. 1 and 2). Since microkinks on the steps are observed even during etching by the step-flow mechanism, they are apparently formed due to the interaction of vacancies with atoms on the steps rather than the interaction of the moving monatomic steps with the negative islands, as discussed above. A thorough analysis of the REM images taken from the same area at different azimuthal angles shows that the microkinks consist of linear portions oriented parallel to $\langle 110 \rangle$ crystallographic directions. The dependence of the length of the linear portions on the time of exposure to the oxygen atmosphere was measured at various temperatures and oxygen pressures. Analyzing the standard deviation of the step shape from the original shape, we find that, at the initial stages of the interaction of oxygen with silicon, the roughness of the steps initially increases with time and then tends to a constant value. The final step roughness and, consequently, the linear dimension of the step kinks increase as crystal temperature or oxygen pressure rises. However, the step roughness is found to decrease sharply near the temperature of the superstructural transition. Consequently, the superstructural reconstruction induces a "faceting" of the monatomic steps during etching of the Si surface at a low oxygen pressure. The appearance of similar kinks on monatomic steps, but with a lower amplitude, was observed on the Si(111) surface undergoing the $(1 \times 1) \longrightarrow (7 \times 7)$ reconstruction, which took place as

temperature decreased to <830°C [19, 20]. The kink formation was attributed to the orientation of superstructural domains and the presence of antiphase boundaries between them.

The island shape was not reliably established because of the uniaxial contraction of the FEM images caused by the small angle of incidence of the electron beam. Therefore, the island shape was analyzed by AFM after the samples were removed from the UHV chamber of the electron microscope. Although a film of natural silicon oxide forms on exposure to the atmosphere, AFM allows reliable images of monatomic steps buried under the natural-oxide layer to be obtained. Figure 5 shows an AFM image of the Si(111) surface after exposure to the oxygen atmosphere under conditions of etching via the 2D-island mechanism. The brighter the contrast, the higher the corresponding surface area. We can see three terraces separated by two monatomic steps, whose height equals one interplane spacing in the ⟨111⟩ direction in Si with an accuracy of 0.1 nm. One can also see faceted triangular islands one interplane spacing deep, which is consistent with the UHV REM observations. Depending on the crystal temperature, the 2D islands varied in shape from regular triangles at low temperatures (<700°C) to perfect circles at high temperatures (>800°C). It should be noted that the monatomic steps contain rectilinear kinks oriented parallel to crystallographic directions (Fig. 5).

Figure 6 shows the time dependences of the intensity of the specularly reflected electron beam, which were measured during etching of the Si surface by oxygen at high temperatures. In the case of epitaxial growth, the oscillations are generally attributed to variations in the surface roughness caused by nucleation and annihilation of 2D islands on the crystal surface [21]. A similar phenomenon takes place during etching of the Si surface by oxygen at high temperatures, with the only difference being that the cyclic change of the relief is caused by the formation of the negative islands. When etching proceeds via the island nucleation, vacancies formed in the central part of a terrace have no time to reach a monatomic step and be incorporated into it. As a result, vacancies coalesce to form a negative island. The islands increase in size due to the further generation of vacancies and their interaction with the islands. The further growth of the islands and their interaction lead to the complete removal of the upper Si layer. In the course of this process, the surface morphology changes from smooth (terraces are free from islands) to rough (islands occupy half the terrace area), which results in the intensity oscillations of the specularly reflected electron beam. Consequently, the period of the oscillations corresponds to the removal of one Si monolayer. At a given oxygen pressure, the period is constant in the temperature range from 540 to 825°C. This finding points to a low activation barrier for the interaction of oxygen with silicon as well as to the fact that etching of the Si surface by oxygen at high temperatures is controlled by the supply of oxygen molecules to the Si surface. An increase in the oxygen pressure in the vacuum chamber enhanced the surface-etching rate determined from the step velocity and reduced the period of the intensity oscillations. It should be noted that the oscillation period is determined by the oxygen pressure, and the temperature range in which the intensity oscillations are observed depends on the oxygen pressure and the interstep distance.

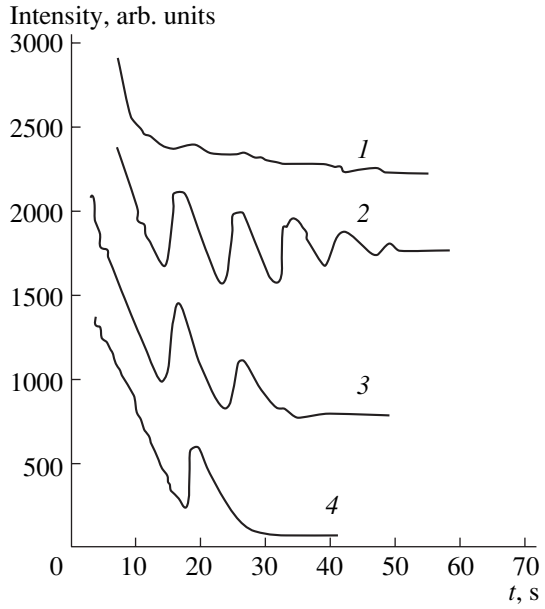


Fig. 6. Time dependences of intensity of the specularly reflected electron beam measured during etching of the Si surface at a constant oxygen pressure and temperatures of (*1*) 835, (*2*) 775, (*3*) 745, and (*4*) 730°C.

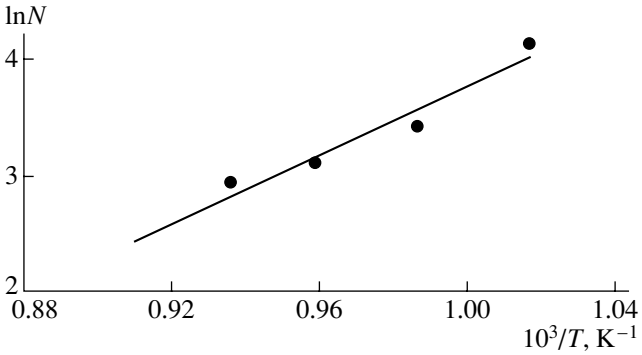


Fig. 7. Arrhenius plot of the number of the negative islands, *N*, nucleated over a given surface area at a constant oxygen pressure.

The density of the 2D islands on the crystal surface specifies an average interisland distance, which under certain conditions is proportional to the migration length of surface vacancies [22]. Direct measurements of the average interisland distance are hampered by two scales of magnification of the REM images caused by

the low angle of view. For this reason, the average distance between the negative islands was determined from a number of islands formed over the same surface area at a constant oxygen pressure. To exclude the effect of the monatomic steps, we chose the surface area located in the central part of the widest terrace. Figure 7 show the Arrhenius plot of the number of the negative islands. Since, as demonstrated above, the rate of the reaction of the SiO formation remains unchanged in the temperature range from 500 to 900°C, the activation energy for the migration of surface vacancies is 1.35 ± 0.15 eV. The same value was obtained from the temperature dependence of the width of the monatomic-step zone free from the 2D islands. This activation energy for the migration of surface vacancies agrees with the values obtained by Shimuzu *et al.* [23] and, within the experimental error, is equal to the activation energy for surface diffusion of Si adatoms on the Si(111) surface [24].

CONCLUSION

Ultrahigh-vacuum reflection electron microscopy was used to study reactions of gases with an atomically clean silicon surface. The initial stages of the interaction of molecular oxygen with the Si(111) surface were studied. Vacancies were demonstrated to form on the Si surface during etching by oxygen at various temperatures and oxygen pressures. The motion of monatomic steps on exposure to the oxygen atmosphere was studied. The activation energy for migration of surface vacancies was estimated from the distribution of the negative 2D islands. Intensity oscillations of the electron beam reflected specularly from the Si surface were observed during etching by oxygen at high temperatures.

ACKNOWLEDGMENTS

This study was supported in part by the Russian Foundation for Basic Research and through the "Physics of Solid-State Structures" and "Surface Atomic Structures" Scientific Programs, project no. 3.5.99.

We are grateful to V.Sh. Aliev for his useful recommendations for the design of the gas-inlet system for admitting the gas into the differential cryogenic system of the electron microscope.

REFERENCES

1. H. Kahata and K. Yagi, Surf. Sci. **220**, 131 (1989).
2. F. M. Ross and J. M. Gibson, Phys. Rev. Lett. **68**, 1782 (1992).
3. K. Murooka, Y. Tanishiro, and K. Takayanagi, Surf. Sci. **275**, 26 (1992).
4. T. Doi, A. Ishizaka, M. Ichikawa, *et al.*, Jpn. J. Appl. Phys. **34**, 2986 (1995).
5. T. Hasegawa, M. Kohno, S. Hosaka, and S. Hosoki, Surf. Sci. **312**, L753 (1994).
6. J. V. Seiple and J. P. Pelz, Phys. Rev. Lett. **73**, 999 (1994).
7. A. V. Latyshev, A. B. Krasilnikov, and A. L. Aseev, Microsc. Res. Tech. **20**, 341 (1992).
8. A. A. Kroshkov, É. A. Baranov, O. A. Yakushenko, *et al.*, Prib. Tekh. Éksp., No. 1, 199 (1985).
9. A. V. Latyshev, A. L. Aseev, A. B. Krasilnikov, and S. I. Stenin, Surf. Sci. **213**, 157 (1989).
10. C. Alfonso, J. C. Heyraund, and J. J. Metois, Surf. Sci. Lett. **291**, L745 (1993).
11. A. V. Latyshev, A. L. Aseev, A. B. Krasil'nikov, *et al.*, Dokl. Akad. Nauk SSSR **300**, 84 (1988) [Sov. Phys. Dokl. **33**, 352 (1988)].
12. Y. Homma, R. J. McClelland, and H. Hibino, Jpn. J. Appl. Phys. **29**, L2254 (1990).
13. R. E. Schlier and H. E. Fransworth, J. Chem. Phys. **30**, 917 (1959).
14. J. J. Lander and J. Morrison, J. Appl. Phys. **33**, 2089 (1962).
15. W. K. Burton, N. Cabrera, and F. C. Frank, Philos. Trans. R. Soc. London, Ser. A **243**, 299 (1951).
16. N. Shimuzu, Y. Tanishiro, K. Kobayashi, *et al.*, Ultramicroscopy **18**, 453 (1985).
17. A. V. Latyshev, A. B. Krasilnikov, and A. L. Aseev, Thin Solid Films **281/282**, 20 (1996).
18. Y. Homma, H. Hibino, T. Ogino, and N. Aizawa, Phys. Rev. B **55**, R10237 (1997).
19. Y. Tanishiro, K. Takayanagi, and K. Yagi, Ultramicroscopy **11**, 95 (1983).
20. A. V. Latyshev, A. B. Krasilnikov, L. V. Sokolov, *et al.*, Surf. Sci. **254**, 90 (1991).
21. J. H. Neave, P. J. Dobson, and B. A. Joyce, Appl. Phys. Lett. **47**, 100 (1985).
22. J. A. Venables, Surf. Sci. **299/300**, 798 (1994).
23. N. Shimuzu, Y. Tanishiro, K. Takayanagi, and K. Yagi, Surf. Sci. **191**, 28 (1987).
24. A. V. Latyshev, A. B. Krasilnikov, and A. L. Aseev, Phys. Rev. B **54**, 2586 (1996).

Translated by N. Izyumskaya

Semiconductors, Vol. 35, No. 9, 2001, pp. 1045–1053. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1092–1101.

SEMICONDUCTOR STRUCTURES, INTERFACES, AND SURFACES

Molecular-Beam Epitaxy of Mercury–Cadmium–Telluride Solid Solutions on Alternative Substrates

Yu. G. Sidorov, S. A. Dvoretskiĭ*, V. S. Varavin, N. N. Mikhaĭlov, M. V. Yakushev, and I. V. Sabinina

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: dvor@isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—Growth processes were considered for heteroepitaxial structures based on a mercury–cadmium–telluride (MCT) solid solution deposited on GaAs and Si alternative substrates by molecular-beam epitaxy. Physical and chemical processes of growth and defect-generation mechanisms were studied for CdZnTe epitaxy on atomically clean singular and vicinal surfaces of gallium-arsenide substrates and CdHgTe films on CdZnTe/GaAs surfaces. ZnTe single-crystalline films were grown on silicon substrates. Methods for reducing the content of defects in CdZnTe/GaAs and CdHgTe films were developed. Equipment for molecular-beam epitaxy was designed for growing the heteroepitaxial structures on large-diameter substrates with a highly uniform composition over the area and their control *in situ*. Heteroepitaxial MCT layers with excellent electrical parameters were grown on GaAs by molecular-beam epitaxy. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Thermal imagers based on the photodetectors for an infrared (IR) wavelength range of 3–12 μm are required for applications both in military equipment for systems of noctovision, detection, and guidance as well as in the national economy for the medical, agricultural, chemical, metallurgical, and fuel industries.

Nowadays, the leading place among materials for the production of IR photodetectors (PDs) is occupied by mercury–cadmium–telluride (MCT) solid solutions. This fact is due to the physical properties of these solutions (high speed, the possibility of varying an MCT band gap within a wide range, and a high quantum efficiency in the range of overlapping wavelengths). For the last 25 years, the technology of MCT production has been developed intensively, which has made it possible to pass from manufacturing bulk single crystals of relatively small diameters (less than 10 mm) to epilayers on large-diameter substrates (up to 75 mm). The MCT epilayers on large-diameter substrates are necessary for the production of IR PD arrays with a large number of elements for enhancing the production efficiency and reducing the cost of devices. According to this, stringent requirements are imposed on the epitaxial technologies of producing such an MCT material. They include a high structural quality and uniformity of photoelectric characteristics over the entire area. The success of epitaxial methods is based both on the progress in increasing the dimensions and improving the structural quality of CdZnTe substrates lattice-matched to the growing MCT film and on the possibility of growing the heteroepitaxial structures (HESs) on the alternative substrates (buffer layers of CdTe and CdZnTe on the substrates of GaAs, Si, and other materials). The use of the alternative substrates makes it possible to markedly reduce the cost of the MCT epitaxial material, especially for the large-scale production of IR PDs. In theory, the design of the systems for processing signals and IR PDs on a single monolithic substrate also makes it possible to improve the characteristics of devices.

Molecular-beam epitaxy (MBE) surpasses other epitaxial methods of growing MCT layers on alternative substrates primarily due to its low growth temperatures (~180°C), which prevents the diffusion of impurities from the substrate and reduces the background doping with these impurities. MBE of MCT attained the level necessary for preparing a material suitable for practical applications [1]. Device-quality material was grown with MCT compositions for the wavelength range from 1 to 20 μm. The hybrid array of 256×256 IR-PD elements was manufactured on the GaAs substrate for the wavelength range of 8–10 μm, and the imagery was demonstrated [2]. Great efforts are being made to grow the epitaxial MCT films on large-diameter silicon substrates [3]. MBE makes it possible to produce MCT heteroepilayers for multicolor IR PDs [4].

In the case of MCT heteroepitaxy on alternative substrates by MBE, it is necessary to solve complicated physical and technological problems.

The physical problems in MCT heteroepitaxy on GaAs and Si substrates are induced by the great differences in the lattice constants of contacting materials, the different nature of chemical bonding, and the low

rates of dissociation and reevaporation of diatomic tellurium molecules. This fact leads to the degradation of the structural quality and to an increase in the dislocation density in MCT HESs by approximately one–two orders of magnitude compared to bulk crystals. An intense defect generation leading to a high density of microtwins and *V* defects can considerably deteriorate the electrical properties of MCT layers and IR PDs based on them.

To overcome the technological difficulties, it is necessary to solve the problems of designing complicated equipment, including an ultrahigh-vacuum MBE system for providing clean growth conditions and methods for *in situ* monitoring (making it possible to control the growth processes in real time). To ensure a high uniformity of MCT composition (X_{CdTe}) over the MCT HES area on GaAs and Si substrates (especially in the case of large-diameter substrates), it is necessary to design systems of molecular fluxes and implement the procedure for *in situ* monitoring, making it possible to attain the desired uniformity and accuracy in setting and maintaining the composition ($\Delta X_{CdTe} \le 0.005$).

The purpose of this study was to gain insight into physical and chemical processes in the heteroepitaxy of MCT layers on alternative substrates deposited by MBE as well as to design and produce the equipment for growing the MCT HESs with a high structural quality and good electrical parameters suitable for the production of highly efficient multielement IR PDs.

2. HETEROEPITAXY OF MERCURY–CADMIUM–TELLURIDE SOLID SOLUTIONS ON ALTERNATIVE SUBSTRATES

In the case of growing the II–VI compounds on GaAs and Si substrates, the defect generation in the growing film is associated with the following factors:

(i) a difference in the lattice parameters of a film and substrates; and

(ii) the chemical interaction between components on the heterointerface.

There are two principal reasons for the formation of structural defects during epitaxy: (a) the presence of mechanical stresses in the epilayer (usually, in the case of heteroepitaxy); and (b) stacking faults formed during crystallization. In the former case, the defects represent the necessary equilibrium components of the crystal because they provide the relaxation of stresses; their type, density, and distribution are independent of the epitaxy conditions. In the latter case, on the contrary, the defect-generation process depends on the film nucleation and growth conditions.

In the case of MCT-film growth on CdZnTe/GaAs and CdZnTe/Si substrates, growth proceeds at low temperatures. In this case, the growth and the defect generation are specified by the rates of the incorporation of components into the crystal lattice and by the rates of their dissociation and reevaporation.

2.1. CdTe/GaAs(001) Heteroepitaxy

The experimental investigation of the processes accompanying the growth of the CdTe films on an atomically clean surface of a GaAs(001) substrate carried out in [5] showed that, at the initial stage, the formation of CdTe(111) or CdTe(001) is possible depending on the state of the GaAs-substrate surface and conditions of performing the process. At low temperatures (<570 K), the growth begins with the appearance of an intense diffuse background at the high-energy electron-diffraction (HEED) pattern, with the reflections and Kikuchi lines from the substrate disappearing. Then, low-intensity streaks from the CdTe(111) film appear. The intensity of these streaks increases with the growth time, while the background intensity decreases. The appearance of streaks from the CdTe(111) films at the earliest stages of growth points to the fact that the CdTe(111) films grow by the two-dimensional mechanism. At a growth temperature above 590 K, the appearance of a diffuse background is not observed. During such growth, along with the reflections from the CdTe(001) film, the reflections from GaAs can be seen. The CdTe(001)-film formation begins with the initiation of isolated islands, which is recognized from the appearance of point reflections in the diffraction pattern. Thus, for a given intensity of molecular flow and depending on a GaAs(001)-substrate temperature, CdTe-film nucleation can occur with the (111) orientation (at relatively low temperatures), with the (001) orientation (at relatively high temperatures), and in the form of a mixture of these orientations within the intermediate temperature range. The change of the CdTe-film orientation depends on the reconstruction of surface complexes based on tellurium and gallium. An increase in the substrate temperature leads to an increased probability of nucleation with the (001) orientation.

According to the existing concepts [6], the growing continuous crystalline layer is pseudomorphic at the initial stage of epitaxy. The calculation shows that no continuous pseudomorphic layer should exist at the initial stage in the case of a lattice mismatch of the contacting materials GaAs/CdTe (~13.6%). The observed pattern of CdTe(111) growth can be explained by the formation of the unordered (amorphous) phase at the initial stage [7]. In fact, there are conditions for which it is possible to obtain the amorphous phase, which is more energetically favorable, instead of the expected epitaxy.

The typical structural defects in the CdTe(111) films grown on the GaAs(001) and on the CdTe(111) substrates are the microtwins in the form of extended lamellas and the dislocation networks [8, 9]. In this case, the twinning planes {111} are parallel to the growth surface. Transmission electron microscopy (TEM) investigations of planar foils parallel to the (111) growth surface, of the cross cuts of the films along the (110) planes, and of thin crystals made it pos-

sible to find a correlation between the dislocation density in networks lying in the (111) planes and the microtwin density in the [111] growth direction. The dislocation density increases with the microtwin density. The defect density increases with a decrease in the epitaxy temperature. Similar patterns are observed in the homoepitaxial CdTe films. This fact testifies that the formation of the observed dislocation networks is not a consequence of heteroepitaxial stresses and is associated with the twinning process. On the singular surface GaAs(100) at the CdTe(111)-film nucleation stage, an intense twinning takes place due to the equal probability of the formation of nuclei turned 180° relative to another around the axis perpendicular to the growth surface [8]. The deviation of the growth-surface orientation from the singular one leads to a decrease in the spacing between the steps. Thereby, the probability of two-dimensional-nucleus formation on the terraces decreases. Therefore, the twinning mechanism becomes suppressed at the steady-state stage, and the twin-lamella density and the dislocation density decrease in the CdTe(111) films.

2.2. Heteroepitaxy on the GaAs Vicinal Faces

The consideration of the processes of CdTe-film growth on GaAs(001) and of defect generation necessitates using the GaAs vicinal surfaces for epitaxy. To grow CdTe with a single orientation on the GaAs surface, it is necessary to begin epitaxy with a compound whose physical and chemical properties are close to those of CdTe but which have less lattice mismatch. This makes ZnTe suitable, since its lattice mismatch with GaAs amounts to ~7%.

It is known that the formation of intermediate compounds of the III_2–VI_3 type [10] occurs on the heterointerface between II–VI and III–V. These compounds can affect the mechanism of growth and defect generation.

In order to gain insight into the chemical interaction with the GaAs vicinal surface and the growth of single-orientation films, we studied the growth of ZnSe on the GaAs surface. The ZnSe–GaAs system can be considered as a model because both the film and the substrate have virtually identical lattice constants, whereas the tendency to form chemical compounds with the film and surface components manifests itself more strongly in this case than for a number of other systems, for example, CdTe–GaAs.

We carried out a crystallochemical consideration of the interaction between the components of the ZnSe/GaAs heterosystem. This consideration was based on the following statements. In the interaction with GaAs, selenium replaces arsenic from the GaAs crystal lattice and forms a chemical bond with gallium [11]. In the crystal lattice with the tetrahedral configuration of bonds (as in GaAs), there must be four valence electrons per atom [12]. As a result of this consideration, the following conclusions were made.

(1) The Se/GaAs surface has saturated bonds.

(2) If the gallium selenide phase is formed on the gallium arsenide surface, terracing by the (111) planes becomes favorable.

(3) On the surface with the saturated bonds, the simultaneous nucleation of the crystal lattice is possible both in the normal and twinned configurations. Therewith, the dimensions of normal and twinned areas do not exceed several interatomic spacings at the initial moment [13].

These experiments validated the above assumptions. Using high-energy electron diffraction and X-ray photoemission spectroscopy, it was established that an excess of selenium in the ZnSe/GaAs heterojunction induces the terracing of the growing film by the (111) planes and twinning [13, 14].

The results of the experimental investigations and the crystallochemical consideration show that the chemical interaction of film components and the substrate in the (II–VI)/GaAs heterosystem plays a part in the formation of the morphology and generation of structural defects. The major driving force for the morphological reconstructions and defect generation at a heterointerface is the disturbance of a mean number of valence electrons per formula unit owing to the formation of Ga–Se bonds in the lattice with the tetrahedral coordination. The tendency of the heterosystem to eliminate the excess valence electrons arising at the heterointerface leads to faceting, twinning, and the violation of the stoichiometry of the growing structure.

The GaAs(112)B surface is inclined by ~19° to the GaAs(111)B surface. We investigated the growth processes for the ZnTe and CdTe films on the GaAs(013) surface, which is inclined to the GaAs(001) surface by ~19° and, as is shown further, is the basic one for the MCT heteroepitaxy on GaAs substrates. With allowance for the above physical and chemical investigations of the growth processes, we determined the optimal conditions for growing structurally perfect CdTe layers on GaAs(013). At the initial stage, ZnTe was grown up to thicknesses of ~1 μm. Further, CdTe was grown up to thicknesses of ~5–10 μm. The HEED patterns obtained during growth show that the (1×1) and (1×2) structures are observed in the azimuth of $[\bar{1}35] \times [\bar{1}3\bar{2}]$. It was established that the (1×2) structure corresponds to an excess of tellurium on the growing surface, while the (1×1) structure corresponds to an excess of cadmium or zinc. In the case of growth with an excess of tellurium, a relief is developed and only point reflections corresponding to the three-dimensional diffraction are observed. In the case of growth with an excess of cadmium or zinc, no relief is developed and two-dimensional diffraction is observed. In the [001] azimuth, the diffraction reflections are extended and their highest intensities are shifted

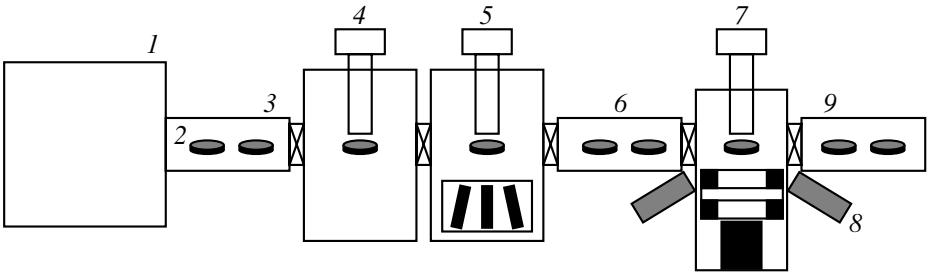


Fig. 1. Schematic diagram of the apparatus for MBE growth of MCT heteroepitaxial structures. (*1*) is the laminar box, (*2*) is the substrate with carrier, (*3*) is the module for loading the substrates, (*4*) is the module for preparing the surface of substrates, (*5*) is the module for growing the buffer layers, (*6*) is the accumulation chamber, (*7*) is the module for growing the MCT layers, (*8*) is the built-in automated ellipsometer, and (*9*) is the module for unloading the epitaxial structures.

towards the sample shadow. This is likely explained by the smoothening of the surface and the formation of a distinct system of equidistant steps. In this case, we obtain mirror-smooth buffer layers with a minimum density of morphological defects. These growth conditions make it possible to grow reproducibly only the (310) buffer-layer orientation and to eliminate the appearance of unordered phases at the initial stage of growth.

2.3. Heteroepitaxy on Silicon Substrates

A silicon substrate is more promising from the viewpoint of developing the IR PDs and signal-reading systems. However, the growth on the silicon substrate is more complicated due to a very large lattice mismatch both for CdTe with Si (~19%) and for ZnTe with Si (~12%). It is also necessary to take into account the change in the balance of valence electrons in the heterojunction, which decisively influences the structure and the morphology of the epilayer formed.

Using ellipsometry and HEED, we studied the processes of ZnTe nucleation on the surfaces of Si substrates treated with As [15] because we failed to obtain a ZnTe single-crystal film on a clean Si surface. The investigations were carried out on the substrates with the (013) and (113) orientations. Based on the known temperatures of molecular tellurium and zinc sources, we found a value x corresponding to the composition of Zn_xTe clusters nucleating on the Si surface. Thus, it was established that x is approximately equal to 2 for the (013) orientation and to 1 for the (113) orientation. It was found that, at the same substrate temperature, lower pressures of zinc and tellurium vapors are needed for the onset of deposition on (013) than on (113). It was also established that epitaxial nucleation takes place only in a certain range of ratios between vapor pressures of components and that these ranges are different for (013) and (113). In the case of nonepitaxial nucleation, the value of x changes, which points to the deposition of various-composition clusters at the initial stage.

For the nucleation of a ZnTe single-crystal layer, it is necessary to alternate two atomic configurations of the Zn_2Te and $ZnTe_2$ types on the surface. Because zinc adsorption is hampered, the nucleation of Zn_2Te clusters is the limiting process at the initial stage of formation of the ZnTe film on (013)Si. To ensure that the conditions are favorable for nucleating clusters with the Zn_2Te composition, it is necessary to have a zinc-vapor pressure which considerably exceeds that of tellurium.

3. EQUIPMENT FOR GROWING THE MCT HETEROEPILAYERS BY MOLECULAR-BEAM EPITAXY

Allowing for the results of physical and chemical investigations, the equipment was designed and manufactured, including the automated system for controlling technological processes and the means for monitoring the MBE-grown MCT layer quality *in situ*. This equipment is intended for the large-scale production of the material, and the basics of its design were described in [16, 17]. Its basis is a "Katun'" MBE system modernized for operation with mercury. The schematic diagram of this system is shown in Fig. 1. It involves three technological chambers, three loading–unloading chambers, and a laminar dust-free box.

For reducing the defect density at the (II–VI)/GaAs heterointerface, it is necessary to prevent the interaction between tellurium vapors and the GaAs substrate at high temperatures. For this purpose, the processes of removing residual oxides and growing a buffer layer are spatially separated to individual technological chambers.

The residual atmosphere in the technological chamber (*4*) of the preepitaxial thermal preparation of the substrate surface contains no tellurium vapors. In this chamber, oxides are thermally removed from the substrate surface. The chamber involves a high-energy electron diffractometer and an incorporated automated ellipsometric goniometer.

The residual atmosphere in the chamber (*5*) for growing buffer layers contains tellurium vapors at a level of 10^{-6} Pa and, therefore, is unsuitable for the

preepitaxial preparation of substrates. The chamber involves a block of molecular effusion sources, a high-energy electron diffractometer, and an incorporated automated ellipsometer. The ellipsometry can be used for monitoring various stages of the technological process of MCT-film deposition, such as the preepitaxial preparation of substrates, the growth of buffer layers, and the synthesis of MCT films [18–21].

In order to grow MCT epilayers, we designed and built a chamber unit (*7*) for growing mercury-containing compounds provided with a nonconventional system of molecular sources, which was intended for growing large-diameter (102 mm) MCT layers homogeneous over an area without substrate rotation. The effusion molecular sources of cadmium and mercury have ring scatterers of fluxes. The cylindrical source of tellurium is mounted coaxially with the scatterer of molecular fluxes of cadmium and mercury. This system of sources provides a high composition homogeneity of MCT epitaxial films over a large area with a minimum expenditure of expensive evaporated materials. The chamber involves the incorporated automated ellipsometer (*8*) used for measuring the MCT growth rate at the initial stage and for monitoring uninterruptedly (in real time) the composition and morphology of the growing MCT layer. The substrate temperature during growth is measured using a special pyrometer [22]. The operation of the pyrometer is based on measuring the difference between the orthogonally polarized components of thermal emission from the substrate surface in the long-wavelength IR region of the spectrum.

The oil-free pumping of the chambers was accomplished by a system consisting of a cryogetter and magnetic-discharge pumps. The chambers *4* and *5* are additionally equipped with titanium-sublimation pumps for obtaining an ultrahigh vacuum. Owing to a high mercury-vapor pressure in the MCT-growth module, there is a special system of liquid-nitrogen traps.

This automated control system for the technological process precisely maintains the conditions for growing the buffer layers on the GaAs substrate and MCT films. The use of an automated ellipsometer makes it possible to monitor the variation in layer composition over thickness according to the given program or to maintain it with a high accuracy by correcting the temperatures of molecular sources.

4. PHYSICAL AND CHEMICAL FEATURES OF GROWING MERCURY–CADMIUM–TELLURIDE FILMS BY MOLECULAR-BEAM EPITAXY

A thermodynamic analysis shows that, in the case of MBE, the MCT films grow under conditions in which the $HgTe_{cr}$ and Te_{cr} phases are stable [23, 24]. In Fig. 2, we show the calculated dependence of the excess free energy on the deposition temperature for HgTe (curve *1*) and for Te (curve *2*) at a mercury-vapor pressure of 10^{-3} Torr and a deposition rate of 1 μm/h. With lowering the temperature below a certain critical level (T_1), the tellurium crystallization turns out to be possible, whereas the crystallization of mercury telluride is still impossible. By further lowering the temperature (below T_2), the nucleation of crystalline mercury telluride turns out to be thermodynamically possible, but the possibility of depositing tellurium atoms is also retained. Thus, if the nucleation of the HgTe phase meets no kinetic obstacles, the probability of nucleation of the elementary-tellurium phase is reduced. If the crystallization of HgTe is hampered, the probability of the nucleation of elementary tellurium increases.

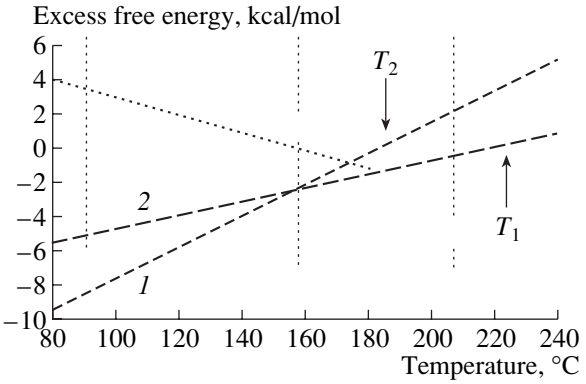


Fig. 2. Temperature dependence of excess free energy: (*1*) for HgTe (deposition is possible at $T < T_2$), and (*2*) for tellurium (deposition is possible at $T < T_1$).

Under these conditions, the nucleation of a particular phase depends on the kinetics of the corresponding process of crystallization, and the nonoptimal choice of the growth conditions leads to the irreversible deterioration of the MCT-film structure in the process of growing. In fact, according to the experimental results, a gradual deterioration in the film structure is possible as the growth duration or the growing-film thickness increases. In Fig. 3, we illustrate the basic possible processes occurring on the surface and involving tellurium. The tellurium molecules incident on the surface can be involved in one of two processes: the dissociation of molecules and the crystallization of a perfect MCT film or tellurium crystallization in the form of an individual phase when the dissociation has no time to occur. In the latter case, the nucleation of the tellurium phase on the surface upsets the MCT crystallographic growth and leads to an avalanche-like multiplication of defects because, according to above, the difficulties in MCT crystallization in damaged areas increase the probability of elementary-tellurium nucleation. The MCT-film growth requires a strict maintenance of the growth conditions and a high quality of the buffer-layer surface. Under nonoptimal growth conditions, for example, with a deficit or excess of mercury, or the original irregularity of the substrate surface (the relief or the pres-

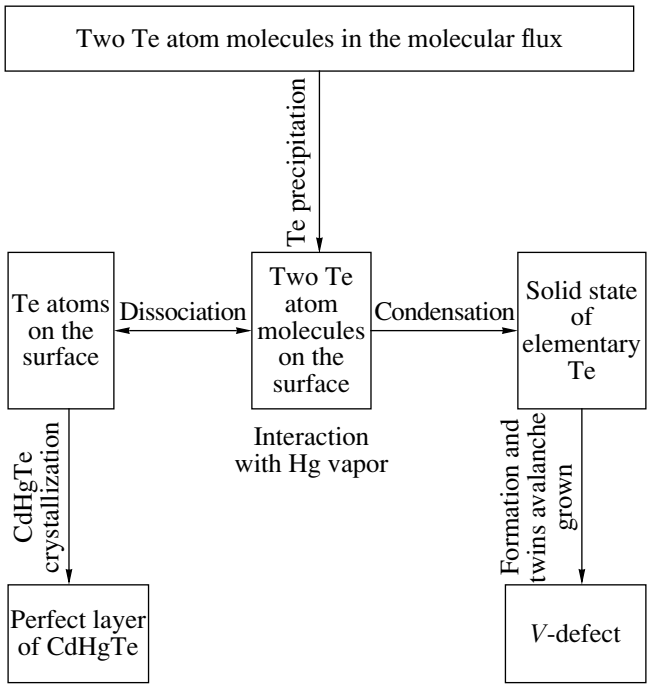


Fig. 3. Schematic diagram of the processes with participation of tellurium occurring on the surface during the MCT growth.

ence of a high density of defects, which can be related to each other), irreversible degradation of the surface and the MCT structure during MBE can occur.

The estimations carried out for tellurium showed that the dissociation of the diatomic tellurium molecule was energetically favorable near a step. It follows from the calculation that, for a suitable step configuration, the dissociation can be accompanied by an energy gain of 10 kcal/mol, whereas the dissociation energy for a tellurium molecule on a singular surface amounts to ~30 kcal/mol.

The above analysis makes it possible to suggest a reliable way to facilitate the tellurium-molecule dissociation, specifically, the use of the substrates deviated from the singular orientations.

The defect structure of films was investigated by optical microscopy and PEM [25]. The *V* defects, whose size increases with distance from the film–substrate interface, are distinctly visible. The interior contains polycrystalline grains, their sizes decreasing closer to the surface. In the grains, we observed intensive twinning along the (111) plane, which is evidenced by the presence of twin reflections in the microdiffraction pattern. The microdefects are bounded by the twin lamellas along the inclined (111) planes. The initial stage of *V*-defect formation is likely the complicated defects, i.e., the clusters of elementary tellurium or tellurium compounds. If the density of tellurium clusters and the film thickness are sufficiently large, the coalescence of neighboring *V* defects takes place with the nucleation of a MCT polycrystalline film.

The possibilities of incorporating an above-stoichiometry tellurium in growing MCT films and the relation of this process with the *V*-defect density were clarified experimentally. The MCT-film composition was measured by an SX50 CAMEBAX X-ray microanalyzer. It was shown that there is an excess content of tellurium in the films of both mercury telluride and MCT [24]. It was found that the concentration of above-stoichiometry tellurium increases with defect density. In the film regions without defects, the composition is close to stoichiometric. In the regions with a high density of *V* defects, there is 3% excess of tellurium. The width of the MCT-homogeneity region is less than 0.01%. Therefore, the excess tellurium is present in the form of second-phase precipitations.

A comparison between the *V*-defect density on different-orientation substrates reveals the region of orientations near the (130) plane on which the defect density is lowest under the conditions used (a sample temperature, a vapor pressure for tellurium and mercury). The films were grown under conditions deviated from optimal. Under these conditions, the growth on the (111)B, (112)B, and (110) planes is accompanied by the *V* defect formation with a high density (10^6 cm^{-2} and higher).

The optimization of growth conditions makes it possible to reduce the *V*-defect density in the MCT films for all the orientations, but the lowest density (10^2 cm^{-2}) [24] is attained on the (103) substrates.

It was experimentally established that the preferential nucleation of *V* defects occurs near the film interface with the buffer layer. This can be caused by the microtwins formed in the buffer layer and also by special features of the initial stages of the MCT-film growth. As was found, at the initial stages of growth, reterracing occurs owing to a mismatch between the surface microrelief and the growth conditions, which reduces the growth rate and adds complexity to the relief, thus leading to an enhanced probability of twinning.

5. CHARACTERISTICS OF HETEROEPITAXIAL MCT STRUCTURES GROWN BY MOLECULAR-BEAM EPITAXY ON GaAs SUBSTRATES

The production of heteroepitaxial MCT structures on gallium-arsenide substrates by MBE invariably includes the consecutive growth of CdZnTe buffer layers and MCT films. Routinely, the buffer-layer thickness is 5–10 μm. The MCT-layer thickness amounts to 5–15 μm and depends on the IR PD structure.

The composition distribution over the area was determined from the transmission spectra measured using a Bruker Fourier spectrometer for the films grown

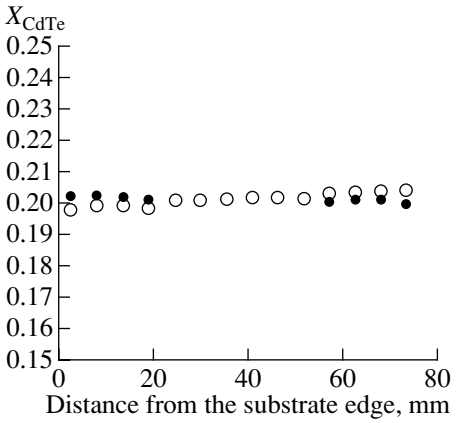


Fig. 4. Uniformity of composition for the MCT film on the GaAs substrate of 76.2 mm in diameter along two mutually perpendicular diameters.

on GaAs substrates of 76.2 mm in diameter (see Fig. 4). The following characteristics of composition distribution are obtained: the mean value $X_{CdTe} = 0.20171$, the standard deviation $\Delta X_{CdTe} = 0.000164$, and the relative deviation $\Delta X/X = \pm 0.654\%$. Such a high uniformity for a heterosystem of 76.2 mm in diameter is comparable with the best foreign data obtained with substrate rotation [26].

Owing to the continuous ellipsometric monitoring *in situ* in the course of growth, the MCT layers can have a specified composition profile over thickness; for example, an actual layer of constant composition and graded-gap layers can be formed. In Fig. 5, we show a variation in composition over thickness measured by an ellipsometer during growth for a typical MBE MCT heterosystem with graded-gap layers. The graded-gap layers with an increased content of CdTe can be used for passivating the surface [27]. An increase in the band gap at heterointerfaces and the MCT-film surface induces the built-in fields that push the nonequilibrium carriers away from the surfaces with an enhanced recombination rate. Thus, we can increase the effective lifetime of the nonequilibrium charge carriers [28].

After growth, the films are *n*-type. A carrier concentration in the *n*-type films is within 10^{14}–10^{15} cm^{-3} with a mobility of (5–15) $\times$ 10^4 cm^2/(V s) for X_{CdTe} = 0.2–0.22. The mobility for the composition with $X_{CdTe} = 0.158$ attains a value of 7.1×10^5 cm^2/(V s) at a temperature of 77 K and is comparable with the mobility in the best bulk material.

The investigation of the uniformity of electrical properties in grown layers was carried out by measuring the magnetofield dependences; we processed experimental data taking into account the possibility of the simultaneous presence of different-type carriers. The results confirm that the MBE makes it possible to obtain a highly homogeneous MCT.

The most important parameter of materials used in the production of photodetectors is the lifetime of the photoexcited charge carriers, whose value is appreciably affected by the presence of the recombination centers in the films. The MCT layers with the graded-gap layers have unprecedentedly large values of minority-carrier lifetimes at 77 K for a material grown on GaAs substrates. The temperature dependence of the minority-carrier lifetime in the film with $X = 0.22$ is shown in Fig. 6. After etching off the upper graded-gap layer, the relaxation time of nonequilibrium carriers decreases by four to five times at 77 K, which testifies that the effect

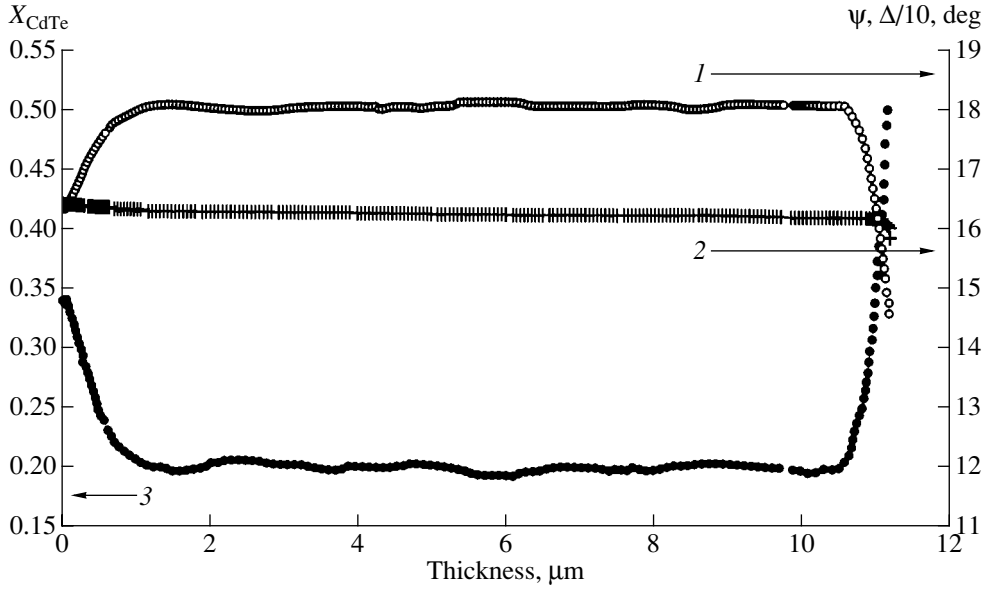


Fig. 5. Results of ellipsometric measurements *in situ* over an MCT-film thickness: (*1*) and (*2*) are the values of ellipsometric angles ψ and Δ, respectively, and (*3*) is the composition profile.

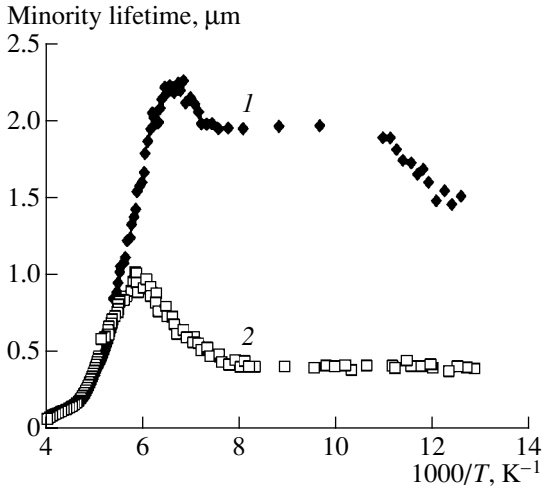


Fig. 6. Temperature dependence of the minority-carrier lifetime in MBE MCT heterosystem for the composition with $\Delta X_{CdTe} \cong 0.22$: (*1*) with graded-gap layers and (*2*) after removing the upper graded-gap layer by chemical etching.

of the graded-gap layers is beneficial. As to the lifetime, the structures grown exceed similar ones described in the literature [29]. It should be noted that the concentration of carriers in the MCT structures with wide-gap layers at the boundaries does not differ from that for the structures without the graded-gap layers. It is possible that the dislocations affect the lifetime not so markedly as assumed [30].

The *n*-type films become *p*-type with a hole concentration of $(0.5–2) \times 10^{16}$ cm^{-3} and a mobility of 300–600 cm^2/V s at an annealing temperature of 230°C, a mercury temperature of 30°C, and an annealing duration of 20 h. For thermal treatments, we used cells filled with a gas (hydrogen or helium). No dependence of the annealing results on the gas type was noted. The cell was placed in a two-zone furnace. One zone was designed for heating a reservoir with mercury, and the second zone, for heating the sample. The conversion to *p*-type is reversible, and annealing at 230°C at a mercury temperature above 180°C again gives rise to the *n*-type of conductivity.

The totality of experimental data on the effect of annealing conditions on the MCT-film properties also suggests the presence of donor centers with a concentration dependent on the growth conditions in addition to mobile accepter centers with a variable concentration in the MCT films when using GaAs substrates. The major donor centers in the MCT films grown by the MBE are likely the antisite tellurium atoms. The experimental data on the variation of the concentration of donor centers in the MCT films with a growth temperature corroborate this assumption. The estimations of equilibrium concentrations of the donor centers introduced both by the intrinsic point defects and the most probable impurities show that the equilibrium concentrations of donor centers under MBE conditions are lower than 10^7–10^{10} cm^{-3}. The model of the nonequilibrium dissolution of defects in MCT with allowance for deviation from the equilibrium predicts an increase in the antisite-tellurium concentration to values of 10^{14}–10^{15} cm^{-3}. As the growth temperature and content of CdTe in the solid-solution state increase, the antisite-tellurium concentration decreases compared with the calculation for the equilibrium state.

From the MBE MCT heterosystems grown, we fabricated photoresistors and photodiodes, individual and in the form of linear and two-dimensional arrays; the latter included as many as 256×256 elements. The devices were sensitive in the spectral ranges of 3–5 and 8–14 μm and operated at 200 K and liquid-nitrogen temperatures with the photoelectric parameters limited by the background radiation [31–33].

6. CONCLUSION

Molecular-beam epitaxy is the most promising method for the preparation of structures on the basis of MCT. This method makes it possible to reduce the cost of MCT in the case of the mass production on GaAs and Si substrates, which is a necessary condition for the wide use of infrared sensors. This method has virtually unlimited potential for improving the performance of infrared sensors and provides opportunities to prepare the structures for new classes of devices.

The investigations of growth mechanisms and defect generation in the films of II–VI compounds grown by MBE on GaAs substrates showed that unordered-phase nucleation and intensive twinning at the heterointerface can occur at the initial growth stages. The methods for suppressing disordering at the initial stages of heteroepitaxy on the basis of the developed concepts about the formation of intermediate compounds at the heterointerface are suggested. These methods open the way for solving the problem of designing high-quality heteroepitaxial structures in the case of heteroepitaxy of nonisovalent semiconductor compounds.

A low rate of dissociation and a low tellurium-vapor pressure at low growth temperatures inherent for molecular-beam epitaxy can lead to an intense generation of defects. The proper choice of substrate orientation and growth conditions makes it possible to increase the probability of nucleation of the MCT-crystalline phase (compared to the elementary-tellurium phase) and to suppress the generation of threading defects (the *V* defects).

The problem of establishing the relation between the growth conditions and electrical properties of MCT heteroepilayers adjoins directly the above problem because the intrinsic point defects in MCT are electrically active and their concentration is directly related to the growth conditions. The use of physical and chemical concepts about the mechanism of defect-structure formation and the nature of electrically active impuri-

ties provides the basis for controlling electrical parameters of MCT heteroepilayers.

The equipment is built for molecular-beam epitaxy of MCT providing the preparation of structures with a highly uniform distribution of composition over area (the gradient of composition does not exceed 0.001 cm^{-1}) on the substrates of 76.2 mm in diameter with the continuous monitoring of composition in the course of growing with an accuracy better than $\Delta X_{CdTe} \le 0.005$. The consecutive use of an ellipsometer without affecting the surface processes for monitoring the critical technological processes increases their reproducibility and opens the way for the automation of growth of MCT heteroepilayers by molecular-beam epitaxy with *in situ* layer-quality control.

Using the developed technology of molecular-beam epitaxy, we grew MCT heteroepilayers on GaAs with a low carrier concentration (conduction-electron concentration is 10^{14}–10^{15} cm^{-3} at 77 K). The minority-carrier lifetimes in the MCT heteroepilayers on GaAs substrates with graded-gap layers (with a large band gap at the film–substrate heterointerface and near the MCT-film surface) amount to 1–3 μs (for ΔX_{CdTe} = 0.2–0.22 at 77 K).

The level of technology and the equipment gives virtually unlimited possibilities for modifying the parameters over thickness of the MCT layer and provides the basis for developing new generations of photodetectors.

ACKNOWLEDGMENTS

We thank A.P. Antsiferov, L.D. Burdina, O.I. Malyshev, A.S. Mardezhov, D.N. Pridachin, and V.A. Shvets for their help in conducting this study.

REFERENCES

1. O. K. Wu, Proc. SPIE **2021**, 79 (1993).
2. A. Kawahara, A. Ajisawa, K. Myamoto, *et al.*, Proc. SPIE **2552**, 411 (1995).
3. T. J. deLyon, R. D. Ravajal, J. A. Vigil, *et al.*, J. Electron. Mater. **27**, 550 (1999).
4. D. M. Jamba, J. E. Jensen, P. D. Brewer, *et al.*, J. Electron. Mater. **27**, 747 (1998).
5. V. A. Zubkov, V. V. Kalinin, V. D. Kuz'min, *et al.*, Poverkhnost, No. 9, 45 (1991).
6. J. W. Matthews, J. Vac. Sci. Technol. **12**, 126 (1975).
7. Yu. G. Sidorov and E. M. Trukhanov, Poverkhnost, No. 6, 106 (1992).
8. S. A. Dvoretskiĭ, V. I. Budarnykh, A. K. Gutakovskiĭ, *et al.*, Dokl. Akad. Nauk SSSR **304**, 604 (1989) [Sov. Phys. Dokl. **34**, 5 (1989)].
9. I. V. Sabinina, A. K. Gutakovsky, V. D. Kuzmin, and Yu. G. Sidorov, Phys. Status Solidi A **126**, 181 (1991).
10. A. Krost, W. Richter, D. R. T. Zahn, *et al.*, Appl. Phys. Lett. **57**, 1981 (1990).
11. T. Scimeca, Y. Watanabe, and F. Maeda, Appl. Phys. Lett. **62**, 1667 (1993).
12. N. A. Goryunova, *The Chemistry of Diamond-like Semiconductors* (Leningr. Gos. Univ., Leningrad, 1963; Chapman and Hall, London, 1965).
13. M. V. Yakushev, Yu. G. Sidorov, and L. V. Sokolov, Poverkhnost, No. 10, 35 (1996).
14. V. G. Kesler, M. V. Yakushev, L. M. Logvinskiĭ, and Yu. G. Sidorov, Poverkhnost, No. 2, 58 (1997).
15. Yu. G. Sidorov, M. V. Yakushev, D. N. Pridachin, *et al.*, Thin Solid Films **367**, 203 (2000).
16. Yu. G. Sidorov, S. A. Dvoretsky, M. V. Yakushev, *et al.*, Thin Solid Films **306**, 253 (1997).
17. Yu. G. Sidorov, S. A. Dvoretskiĭ, N. N. Mikhaĭlov, *et al.*, Opt. Zh. **67** (1), 39 (2000) [J. Opt. Technol. **67**, 31 (2000)].
18. K. K. Svitashev, V. A. Shvets, A. S. Mardezhov, *et al.*, Avtometriya, No. 4, 100 (1996).
19. A. S. Mardezhov, N. N. Mikhaĭlov, and V. A. Shvets, Poverkhnost, No. 12, 92 (1990).
20. K. K. Svitashev, S. A. Dvoretsky, Yu. G. Sidorov, *et al.*, Cryst. Res. Technol. **29**, 931 (1994).
21. K. K. Svitashev, V. A. Shvets, A. S. Mardezhov, *et al.*, Zh. Tekh. Fiz. **65** (9), 110 (1995) [Tech. Phys. **40**, 924 (1995)].
22. S. A. Dvoretskiĭ, S. A. Dulin, N. N. Mikhaĭlov, S. V. Rykhlitskiĭ, and Yu. G. Sidorov, RF Patent No. 2149366 (2000).
23. V. S. Varavin, S. A. Dvoretsky, V. I. Liberman, *et al.*, Thin Solid Films **267**, 121 (1995).
24. V. S. Varavin, S. A. Dvoretsky, V. I. Liberman, *et al.*, J. Cryst. Growth **159**, 1161 (1996).
25. I. V. Sabinina, A. K. Gutakovsky, Yu. G. Sidorov, *et al.*, J. Cryst. Growth **117**, 238 (1992).
26. P. S. Wijewarnasuriya, M. Zandian, D. D. Edwall, *et al.*, J. Electron. Mater. **27**, 546 (1998).
27. R. K. Bhan, V. Dhar, and P. K. Chaudhury, Appl. Phys. Lett. **68**, 2453 (1996).
28. A. V. Voĭtsekhovskiĭ, Yu. A. Denisov, A. P. Kokhanenko, *et al.*, Avtometriya, No. 4, 51 (1996).
29. J. M. Arias, J. G. Pasko, M. Zandian, *et al.*, Proc. SPIE **2228**, 210 (1994).
30. J. M. Arias, M. Zandian, S. H. Shin, *et al.*, J. Vac. Sci. Technol. B **9**, 1646 (1991).
31. E. V. Susov, Yu. G. Sidorov, V. N. Severtsev, *et al.*, Avtometriya, No. 4, 40 (1996).
32. V. N. Severtsev, E. V. Susov, V. S. Varavin, *et al.*, Avtometriya, No. 4, 21 (1998).
33. V. N. Ovsyuk, Yu. G. Sidorov, V. V. Vasil'ev, and V. V. Shashkin, Prikl. Fiz., No. 5, 58 (2000).

Translated by V. Bukhanov

Semiconductors, Vol. 35, No. 9, 2001, pp. 1054–1062. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1102–1110.

SEMICONDUCTOR STRUCTURES, INTERFACES, AND SURFACES

Epitaxial Growth, Electronic Properties, and Photocathode Applications of Strained Pseudomorphic InGaAsP/GaAs Layers

V. L. Alperovich, Yu. B. Bolkhovityanov*, S. I. Chikichev, A. G. Paulish, A. S. Terekhov, and A. S. Yaroshevich

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: bolkhov@isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—The results of experimental and theoretical investigations directed toward the development of highly efficient sources of spin-polarized electrons are reported. The sources are based on heteroepitaxial elastically strained films of the quaternary InGaAsP solid solution grown by liquid-phase epitaxy on GaAs substrates. The InGaAsP films synthesized were 0.1–0.2 μm thick with the band gap being within the range of 1.4–1.9 eV and having elastic strains as high as 1%. This provided splitting of the valence band top by 40–60 meV and a degree of the spin polarization P of the electrons photoemitted as high as 80%. The films have a high quantum yield of photoemission Y upon activating to the negative electron affinity state due to the adsorption of Cs and O. Record values for the effective figure of merit P^2Y are achieved. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Anatoliĭ Vasil'evich Rzhanov was the initiator of scientific investigations and technological developments in the field of semiconductor cathodes with a negative electron affinity at the Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences. Only in this field does the interrelation of the atomic and electronic properties of the surface play an important role. Some recent advances in the development of photoemitters with an effective negative electron affinity (NEA photocathodes), which are used as sources of spin-polarized electrons, are described in this paper.

For several decades, the growth of lattice-matched heterostructures was the major focus of all epitaxial technology in multilayer semiconductor heterostructures. Among a variety of heterostructures, those based on the quaternary InGaAsP semiconductor solid solutions occupy a particular place. The sources and detectors of optical radiation for fiber links operating in the near-infrared region at the wavelength of 1.3–1.5 μm have been developed on the basis of this material. Wide-gap InGaAsP solid solutions also found application in the development of lasers and light-emitting diodes for the visible region of the spectrum. For this reason, it is not surprising that a great body of information on physical properties and methods of growth has been accumulated for this quaternary system. All conventional epitaxial techniques were successfully used for the synthesis of lattice-matched heterostructures in the InGaAsP/GaAs and InGaAsP/InP systems. This is true for gaseous-phase epitaxy with the use of chlorides of the Group III metals and hydrides of Group V elements [1], for molecular-beam epitaxy [2], for metallo-organic chemical vapor deposition (MOCVD) [3], and of course for liquid-phase epitaxy (LPE) [4].

In recent years, however, the problem of obtaining heterostructures with the lattice-mismatched film and substrate appeared in connection with the development of the sources of spin-polarized electrons [5]. The sources mentioned are based on the use of the phenomenon of negative electron affinity (NEA), which is achievable due to the adsorption of Cs and O on the atomically clean surface of GaAs and other semiconductors. These sources are widely used not only in high-energy physics but also in atomic physics, as well as in solid-state physics [6–8]. Due to the degeneracy of the valence band at $k = 0$, the theoretical degree of spin polarization of electrons photoemitted for GaAs and other III–V compounds with the zinc-blende structure is 50%. This limit may be exceeded using the strained pseudomorphic films grown on lattice-mismatched substrates. Strains relieve the degeneracy of a valence band. Basically, this permits the degree of spin polarization to be obtained at a rate as high as 100%. In recent years, a considerable increase in the degree of spin polarization was obtained by using strained photocathodes [8–10]. However, the device parameters are still far from optimal. Specifically, high degrees of spin polarization are achievable for structures with a low quantum yield ($Y \approx (1–2) \times 10^{-3}$), which is related to the small thickness of strained films. In a high-quality photocathode, the strain of the active layer should be as

large as possible; along with this, the active layer should be as thick as possible. In addition, dislocations should be absent, the surface should be smooth, and the thickness and doping should be uniform over an area of ~1 cm^2. In order to use fixed-frequency lasers for excitation, it is also desirable to have the ability to vary the band gap of the active layer. Taking into account the totality of the requirements mentioned above, searching for new strained heterostructures for the development of highly efficient NEA photocathodes, and developing sources of spin-polarized electrons with improved characteristics on the basis of these heterostructures constitute an important problem in physics and technology. The resolution of this issue is closely related to various fundamental problems of semiconductor heteroepitaxy. These are the problems of the critical thickness of the heteroepitaxial film, the stability of the morphology of strained layers, and the influence of strains and a substrate orientation on the phase equilibria in multicomponent systems.

In this paper, the growth of strained InGaAsP layers on the GaAs substrates, the results of the investigation into their electronic properties, and the application of the films obtained as sources of spin-polarized electrons (SSPE) are described. The sources were tested in cooperation with the Institute of Physics of the Johannes Gutenberg University (Mainz, Germany), and the Institute of Nuclear and High-Energy Physics (Amsterdam, Holland).

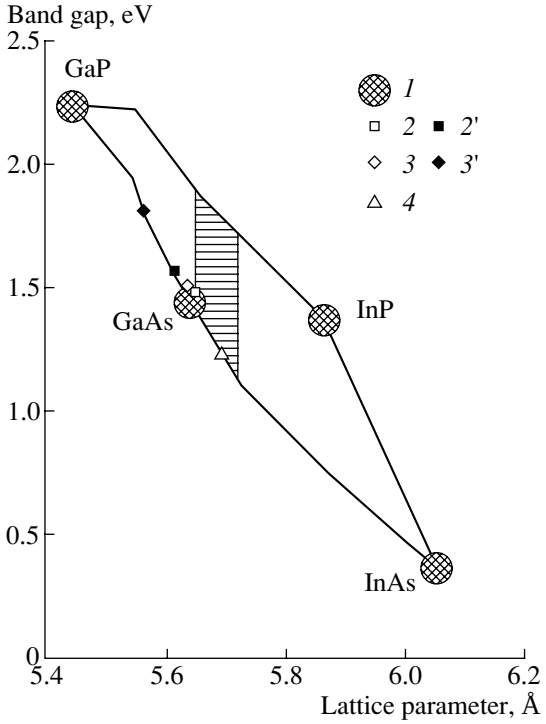


Fig. 1. Band gap for the quaternary InGaAsP solid solution as a function of the lattice parameter. (*1*) Binary compounds; the lines connecting them correspond to the relevant solid solutions. (*2–4*) Experimental data from publications: (*2*) from [8], (*3*) from [16], and (*4*) from [17]. The symbols *2*', *3*' indicate solid solutions considered in this study. The hatched area corresponds to the InGaAsP/GaAs quaternary alloys considered in this paper.

2. CHOICE OF HETEROSTRUCTURES AND SYNTHESIS TECHNIQUES

We selected the technique of the synthesis of heterostructures based on a long-standing experience accumulated with respect to the development of the LPE of III–V semiconductor compounds at the Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences. LPE was originally used for the growth of the quaternary InGaAsP solid solution more than 25 years ago [11]. Compared to other epitaxial techniques, LPE is comparatively simple and relatively cheap. It will be demonstrated below that LPE permits the strained InGaAsP/GaAs heterostructures to be obtained, which are not only highly competitive with the structures grown using other methods but also exceed them by the totality of required properties.

As was noted above, the essential factor which permits an increase in the yield of spin-polarized electrons is the relief of the degeneracy of the valence band at the center of the Brillouin zone. The elastic uniaxial deformation is only one of the methods for solving this problem. Among other methods, the use of superlattices, for which degeneracy is relieved due to quantum-well effects, may be mentioned [12, 13]. For the same purpose, the effect of spontaneous ordering in InGaP layers grown by MOCVD may be also used [14, 15]. The advantage of semiconductor solid solutions is the possibility of varying their band parameters or, more precisely, all their physical properties, by varying the composition. The variations in the band gap E_g for the $In_xGa_{1-x}As_yP_{1-y}$ solid solution over the entire range of the variation of the lattice parameter a is shown in Fig. 1. As can be seen from Fig. 1, the quaternary system encompasses a wide range of $E_g(x, y)$ and $a(x, y)$, and from this point of view, is highly advantageous for SSPE design. In order to obtain high degrees of spin polarization, it is necessary to provide the compressive strain in the plane of the active layer. Consequently, the native lattice parameter for the photoemitting layer should be larger than that for the substrate. With this purpose, the intermediate $GaAs_{1-x}P_x$ buffer layer, on which the active *p*-GaAs layer was formed, was grown on the GaAs substrate [8, 9, 16]. The variation in the composition over the buffer layer thickness from 0 (at the substrate) to x (at the interface with the active layer) and related dislocation generation were observed. These phenomena led to the compression of the active layer in the growth plane and to tetragonal distortions in the layer along the normal direction. In Fig. 1, symbol *2*' in the line of the ternary GaAsP solid solution indicates the final composition of the buffer layer used in study [8]; symbol *2* in Fig. 1 corresponds to the strained GaAs film grown on this buffer layer. It was

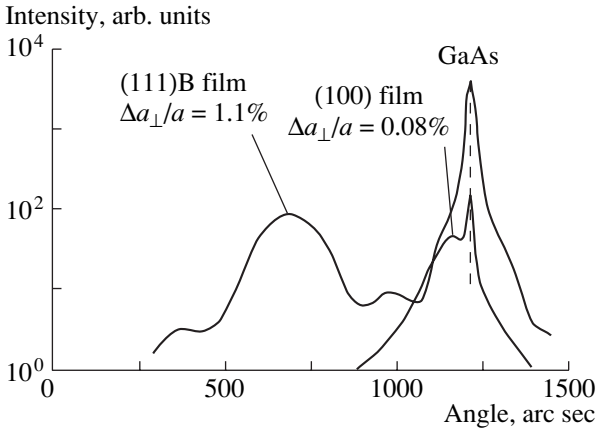


Fig. 2. X-ray rocking curves for two InGaAsP films simultaneously grown on the GaAs(100) and (111)B substrates. Melt composition: P, 0.46% Ga, 4%; and As, 11%; and the rest is In.

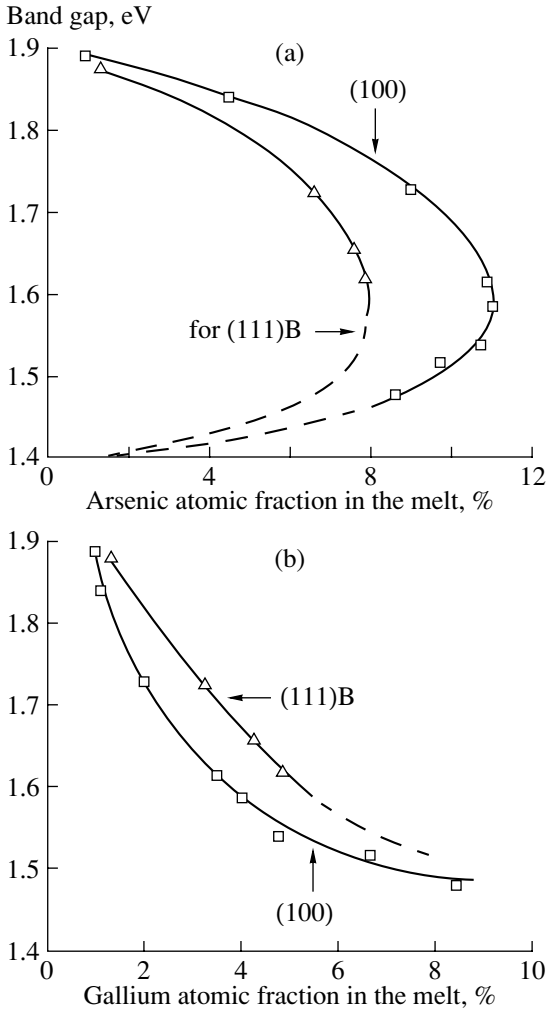


Fig. 3. Band gap for the InGaAsP films lattice-matched to GaAs(100) and (111)B in relation to the (a) As and (b) Ga content in the melt.

reported that the degree of spin polarization for this structure is as high as 90% [8, 9]. Symbol 3 in Fig. 1 demonstrates the experimental results reported in [16], in which case the $GaAs_{1-x}P_y$ active layer grown on the GaAs substrate with the $GaAs_{1-x}P_x$ buffer layer was used. It can be seen that the possibility of varying the band gap in such structures is very limited. Another disadvantage of the photocathode structures described is the impossibility of obtaining layers with a sufficient thickness and a high level of elastic strains. This is related to the fact that the density of threading dislocations, which are inherited by the active layer and inevitably cause its plastic relaxation, is high in the buffer layers. For this reason, the direct growth of the lattice-mismatched layer on the surface is more favorable. Thus, the pseudomorphic $In_xGa_{1-x}As$ film was grown directly on the GaAs substrate and was used as the active layer (Fig. 1, curve *4*) [17]. Such an approach permits early plastic relaxation to be prevented. However, a decrease in the band gap is accompanied by a decrease in the quantum yield of photoemission.

We previously suggested using pseudomorphic layers of the wide-gap ternary $In_xGa_{1-x}P$ and quaternary $In_xGa_{1-x}As_yP_{1-y}$ solid solutions grown on GaAs substrates as NEA photoemitters with a high quantum yield [18, 19]. The highly strained films were grown by LPE [20–23]. The hatched region in Fig. 1 demonstrates the range of $In_xGa_{1-x}As_yP_{1-y}$ compositions suitable for the development of strained photoemitters. Selecting the necessary composition, we were able to provide the lattice-mismatched layers with the band gap from 1.4 to 1.9 eV. The influence of elastic strains on the InGaAsP band structure was investigated in detail. As a result, the stable NEA photocathodes with a high quantum yield and a high degree of spin polarization of photoemitted electrons were developed [24].

3. LIQUID-PHASE EPITAXY OF ELASTICALLY STRAINED InGaAsP LAYERS

The major technological conditions for LPE and the equipment used were described in detail previously [23, 25]. For this reason, they are described below only briefly. The InGaAsP films were grown by a step-cooling technique ($\Delta T = 10$ K) at 1043 K on the GaAs(100), (111)A, and (111)B substrates. Elastic strains in these films were determined from measurements of the lattice parameter along the growth direction $\Delta a_\perp/a_s = (a_f - a_s)/a_s$ using a double-crystal X-ray diffractometer (CuK_α radiation). The reflections (004) and (111) were conventionally used for the (100) and (111) films, respectively. In order to reveal the possible plastic relaxation in the layers, X-ray rocking curves were recorded for both symmetric (111) and asymmetric (133), (422), and (115) GaAs reflections. The layer thickness was determined using an optical microscope on the cross-sectional cleavages of the structures. For thin films (<0.5 μm), an electron microscope was used.

If the structural quality of thin strained layers is high, the interference of X-ray waves is observed in them; it may also be used for calculating the thickness.

It was found in our experiments that the InGaAsP films grown in identical conditions on the (100) and (111)B substrates have substantially different compositions, i.e., different lattice parameters and band gaps. The aforesaid is illustrated by Fig. 2, which demonstrates X-ray rocking curves for two films simultaneously grown on the GaAs(100) and GaAs(111)B substrates, as well as by Fig. 3, which represents the dependence of E_g of the layers on the composition of the liquid phase.

It is known that the coherent growth of lattice-mismatched layers is possible up to a definite thickness only, which is called the critical thickness of the heteroepitaxial film. On exceeding the thickness mentioned, misfit dislocations are introduced into the film and the level of elastic strains is decreased in the structure. In this case, the structural quality of the films is abruptly lowered, which manifests itself as a broadening of the peaks in X-ray rocking curves and as the disappearance of interference phenomena (Figs. 4 and 5). Such layers are unsuitable for photocathodes.

Systematic investigations of the layers by optical and electron microscopy, as well as by X-ray diffractometry, allowed us to find the synthesis conditions for InGaAsP pseudomorphic films, which allowed heterostructures to be obtained with a record level of elastic strains at a thickness of 0.1–0.2 μm and band gap of 1.4–1.9 eV. The dependence of the critical thickness of the grown films on the lattice mismatch is shown in Fig. 6. The theoretical curve for the critical thickness of the heteroepitaxial film calculated from the Matthews–Blackeslee model is also shown here [26]. It can be seen that the experimental results may substantially exceed the theoretical predictions of the model. It is known that the critical thickness for the onset of introducing the misfit dislocations into the growing strained layer may be exceeded. In this case, certain conditions which are favorable for the inhibition of nucleation and propagation of misfit dislocations should be met. This problem is considered in more detail in our previous study [23].

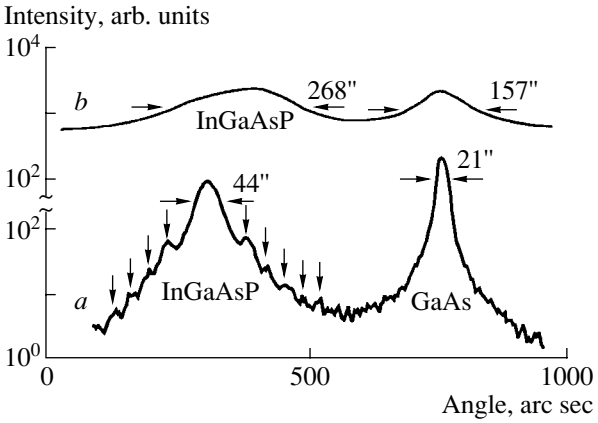


Fig. 4. X-ray rocking curves for two strained InGaAsP films ($\Delta a/a_\perp = 0.7\%$) of the same composition but different thickness h equal to (*a*) 0.47 and (*b*) 2.1 μm. The full width at half-maximum (FWHM) values are given.

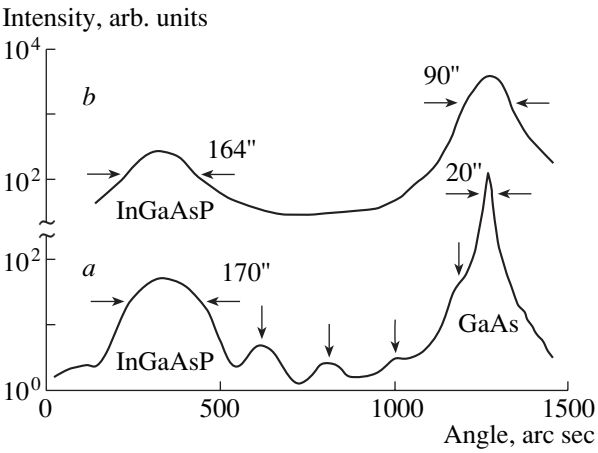


Fig. 5. X-ray rocking curves for two highly strained InGaAsP films ($\Delta a/a_\perp = 1.75\%$) of the same composition but different thickness h equal to (*a*) 0.09 and (*b*) 0.25 μm. The full width at half-maximum (FWHM) values are given.

4. ELECTRONIC PROPERTIES OF STRAINED LAYERS

The effects of elastic strains on the band structure and thresholds of fundamental optical transitions are shown in Fig. 7. A free unstrained layer with the lattice parameter a_0 is schematically depicted in the upper part of Fig. 7a. The same layer lattice-matched to the substrate with a smaller lattice parameter $a_s < a_0$ is depicted below. Due to the difference $\Delta a = a_0 - a_s$, the pseudomorphic film is compressed in the layer plane and stretched along the growth direction. The modification of the band structure related to these strains is shown in Fig. 7b. For unstrained semiconductors with a sphalerite structure, the maximum of the valence band located in the Brillouin zone comprises fourfold-degenerate multiplet $P_{3/2}$ with $J = 3/2$ and $m_J = \pm 3/2$ (subband of heavy holes) and $J = 3/2$ and $m_J = \pm 1/2$ (subband of light holes). The elastic strain leads to a total shift of all bands due to the hydrostatic (isotropic) component. It also splits the initially degenerate subbands into two subbands each with the components of the angular momentum $m_J = \pm 3/2$ and $m_J = \pm 1/2$ due to the shear component of deformation. Optical absorption spectra for the layers and first derivatives of these spectra are schematically shown in Fig. 7c.

To determine the band gap splitting and valence band splitting experimentally, we used photocurrent and photoluminescence (PL) spectroscopy. The photocurrent spectra were measured on the semiconductor–

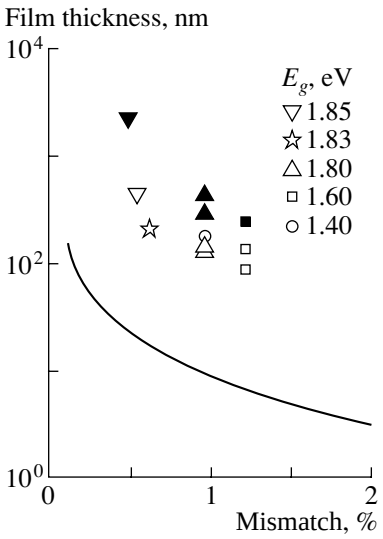


Fig. 6. Critical thickness of InGaAsP epilayers as a function of the lattice mismatch. The line corresponds to values calculated from the model [26]. Blank symbols are related to the pseudomorphic films with a negligibly small plastic relaxation, and black symbols are related to the films with a plastic deformation of 0.01–0.02%. The band gap is given for each symbol in the inset.

electrolyte contact at room temperature [27]. The PL spectra were excited using an argon laser (λ = 488 nm). The excitation power density was no higher than 20–40 W/cm^2. The spectra were measured using a double monochromator and a conventional system for photon counting. The energy of the PL peak was taken as the band gap E_g. The E_g values obtained from the photocurrent and PL spectra differed by no more than 5 meV.

The photocurrent spectrum $I_{ph}(E)$ and its first derivative $dI_{ph}(E)/dE$ are shown in Fig. 8. The derivative was determined by numerical differentiation of the measured spectrum. The derivative spectrum includes two clearly distinguishable peaks, which correspond to two split subbands of the valence band. The strain dependence of splitting for several InGaP(111) films is shown in Fig. 9. As can be seen from Fig. 9, splitting depends nonlinearly on the strain and does not exceed 40 meV. The nonlinearity of deformational splitting is related to the small magnitude of spin–orbit splitting of the valence band of InGaAsP, which is in turn determined by spin–orbit splitting for the Group V nonmetal atom. It is known that this magnitude for atoms is proportional to the squared nuclear charge (z^2). For this rea-

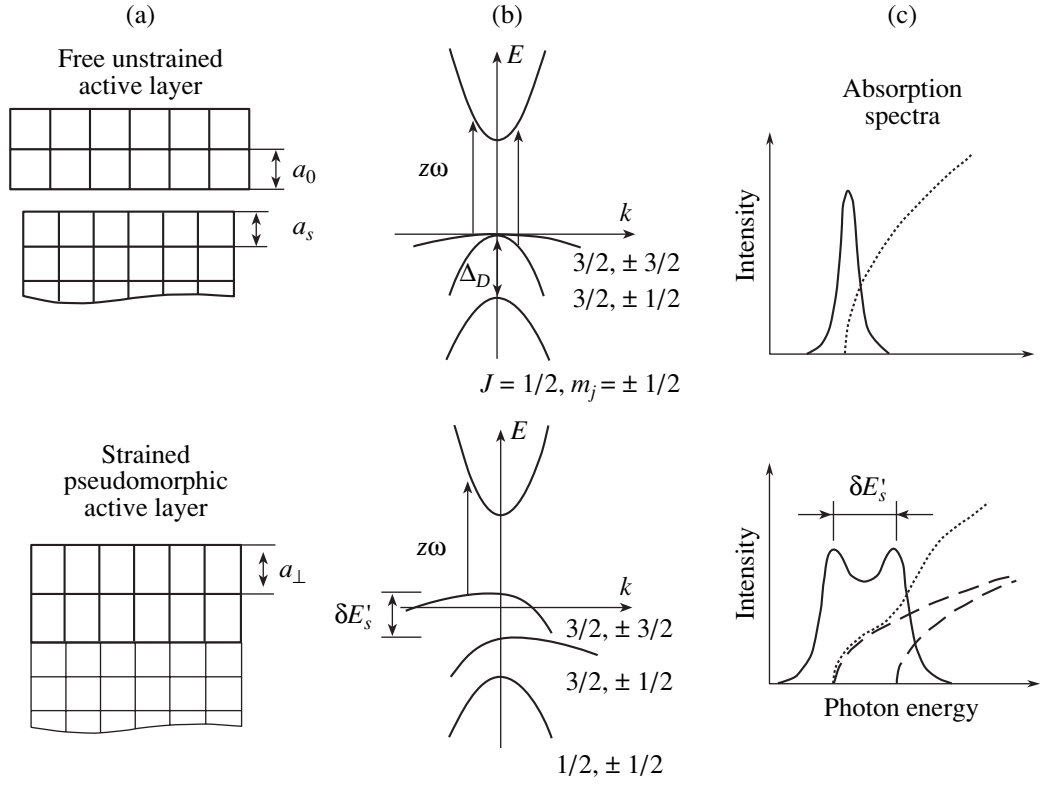


Fig. 7. (a) Schematic representation of strain effects in heteroepitaxial semiconductor layers with the sphalerite structure, (b) modification of the band structure related to strains, and (c) optical absorption spectra (dotted lines) and their derivatives (solid lines).

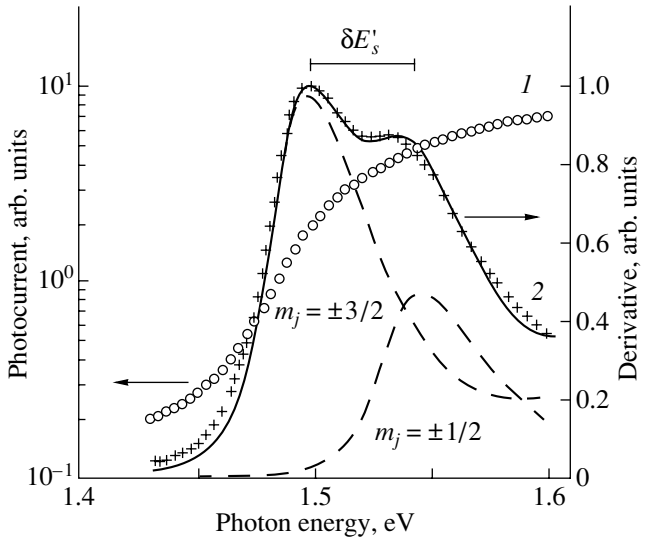


Fig. 8. (*1*) Photocurrent spectrum and (*2*) its derivative measured for the pseudomorphic InGaAsP film. Dashed lines represent the contribution of subbands splitted with $m_J = \pm 1/2$ and $m_J = \pm 3/2$ to the derivative.

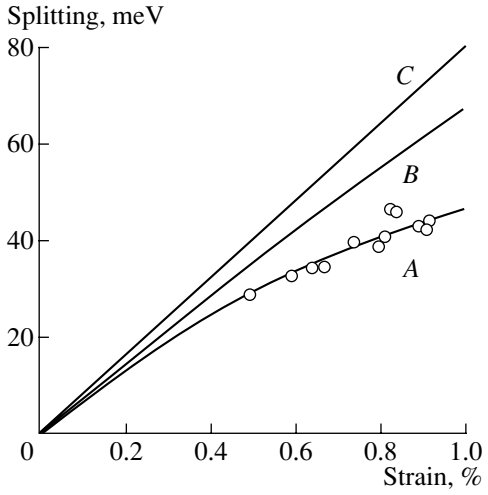


Fig. 9. Valence band splitting as a function of strain. Symbols correspond to the experimental results. Curves *A* and *B* represent the theoretical calculation for InGaP(111) and GaAs(111), respectively; and curve *C* corresponds to the calculation for GaAs to a linear approximation.

son, going from the ternary InGaP solid solution (Z_P = 15) to the quaternary InGaAsP solid solution (Z_{As} = 33) provides an increase in the deformational splitting of the valence band. Figure 10 demonstrates that this effect really exists.

5. PREPARATION OF AN ATOMICALLY CLEAN SURFACE AND ACTIVATION TO THE STATE WITH THE NEGATIVE ELECTRON AFFINITY

The behavior of the surface of the quaternary InGaAsP solid solution under heating in an ultrahigh vacuum is considerably complicated by the presence of two Group V elements with differing volatilities. In contrast to binary semiconductors, the data on the methods of cleaning the InGaAsP surface are actually lacking in the available publications. We used the following procedure of chemical treatment of the surface prior to heating in vacuum. First, the samples were rinsed in boiling isopropyl alcohol and then were placed into a hermetic box with an N_2 atmosphere, where the surface oxides were removed in the isopropanol solution of HCl. After that, the samples were mounted on a cermet holder and were transferred to an ESCALAB Mk-II photoelectron spectrometer. The samples were heated by a constant current. The surface composition subsequent to chemical treatment and vacuum heating was determined from X-ray photoelectron spectra excited by the AlK_α line with the energy of 1486.6 eV. The energy resolution of the spectrometer was 1 eV. In order to increase the sensitivity with respect to surface layers, the spectra were measured at various angular positions of the detector relative to the normal to the surface.

After chemical treatment, we observed O and C on the InGaAsP surface; their amount was as large as half a monolayer. On heating, desorption of O from the surface started at $T \approx 400$°C, and was completed at $T \approx 600$°C. The residual amount of C was a 0.1 monolayer. In contrast to GaAs, no layer of elemental As is formed on the InGaAsP surface subsequent to chemical treatment. In the case of GaAs, As plays the role of a passivating coating and cleans the surface of O and C by heating at 400°C.

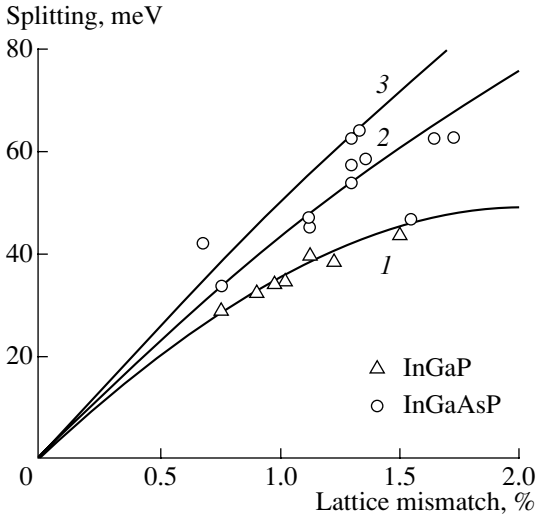


Fig. 10. Valence band splitting as a function of the lattice parameters along the growth direction for the InGaP and InGaAsP solid solutions. (*1*), (*2*), and (*3*) are the theoretical curves for InGaP, InGaAsP, and GaAs, respectively.

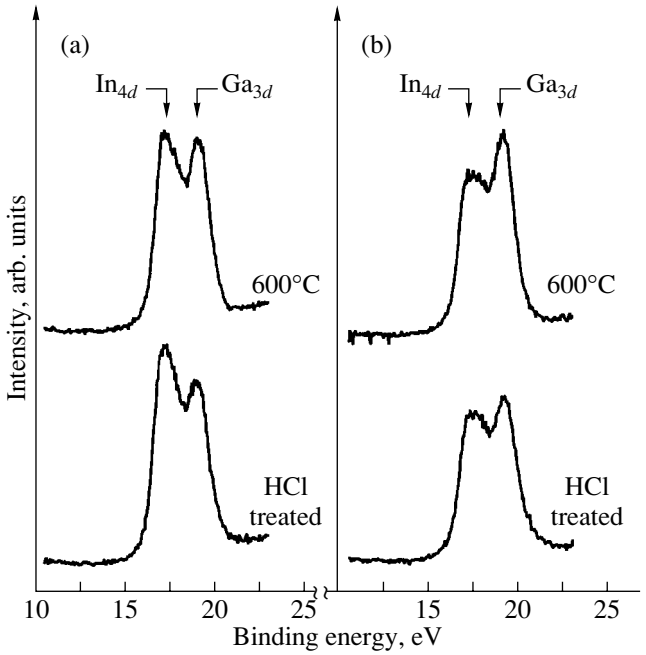


Fig. 11. X-ray photoelectron spectra of the InGaP(100) surface upon treatment in an alcohol solution of HCl (lower curves) and subsequent to heating in vacuum at 600°C for 10 min (upper curves). The spectra were measured for two angles of the electron escape: (a) 0° and (b) 75°.

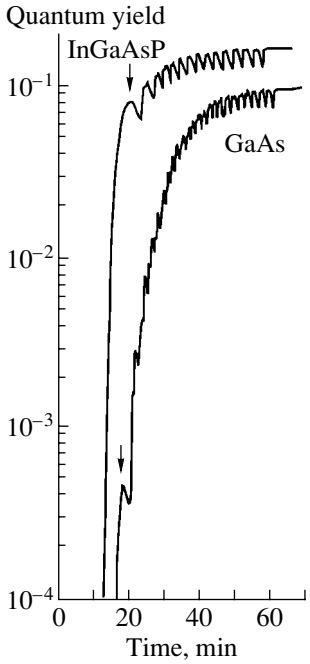


Fig. 12. Curves of activating of GaAs and InGaAsP to the state of negative electron affinity at coadsorption of Cs (continuous flux) and O (pulsed dosing).

As an example, let us consider the variation in the surface composition for the ternary InGaP solid solution. Figure 11 demonstrates the photoelectron spectra of the In_{4d} and Ga_{3d} levels recorded upon chemical treatment in HCl (lower curves in Fig. 11) and upon heating in a vacuum at 600°C for 10 min (upper curves in Fig. 11). The spectra were recorded for photoelectrons emitted along the normal to the surface ($\alpha = 0°$) and at an angle to the surface ($\alpha = 75°$), which corresponded to the escape depth of ~2.5 and ~0.6 nm, respectively. It can be seen from Fig. 11 that, upon chemical treatment, the measurement in the configuration sensitive to the bulk properties of the material yields a smaller amplitude for the Ga_{3d} peak compared to the In_{4d} peak. For the configuration sensitive to the composition of surface layers, the situation is inversed; i.e., the Ga peak is more intense. Upon annealing, the amplitude of the Ga_{3d} peak increases in relation to the In_{4d} peak. These data demonstrate that treatment of HCl in an alcohol solution brings about enrichment with Ga and depletion in In, and annealing enhances this trend.

The clean InGaAsP surface was activated to the state of the efficient negative electron affinity in a special chamber equipped with Cs and O sources with a residual pressure of 5×10^{-9} Pa. To avoid the degradation of photoemission properties of InGaAsP, which are related to the variation in the surface composition, a prolonged (1 h) annealing at a relatively low temperature (400°C) and rapid (1 min) heating to 600°C were used. Such a procedure provided reproducible high quantum yields for both GaAs and InGaAsP photocathodes.

The samples were activated by the adsorption of Cs and O. The cesium inlet was continuous, and Cs pressure in the chamber was no higher than 7×10^{-9} Pa, whereas the O inlet was pulsed ($P \lesssim 10^{-7}$ Pa). For measurements of the photoemission spectra, a halogen lamp and a monochromator were used. In experiments with large photocurrents, a continuous Ar laser with a luminous power as high as 100 mW in a spot 3 mm in diameter was used as an optical source.

For unstrained InGaAsP films ~1 μm thick, the quantum yield was 10–15% for the (111) orientation and 20–25% for the (100) orientation. Typical activation curves for InGaAsP(111)A surfaces with E_g = 1.89 eV and for GaAs are shown in Fig. 12. It can be seen that the first peak of the quantum yield Y_1 for InGaAsP amounts to about half of the final value, whereas for GaAs $Y_1 \approx 0.005 Y_{max}$. This demonstrates that, in contrast with GaAs, the NEA-state on the surface of a wide-gap solid solution may be obtained due to the adsorption of only Cs with no O. The spectral dependences of quantum yield for the GaAs and InGaP photocathodes are shown in Fig. 13.

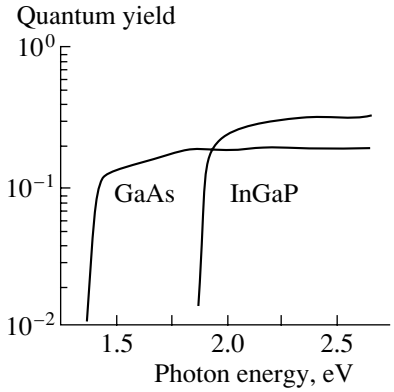


Fig. 13. Photoemission spectra of GaAs and $In_{0.49}Ga_{0.51}P$ activated surfaces. The threshold values of photocurrent correspond to the band gap E_g = 1.43 eV for GaAs and E_g = 1.92 eV for InGaP.

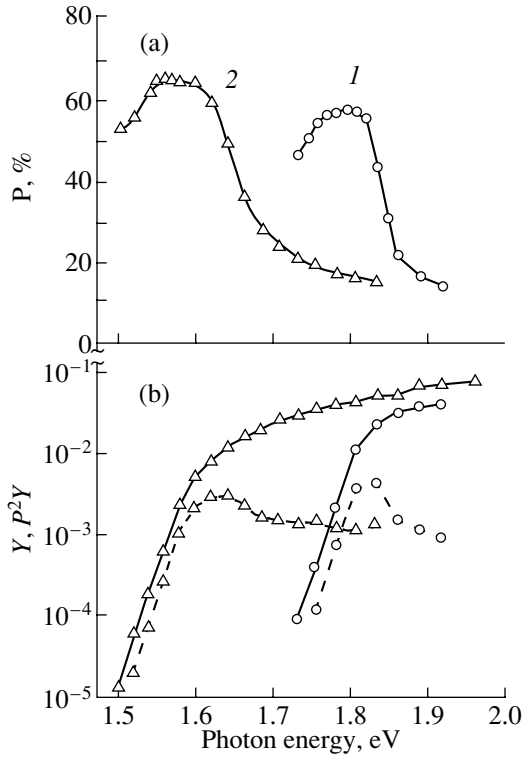


Fig. 14. (a) Spectral dependences of the degree of spin polarization P for photoelectrons emitted from the (*1*) InGaP and (*2*) InGaAsP surface. (b) Spectral dependences of the quantum yield Y (solid curves) and effective figure of merit P^2Y (dashed lines) for the same samples.

6. APPLICATION OF InGaAsP PHOTOCATHODES FOR THE SOURCES OF SPIN-POLARIZED ELECTRONS

Spin-polarized spectra were measured at the Institute of Physics of the Johannes Gutenberg University (Mainz, Germany), and at the Institute of Nuclear and High-Energy Physics (Amsterdam, Holland). The photocathodes were fabricated on the basis of pseudomorphic InGaAsP/GaAs heterostructures according to the technology described above. The spectra of spin polarization $P(E)$ were measured using a Mott polarimeter with an energy of 40 keV [24]. The systematic and random errors of $P(E)$ measurements were no larger than 3 and 1%, respectively.

For pseudomorphic highly strained $In_xGa_{1-x}As_yP_{1-y}$ films ($x \approx 0.35$, $y \approx 0.5$, and $\varepsilon \approx 1\%$), even the first measurements demonstrated that the largest values of P_{max} are in the range of 0.58–0.66, as can be seen from Fig. 14a. Thus, these values substantially exceed the theoretical limit for unstrained photocathodes. For numerous applications, the figure of merit P^2I for the beam of spin-polarized electrons is important [7]. Here, I is the beam current, which is proportional to the quantum yield of the photocathode and to the intensity of laser pumping. In the continuous mode, an increase in the current due to an increase in the pumping intensity is often restricted by the power of the laser used or by the heating of the photocathode. For this reason, it is convenient to use the effective figure of merit P^2Y to characterize the source. The P^2Y quantity is shown in Fig. 14b. It can be seen from Fig. 14b that the P^2Y magnitudes are as large as 4.7×10^{-3} and 3.3×10^{-3} for the photocathode with a wider and narrower band gap, respectively. We note that, as regards this parameter, the photocathodes obtained are highly competitive with the best results published in the literature [28, 13].

The further optimization of the film thickness, band gap, and elastic strain provided the development of the InGaAsP photocathodes with E_g = 1.67–1.72 eV. Being tested on an accelerator in the Institute of Nuclear and High-Energy Physics (Amsterdam), these photocathodes demonstrated a degree of spin polarization in the range of 65–80% [29]. The frequency-controlled $Ti:Al_2O_3$ laser generating pulses of spin-polarized electrons with a duration of 2 μs and repetition frequency of 1 Hz was used as a pumping source. For these photocathodes, the estimation of the P^2Y factor yields a value of about 10^{-2} at the photoemission current as high as 150 mA. The photocathode lifetime (i.e., the duration of operating at $I > 15$ mA and $P \gtrsim 65\%$) was 4–5 weeks.

7. CONCLUSION

Thus, the growth technology for strained pseudomorphic InGaAsP/GaAs and InGaP/GaAs heterostructures was developed, and their electronic properties were investigated. Based on these structures, the photoemission electron sources with a degree of spin polarization as high as 80% were developed. With respect to the high quantum yield of photoemission for wide-gap solid solutions, the integrated figure of merit P^2Y for the source of spin-polarized electrons was $(3–10) \times 10^{-3}$. Taking into account the possibility of tuning the band gap to a specified wavelength of a pumping laser and

the high stability of the photocathodes, we may conclude that the InGaAsP/GaAs heterostructures are promising for the development of highly efficient sources of spin-polarized electrons.

REFERENCES

1. G. H. Olsen and T. J. Zamerowski, Prog. Cryst. Growth Charact. **2**, 309 (1979).
2. M. B. Panich, Prog. Cryst. Growth Charact. **12**, 1 (1986).
3. M. Razeghi, *The MOCVD Challenge* (Inst. of Physics Publ., Bristol, 1995), Vol. 2.
4. K. Nakajima, in *GaInAsP Alloy Semiconductors*, Ed. by T. P. Pearsall (Wiley, New York, 1982), p. 43.
5. R. Prepost and T. Maruyama, Annu. Rev. Nucl. Part. Sci. **45**, 41 (1995).
6. G. Lampel and C. Weisbuch, Solid State Commun. **16**, 877 (1975).
7. D. T. Pierce, R. J. Celotta, G. G. Wang, *et al.*, Rev. Sci. Instrum. **51**, 478 (1980).
8. T. Nakanishi, H. Aoyagi, H. Horinaka, *et al.*, Phys. Lett. A **158**, 345 (1991).
9. R. Alley, H. Aoyagi, J. Clendenin, *et al.*, Nucl. Instrum. Methods Phys. Res. A **365**, 1 (1995).
10. Yu. A. Mamaev, A. V. Subashiev, Yu. P. Yashin, *et al.*, Solid State Commun. **114**, 401 (2000).
11. G. A. Antypas and R. L. Moon, J. Electrochem. Soc. **120**, 1574 (1973).
12. T. Omori, Y. Kurihara, T. Nakanishi, *et al.*, Phys. Rev. Lett. **67**, 3294 (1991).
13. Y. Kurihara, T. Omori, Y. Takeuchi, *et al.*, KEK Preprint 94-59; SLAC-PUB-6530 (1994).
14. A. Gomio, T. Suzuki, and S. Iijima, Phys. Rev. Lett. **60**, 2645 (1988).
15. S. H. Lee and G. B. Stringfellow, J. Appl. Phys. **83**, 3620 (1998).
16. Yu. A. Mamaev, Yu. A. Yashin, A. V. Subashiev, *et al.*, Phys. Low-Dimens. Struct. **27**, 7 (1994).
17. T. Maruyama, E. L. Garwin, R. Prepost, *et al.*, Phys. Rev. Lett. **66**, 2376 (1991).
18. V. L. Al'perovich, Yu. B. Bolkhovityanov, A. G. Paulish, and A. S. Terekhov, Pis'ma Zh. Tekh. Fiz. **18** (22), 67 (1992) [Sov. Tech. Phys. Lett. **18**, 751 (1992)].
19. V. L. Alperovich, Yu. B. Bolkhovityanov, A. G. Paulish, and A. S. Terekhov, Nucl. Instrum. Methods Phys. Res. A **340**, 429 (1994).
20. Yu. B. Bolkhovityanov, V. L. Alperovich, A. S. Jaroshevich, *et al.*, J. Cryst. Growth **146**, 310 (1995).
21. Yu. B. Bolkhovityanov, A. M. Gilinsky, N. V. Nomerotsky, *et al.*, J. Cryst. Growth **149**, 17 (1995).
22. V. L. Alperovich, Yu. B. Bolkhovityanov, A. S. Jaroshevich, *et al.*, J. Appl. Phys. **82**, 1214 (1997).
23. V. L. Alperovich, Yu. B. Bolkhovityanov, S. I. Chikichev, A. S. Jaroshevich, A. G. Paulish, and A. S. Terekhov, in *Indium Phosphide and Related Compounds: Materials, Applications and Devices*, Ed. by M. O. Manasreh (Singapore, 2000), Vol. 9, p. 651.
24. P. Drescher, S. Pluetzer, E. Reichert, *et al.*, Nucl. Instrum. Methods Phys. Res. A **381**, 169 (1996).
25. Yu. B. Bolkhovityanov, R. I. Bolkhovityanova, and S. I. Chikichev, J. Electron. Mater. **12**, 525 (1983).
26. J. W. Matthews and A. E. Blakeslee, J. Cryst. Growth **27**, 118 (1974).
27. Yu. B. Bolkhovityanov, F. S. Jaroshevich, M. A. Revenko, *et al.*, Semicond. Sci. Technol. **11**, 1847 (1996).
28. T. Saka, T. Kato, T. Nakanishi, *et al.*, Jpn. J. Appl. Phys. **32**, L1837 (1993).
29. Yu. B. Bolkhovityanov, A. M. Gilinsky, C. W. de Jager, *et al.*, in *Proceedings of the 12th International Symposium on High–Energy Spin Physics, Amsterdam, 1997*, p. 700.

Translated by N. Korovin

Semiconductors, Vol. 35, No. 9, 2001, pp. 1063–1071. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1111–1119.

SEMICONDUCTOR STRUCTURES, INTERFACES, AND SURFACES

Electronic Properties of InAs-Based Metal–Insulator–Semiconductor (MIS) Structures

G. L. Kuryshev*, A. P. Kovchavtsev, and N. A. Valisheva

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

**e-mail: kur@isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—Special features of electronic processes in InAs-based MIS structures operating in the charge injection mode were investigated. These structures are used as photodetectors in the spectral range of 2.5–3.05 μm. A double-layer system consisting of an anodic oxide layer and a low-temperature silicon dioxide layer was used as an insulator. It was shown that fluorine-containing components, which were introduced into electrolyte, reduced the value of the built-in charge and the surface-state density to minimal measurable values of $\lesssim 2 \times 10^{10}$ cm^{-2} eV^{-1}. Physical and chemical characteristics of the surface states at the InAs–insulator interface and the possible causes of their absence were discussed on the basis of the phase composition data of anodic oxide obtained by X-ray photoelectron spectroscopy. An anomalous field generation was observed under the nonequilibrium depletion of the semiconductor. The processes of tunneling generation, which are important at large amplitudes of the depletion pulse, were considered. The noise behavior of MIS structures under a nonequilibrium depletion was investigated. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Physicotechnological Features of InAs-Based MIS Structures

InAs-based MIS structures form the basis for the development of the multielement infrared (IR) detector arrays, which have a detectivity limit of $\gtrsim 10^{12}$ cm Hz$^{1/2}$ W^{-1} in the wavelength range λ = 2.5–3.05 μm. The electronic processes at the semiconductor–insulator interface define the device characteristics of MIS structures [1].

The main task in the development of the InAs-based perfect MIS structures is to obtain semiconductor–insulator interfaces with predetermined and reproducible electrical parameters: a surface state density of $\lesssim 10^{11}$ cm^{-2} eV^{-1}; a relaxation time of the surface potential in the MIS structures at the nonequilibrium pulse depletion of >15–20 ms (at a temperature T = 150 K); and a charge density of slow states near the semiconductor–insulator interface, which are responsible for hysteresis in the capacitance–voltage characteristics, of $< 5 \times 10^{-8}$ C cm^{-2}.

All technological processes should be carried out at temperatures no higher than 200–250°C, because at higher temperatures the surface stoichiometry is disturbed due to the evaporation of the Group V element. The insulator properties of SiO_2 layers grown at these temperatures are inferior to those of a high-temperature silicon dioxide (450–700°C). In this case, the electrical parameters of MIS structures are poorly reproducible. This fact is usually related to a native oxide layer at the surface of the semiconductor substrate. It is assumed that the transition layer provides a high surface-state density. At an atomically clean surface, obtained by cleaving in an ultrahigh vacuum, the density of states is low but the oxygen adsorption at the surface forms an interface with a high density of states [2].

Anodic oxidation of the InAs surface makes it possible to obtain satisfactory properties of the interface [3]. However, properties of the anodic oxide (AO) layer do not satisfy the insulator strength requirements. They also have a low mechanical and chemical strength, which makes it difficult to fabricate structures by planar technology. A drawback of the anodic oxides consists in the injection instability of MIS structures in electric fields, which is related to small heights of the potential barriers for electrons and holes at the boundary with the semiconductor because the insulating film consists of a mixture of In_2O_3 oxide with a band gap of 2.8–3.75 eV [4, 5] and As_2O_3 oxide with a band gap of about 4 eV [6]. According to [7–10], the oxide layers at the InAs surface have a complicated composition and, depending on the production conditions, they consist of elemental arsenic as well as indium and In_2O_3, As_2O_3, $InAsO_4$, $In(AsO_3)_3$, As_2O_4, and As_2O_5 oxide mixtures with the common nonstoichiometric composition.

The double-layer insulator system AO–SiO_2 [11] makes it possible to develop MIS structures useful for fabricating device structures. Varying the composition of the oxide agents enables the electrical parameters of the InAs–SiO_2 interface to be changed [12–14].

MIS structures were fabricated on the InAs autoepitaxial structures of ÉSAI-40-35-10 or ÉSAI-92-35-10 Yae 0.040 TU types. The *n*-type epilayers grown on

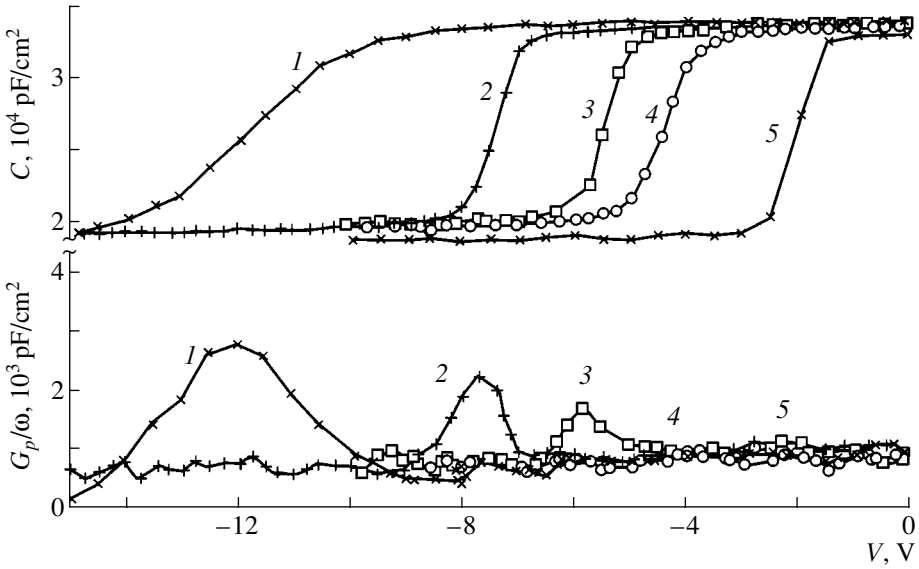


Fig. 1. Capacitance–voltage $C(V)$ characteristic and the normalized conductance G_p/ω in relation to the voltage V at a frequency of 50 kHz. (*1*) Oxidation without fluorine, (*2–5*) NH_4F is introduced into the electrolyte in amounts of 0.044, 0.22, 0.44, and 2.17 vol %, respectively.

heavily doped n^{++} substrates (~10^{18} cm^{-3}) with the (111)A orientation had the following parameters: the concentration of the majority charge carriers (electrons) was $(1–6) \times 10^{15}$ cm^{-3} and the lifetime in the layers was 0.3–1.8 μs. Silicon dioxide layers 140 nm thick obtained by the oxidation of monosilane in an oxygen atmosphere in a reactor with a lowered pressure (RLP) at T = 200–240°C were used as the gate insulator. The RLP-SiO_2 layers contain a small amount of adsorbed water (<2 vol %), and the breakdown fields of the layers are $>6 \times 10^6$ V cm^{-1}. Doped In_2O_3 layers 120 nm thick obtained by the ion–plasma sputtering of the target containing 8% SnO_2 were used as the field electrodes.

EXPERIMENT

1. Chemical Preparation and Modification of the InAs Surface

To obtain a stable InAs–insulator interface with a minimal surface-state density, we studied different methods of the InAs substrate-surface treatment before deposition of the gate insulator: degreasing and etching in polishing etchants with the subsequent removal of the residual oxide layer in various solutions. We found that, after optimizing the technological conditions of the silicon dioxide layer deposition, the chemical treatment of the InAs substrate surface did not profoundly affect the parameters of the MIS structure. In this case, the surface state density (SSD) is about 1.2×10^{11} cm^{-2} eV^{-1}.

We obtained an efficient improvement of the InAs–insulator interface parameters if, before the insulator deposition, the InAs substrate surface was oxidized in an electrolyte which contained fluoride ions introduced into the electrolyte by adding the ammonium fluoride. Oxidation was performed at room temperature in a two-electrode cell in the galvanostatic mode at a current density of 0.5 mA cm^{-2}. For the electrolyte, we used a concentrated ammonia solution in ethylene glycol with a volume ratio of 1 : 5.

Successive variations in the capacitance–voltage characteristics ($C–V$) $C(V)$ and the normalized conductance dependences (G_p/ω) on voltage (V) with the fluoride-ion concentration in the electrolyte as the parameter are shown in Fig. 1. One can see that, as the fluoride-ion concentration increases, the built-in charge in the oxide of the MIS structure decreases (the $C–V$ characteristic is shifted to zero voltage), and SSD decreases (the slope of this characteristic increases and the amplitude of the conductance peak at an alternating signal decreases). The effect of fluorine on the capacitance frequency dispersion and small-signal conductance of MIS structures is illustrated in Fig. 2 [15]. Samples with the anodic oxide sublayer grown in the electrolyte not containing fluoride ions (Fig. 2a) are characterized by a large built-in charge and SSD. This circumstance manifests itself in the shift of the $C–V$ characteristic to negative biases (up to 15–20 V), the frequency dispersion of this characteristic, and in the appearance of a peak in the dependences of the normalized conductance on voltage. Estimation of SSD by the peak magnitude from the $G_p/\omega = f(V)$ curve yields $N_{ss} \approx 3 \times 10^{11}$ cm^{-2} eV^{-1}. In the samples with the anodic oxide sublayer (Fig. 2b) grown in the fluorine-containing electrolyte with an optimal fluorine concentration, the built-in charge decreases (the voltage corresponding to flat bands is approximately equal to –3V), the capacitance frequency dispersion is absent, and the peak amplitude in the normalized-conductance dependence on voltage

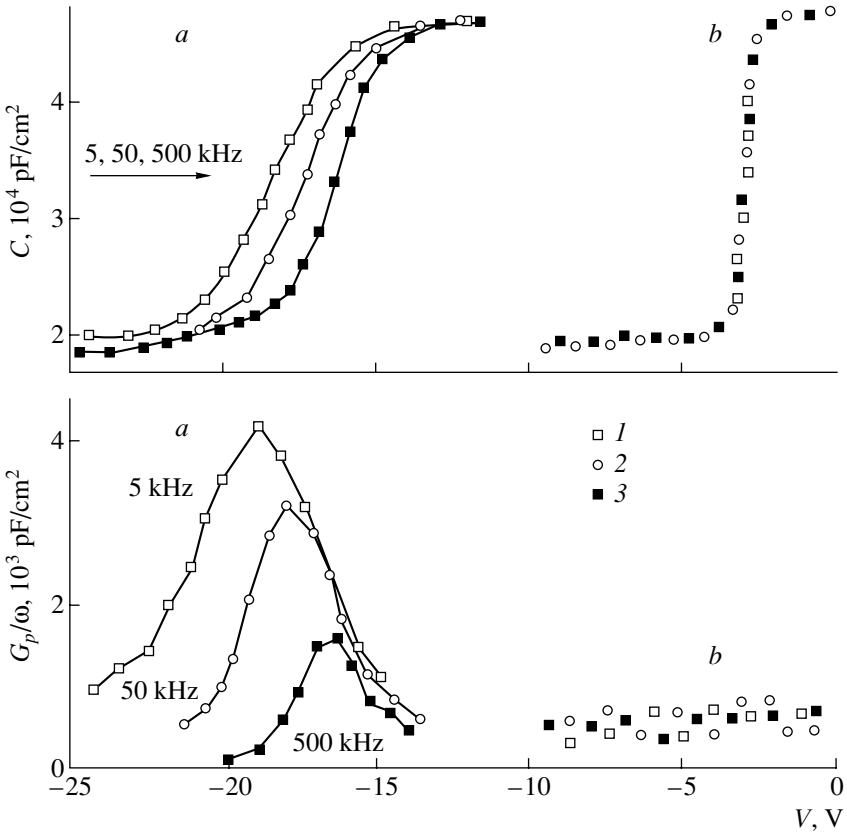


Fig. 2. The frequency dispersion of the capacitance (C) and normalized conductance (G_p/ω) of MIS structures fabricated with an anodic sublayer (*a*) without fluorine and (*b*) with the fluorine addition.

becomes smaller than the ultimate sensitivity of the measuring system. According to estimates, SSD is equal to $N_{ss} < 2 \times 10^{10}$ cm^{-2} eV^{-1}.

An analysis of X-ray photoelectron spectra showed that compositions of the anodic oxide layers with fluorine and without it are quite different [16]. The oxide layers without fluorine consist of a mixture of In_2O_3 and As_2O_3 (Fig. 3, curve *b*). Nonstoichiometric oxyfluoride of pentavalent arsenic and InF_3 are present in the fluorine-containing layers of the anodic films (Fig. 3, curve *c*). An analysis of the compound distribution with depth, which is part of the fluorinated anodic oxide layer, shows that the [In]/[As] concentration ratio dependences are similar to the distribution profile in AO without fluorine. They differ mostly in the surface layer. In the fluorinated layer, in contrast to AO without fluorine, the surface layers are not enriched with In compounds. Fluorine is found across the entire film thickness, with its concentration increasing as the oxide–InAs interface is approached. The highest ratio of the fluorine and oxygen concentrations [F]/[O] in the transition AO–InAs layer decreases from 0.47 to 0.2 as the oxide thickness decreases from 17 to 7 nm. Because the fluorine concentration increases with an increase in the growth time (i.e., with the film thickness increasing at constant current density), the fluorine accumulation at the interface is probably related to a high (compared to the oxygen ions) diffusion rate of fluorine ions in growing AO.

In [17], the features of the chemical bond in the molecules of oxides of Group V elements, with these oxides being formed at the III–V semiconductor surface, were considered. It follows from [17] that native oxide layers on the phosphides (InP and GaP) consist of phosphates of the Group III elements and are more stable than those on arsenides and antimonides due to the donor–acceptor pair P=O. In addition, this bond is multiple due to the $d\pi$–$p\pi$ bonding formed under the interaction of free d orbitals of the phosphorus atom with occupied p orbitals of the oxygen atom. Arsenic forms a native oxide As_2O_3 at the InAs and GaAs surfaces. The As_2O_3 atomic structure differs from the phosphate atomic structure on InP and GaP in the existence of unshared electronic pairs. This pair, according to [13], is responsible for the GaAs surface states whose density is equal to the number of As_2O_3 molecules at the

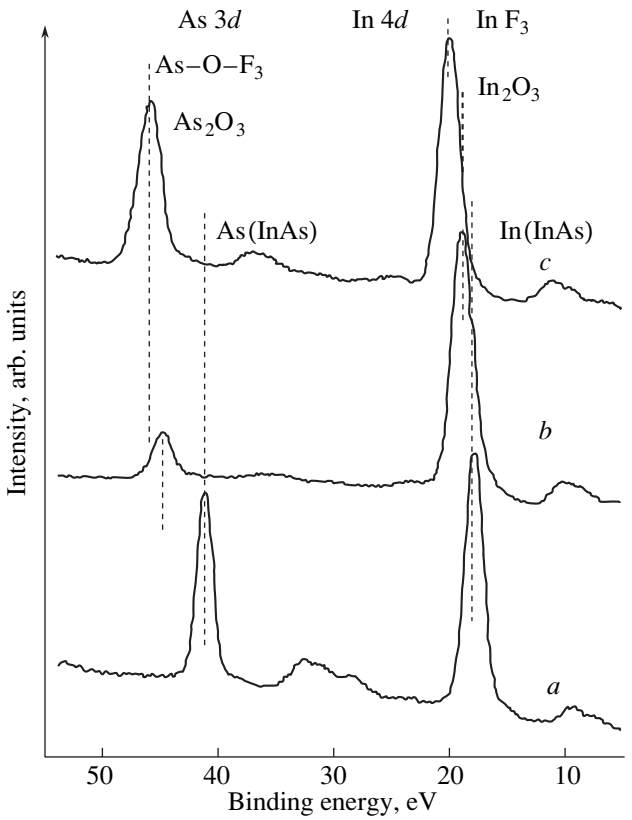


Fig. 3. X-ray photoelectron spectra of the InAs surface: (*a*) after ion treatment; (*b*) with the anodic oxide 18 nm thick, without fluorine; and (*c*) with the anodic oxide 7 nm thick, with fluorine.

interface. On adding fluorides into the oxidizing medium, arsenic is oxidized to a pentavalent state forming oxyfluoride $AsOF_3$. This process becomes possible due to the specific chemical properties of fluorine: it forms fluorides of elements in the highest oxidation degree. In this case, the source of the surface states (unbound occupied orbitals) is removed. Such a molecule has bonds in the nearest As surroundings that are similar to the bonds in the nearest P surroundings in the P_2O_5 compound. The structure of the X_3BO molecules (where X is the halide-ion and B is the Group V element) is a distorted tetrahedron, where three σ bondings B–X form a pyramidal group and the fourth forms a double bond with the oxygen atom As=O similar to the P=O bond.

It follows from the comparison of the nearest As surroundings in InAs and in its oxyfluoride that their configuration is identical; i.e., it is of a tetrahedron type. The types of bonds are also identical. An As atom in InAs is bonded with three In atoms via σ bonding and with the fourth atom via the donor–acceptor bond. The arsenic atom in the oxyfluoride also has four neighboring atoms with whom it is bonded by three σ-bonds (As–F) and one donor–acceptor bond (As=O). Then the molecules, isoelectronic and isostructural in the nearest surrounding to the compound of As with In in the semiconductor volume, are formed during oxidation in the fluorine-containing medium at the semiconductor surface. To put it differently, the compounds formed at the surface have bonds similar in type, energies, and space arrangement to the bonds in the bulk of the semiconductor. Due to this fact, the dangling bonds are removed from the interface. This results in a low SSD.

Thus, using the system (fluorinated anodic sublayer)–RLP-SiO_2, we succeeded in obtaining an InAs–insulator interface comparable to the Si–SiO_2 interface.

2. Tunneling Generation of the Charge Carriers

A band-to-band tunnel current in the field of the injection pulse flows from the inversion layer of the InAs valence band into the conduction band through the potential barrier formed by the space charge region (SCR) of the depletion semiconductor region. An additional (changing in time) component of dark current appears in SCR. In narrow-gap semiconductors, this component makes an appreciable contribution.

It is possible to calculate the tunnel current density through the band gap of the semiconductor under a nonequilibrium depletion in a one-dimensional quasi-classical approximation [18], when holes from the inversion layer tunnel through the depletion semiconductor region into a quasi-neutral bulk. The tunnel current is expressed by the formula

$$I_t = 4\pi m_p^* q/h^3 \int_0^\infty dE(f_1 - f_2)\int_0^E dE_\perp P_T, \quad (E, E_\perp), \quad (1)$$

where f_1 and f_2 are the Fermi–Dirac distribution function on the left and right (the metal is to the left), respectively; m_p^* is the hole effective mass; E is the total energy of the tunneling particle; and $E_\perp$ is the energy component perpendicular to the tunneling direction. We assume that it is equal to 0. The quantity P_T is the tunneling probability through the semiconductor band gap, which is given in the Wentzel–Kramers–Brillouin approximation by [19]

$$P_T = \exp\left(-2\int_{x_1}^{x_2} K_x(x)dx\right). \quad (2)$$

Here, x_1 and x_2 are classical turning points and $K_x(x)$ is the wave vector component in the tunneling direction in the classical forbidden region.

The Franz law [20] typically used for the dispersion relation in the potential barrier can be expressed as

$$K_x^2 = (2m_x^*/\hbar^2)\varphi(x)[1 - \varphi(x)/E_g], \quad (3)$$

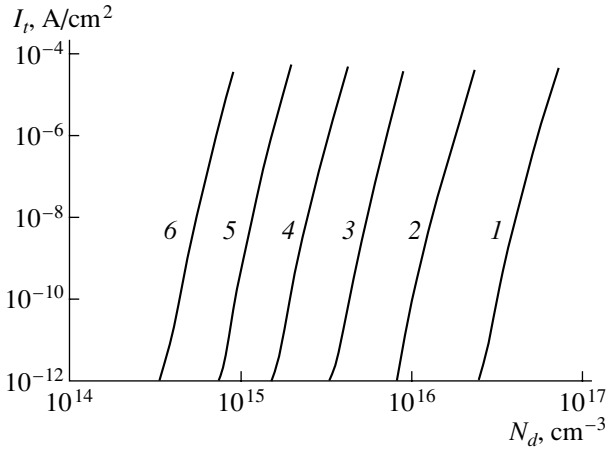


Fig. 4. Dependence of the tunnel band-to-band current density on concentration of the donor impurity at $T = 150$ K for different initial band bendings φ_{s0}: (*1*) 0.5, (*2*) 1, (*3*) 2, (*4*) 4, (*5*) 8, and (*6*) 10 V.

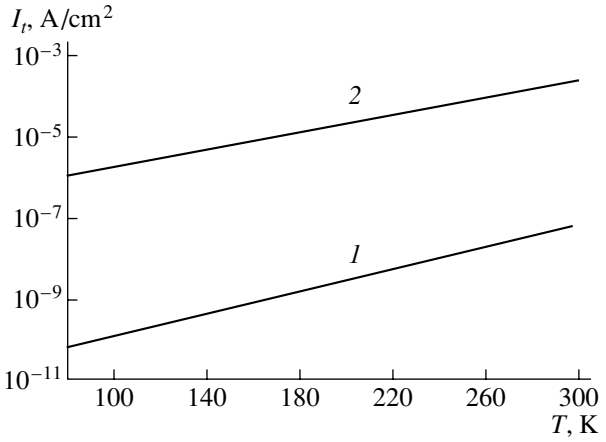


Fig. 5. The temperature dependence of the tunnel current density at φ_{s0}: (*1*) 3 and (*2*) 5 V. $N_d = 3 \times 10^{15}$ cm^{-3}.

where $\varphi(x)$ is the band bending in SCR and E_g is the band gap. The Franz law enables one to take into account the effect of the second energy band on the tunneling of the charge carrier through the band gap of the semiconductor. However, this law does not take into account the coordinate dependence of the tunnel effective mass (m_x^*) (the effective mass should change from that of electron m_e^* to that of hole m_p^* as a result of the carrier tunneling from the conduction band to the valence band). Matching the tunnel effective masses at the band gap edges, we use the expression in [21] to obtain

$$K_x^2 = (2m_e^* m_p^*/\hbar^2) \times \{[[E_g - \varphi(x)]\varphi(x)]/[m_e^*\varphi(x) + m_p^*[E_g - \varphi(x)]]\}. \quad (4)$$

If the electron and hole masses are equal to $m_e^* = m_p^* = m_x^*$, this relation is transformed into the Franz dispersion relation (3). We performed a numerical calculation of the tunnel current density by substituting relation (4) into (1) and taking into account the parabolic dependence of the band bending on the coordinate in the approximation of a uniform distribution of the doping impurity over the semiconductor depth.

The tunnel model was used in calculating the parameters of the photosensitive structures within the context of a simplified linear model. Let us consider the effect of the tunnel current on the parameters of the photodetector. The results of the tunnel current density I_t calculation for different injection pulse amplitudes, which cause different initial band bending φ_{s0}, in relation to the donor impurity concentration N_d and temperature T are shown in Figs. 4 and 5. The impact of the tunnel current becomes appreciable for the injection pulse amplitude $V_i > 6$ V, when I_t becomes comparable to the density of the generation and diffusion components; i.e., when $I_t > (10^{-5}$–$10^{-4})$ A cm^{-2}. The effect of the tunnel current on the time (t) dependence of the surface potential relaxation in the dark and under illumination (theory and experiment) is illustrated in Fig. 6. The arrow indicates the point corresponding to the band bending in which the tunnel component of the dark current becomes comparable to the current of thermal generation (~6 V). This point corresponds to the onset of the tunnel current switching-off due to its sharp dependence on the electric field strength in SCR. At a smaller band bending ($t > 1$ ms), the tunnel current is several orders of magnitude lower than the thermal generation current and, thus, does not exert a pronounced effect on the characteristics of actual photodetectors (PDs). The relaxation curve has a conventional shape (without a tunneling component) (curves *1*). For larger band bending ($t < 1$ ms) or for higher impurity concentrations, the tunnel current significantly exceeds the other thermal currents and affects the PD parameters.

A comparison of theory and experiment shows that the tunneling model explains satisfactorily the experimental results. Because the device tunnel current is stray, its contribution manifests itself in a noise increase at small collection times and in the attenuation of the photosignal due to a partial filling of the quantum well by the dark current carriers. If the tunnel current exists, the MIS structure relaxes most of the time at a smaller band bending in comparison with the MIS structure relaxation when the tunnel component is absent. These circumstances may lead to a decrease in the collection coefficient of photogenerated carriers, especially for small absorption coefficients because the higher the tunnel current density, the faster the MIS structure relaxes to the equilibrium state. In this case, a smaller charge of the photogenerated carriers will be accumulated in the potential well during the collection time. As the calculation showed, the signal can decrease by approximately 1.5 times for certain parameters if the

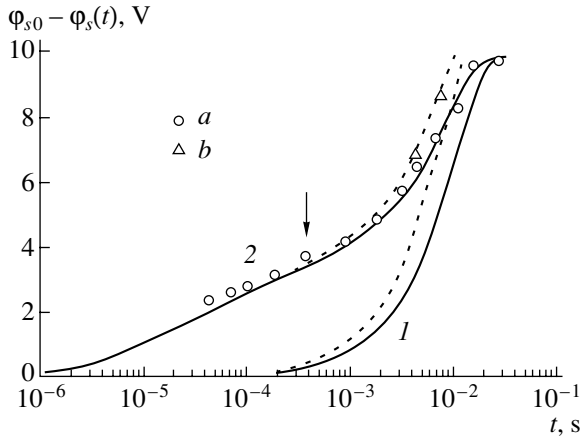


Fig. 6. The tunnel current effect on the relaxation curve of the surface potential. Parameters of calculation: $S = 3 \times 10^{-5}$ cm^2, $C_n = 50$ pF, $N_d = 3 \times 10^{15}$ cm^{-3}, the diffusion length of the minority carriers in the quasi-neutral InAs volume $L_p = 5$ μm, the absorption coefficient $\alpha = 5 \times 10^3$ cm^{-1} at the wavelength $\lambda = 3$ μm. The arrow indicates the point at which I_t is equal to the density of the thermal generation current. Solid curves are theoretical calculations with (*1*) disregarding tunneling and (*2*) considering tunneling. Dotted lines are the calculation for the luminous power of 10^{-5} W cm^{-2} ($\lambda = 3$ μm). Dots are the experimental results (a) in the dark and (b) under illumination.

current of photogenerated carriers is much lower than the tunnel current density (small-signal approximation). The tunnel current also enhances the noise in MIS structures for small collection times because, during the same time, the concentration of the collected particles is much higher than when the tunnel current is absent.

3. Noise in MIS Structures at Nonequilibrium Pulse Depletion

Usually, the relation $\overline{\Delta N^2} = \overline{N}$ [22], where $\overline{\Delta N^2}$ is the root-mean-square (rms) fluctuation of the collected carriers and $\overline{N}$ is the average number of the collected carriers, is used for calculating the threshold characteristics of CID-photodetectors (charge-injection devices) for estimating the rms fluctuation of a dark and background charge caused by the shot character of the generation current. The use of this relation is based on the assumption that the generation process is of the Poisson type. This means that the generation rate remains constant during the charge accumulation by the MIS structure. Strictly speaking, this assumption is valid only for a small value of the collected charge, when it does not markedly affect the depth of the nonequilibrium depletion region. During the relaxation of the MIS capacitor, the charge accumulation in the inversion semiconductor layer leads to a decrease in the depletion region depth and, correspondingly, to a variation in the generation rate. This means that the generation probability of a minority carrier at a specific moment depends on the total number of previous generation events; i.e., the generation process cannot be considered as Poisson.

The simplest generalization of the Poisson process to the case when the probability of the next transition of the structure to a new state is related to the generation rate and depends only on the current structure state is considered in [23]. The formulas for rms fluctuation of the accumulated charge in the MIS structure are given as the result of the consideration of the physical system in the model of the Markovian process.

The fluctuation of the noise voltage can be expressed as

$$\delta U_n = (\delta\varphi_s/\delta N)\delta N, \tag{5}$$

where $\delta\varphi_s$ and δN are fluctuations of the surface potential and of the number of collected particles in the inversion layer due to generation of the charge carriers, respectively. The surface band bending is given by

$$\varphi_s(N) = V_k/2 + V_i - q\overline{N}/C_0 - \sqrt{V_k(V_i - q\overline{N}/C_0 + V_k/4)}. \tag{6}$$

Here, $V_k = 2\varepsilon_0\varepsilon_s N_d q S^2/C_0^2$ and $C_0 = C_n C_{ox}/(C_n + C_{ox})$, C_{ox} is the geometric capacitance of the gate insulator, C_n is the capacitive load, S is the area of the MIS structure, V_i is the injection pulse amplitude, and $q\overline{N}$ is the charge collected in the inversion layer. We calculated the noise-voltage standard deviation (SD) of the MIS structure according to the following formula:

$$U_n = -(q/C_0)\{[V_k/(2\sqrt{V_k(V_k/4 + V_i - q\overline{N}/C_0)})] - 1\} \times \sqrt{\int_0^{\tau_q} [I^2(\tau_q)/I(t)]dt}. \tag{7}$$

Here, τ_q is the collection time and I is the total current density.

The noise voltage was measured as $U_n = \sum_{n=1}^{M} U_n^*/(M - 1)$, where U_n^* is the instantaneous value of the noise sampling and M is the number of samplings.

We experimentally determined the noise SD of the MIS structure using five implementations (1024 samplings in each) for certain values of dc bias, the injection pulse, and the collection time. The square root of the differences of the noise SDs squared was taken as the noise SD. A minimum measured noise by estimations amounts to 75 charge carriers in the potential well of the MIS structure.

All the studied elements may be divided into two types: normal and anomalous elements. The noise of normal elements is defined by the potential well occupation by the charge carriers due to the dark generation, largely of a thermal character, and by the external back-

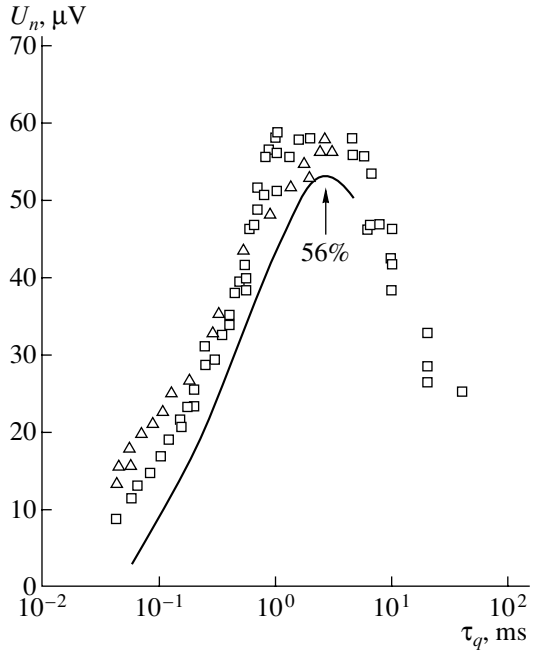


Fig. 7. The noise SD dependence on the collection time at $T = 150$ K and $V_i = 2$ V for two typical normal elements of 4×10^{-5} cm^2 in area, measured in the dark after the gate formation but without connection to a metallic wiring. The solid line is the theoretical calculation with the parameters: generation lifetime of the minority carriers $\tau_g = 2.5 \times 10^{-8}$ s, the thickness of the gate insulator $d_{ox} = 120$ nm, the diffusion length of the minority carriers in the InAs quasi-neutral bulk $L_p = 5$ μm, and $N_d = 3 \times 10^{15}$ cm^{-3}. The percentage of the potential well occupation in a CID-element by the dark charge carriers is indicated.

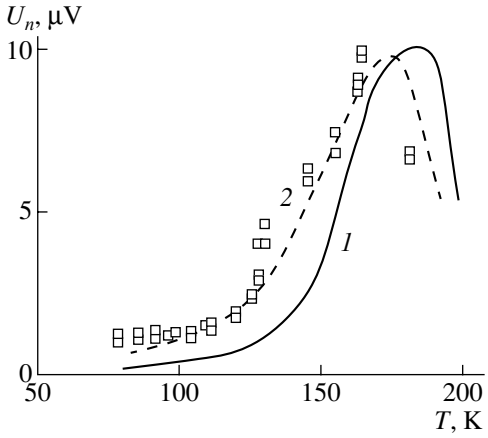


Fig. 8. The temperature dependence of the noise SD of a typical normal element of 4×10^{-5} cm^2 in area at $V_i = 1.3$ V measured at a load capacitor of 4 pF. (*1*) Results of calculations; (*2*) the results of semiempirical calculation with the excess currents shown in Fig. 9 taken into account.

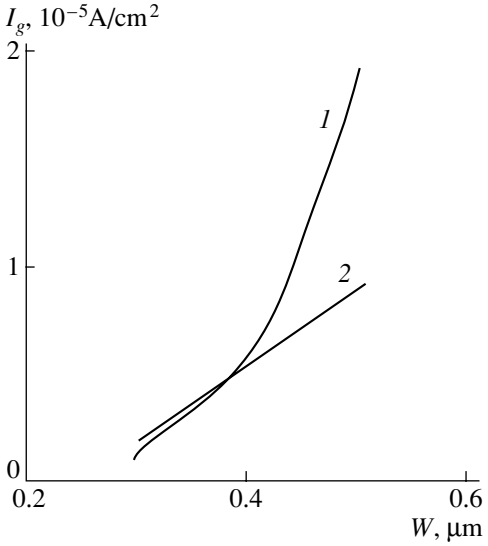


Fig. 9. The Zerbst curves: (*1*) obtained from the measurements of the capacitance relaxation and the surface potential of an MIS structure 4×10^{-5} cm^2 in area at $T = 150$ K; (*2*) theoretical dependence. W is the SCR width.

ground (infrared radiation incident on the element). The main criterion of the element normality is the noise spectrum similar to that of a white Poisson's noise. The noise of an anomalous element is significantly higher, and it does not correlate with the CID-element potential well occupation by the charge carriers. The component $1/f$ is obviously present in the noise spectrum.

Experimental and theoretical noise SD dependences for two typical normal elements on the collection time τ_q (at $T = 150$ K) and temperature are shown in Figs. 7 and 8. The noise peak is obvious in the curves. It approximately corresponds to a 55% occupancy of the element potential well by charges. The peak in these curves is satisfactorily described by the theoretical model concerning the CID-element noise. Experimental noise signals exceed the theoretical value at a small collection time by three or more times. The disagreement between theory and experiment in this region can be diminished if an excess generation current is taken into account as is shown in the Zerbst curves (Fig. 9). The generation current I_g was calculated from the capacitance relaxation curves and the surface potential of the MIS structure.

Thermal generation currents in the semiconductor SCR decrease appreciably as temperature decreases. Let us examine the MIS structure noise variation as a result of the occupation of the potential well under the effect of the background IR radiation in relation to the collection time variation. The experimental and theoretical dependences of the noise voltage rms on the collection time for a normal photodetector at $T = 120$ K are shown in Fig. 10. At this temperature, the thermal generation is suppressed and only the charge carriers generated by the background light contribute to the noise. A good agreement with the theoretical calculations is obvious (solid line). Arrows indicate the percentage

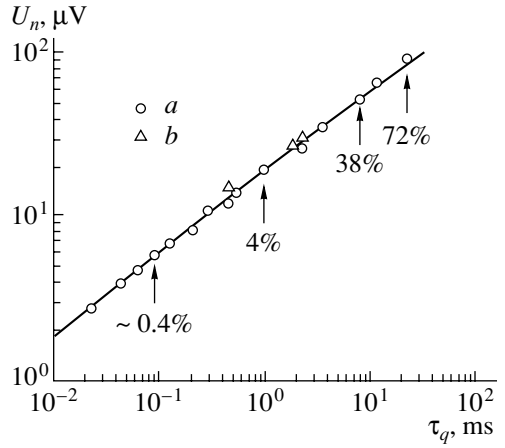


Fig. 10. The dependence of noise SD on the collection time at $T = 120$ K for the elements of 4×10^{-5} cm^2 in the area under illumination with (*a*) an unmodulated flux (background flux of 10^{-6} W cm^{-2}), (*b*) the noise of photodetectors combined into 2×128 array at collection times of 0.65, 1.3 and 2.6 ms. The solid line is the result of theoretical calculation.

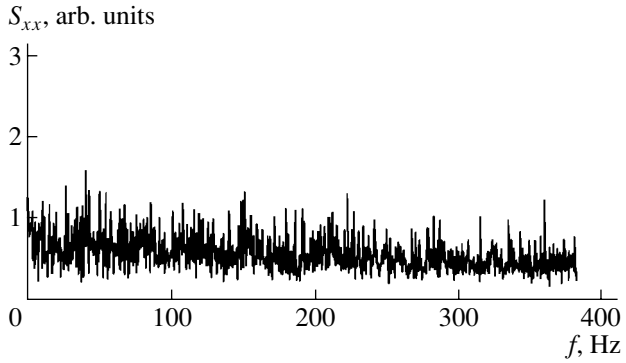


Fig. 11. The noise spectrum of the normal element measured at $T = 120$ K and at the collection time $\tau_q = 1.3$ ms without background illumination. The noise SD obtained by integration of the noise spectrum is 13 μV (in the MIS structure).

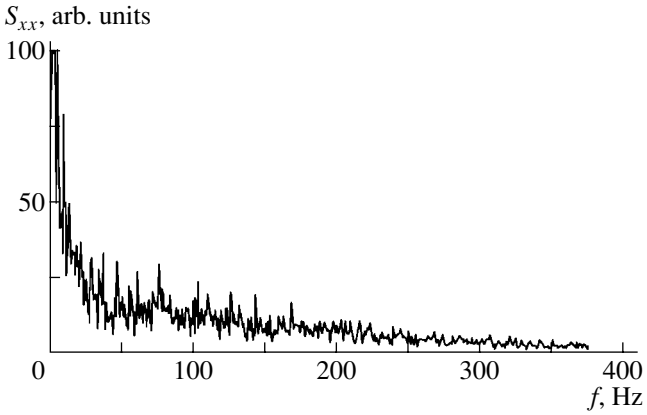


Fig. 12. The noise spectrum of an anomalous element measured at $T = 120$ K and the collection time $\tau_q = 1.3$ ms without background illumination. The noise SD obtained by integration of the noise spectral dependence is 43 μV (in the MIS structure).

occupation of the MIS-structure potential well by charges. The accumulated charge is proportional to the collection time τ_q. The good agreement of the experimental data and the $U_n \propto \tau_q^{1/2}$ relation shows that MIS structures are in the mode limited by the background fluctuations.

The noise spectrum $S_{xx}(f)$ (S_{xx} is the signal autocorrelation function and f is frequency) was measured in typical elements in the collection time range from 40 μs to 5 ms. An algorithm of the fast Fourier transform with the following convolution in the frequency band $1/\tau_q$ was used to calculate the noise-signal spectral power. A rectangularly shaped spectral window is chosen. The spectrum is obtained by averaging five representations with 1024 samplings in each. To obtain the spectrum noise density, the array of numbers was inputted into a computer and was subjected to the discrete Fourier transform. The spectrum was additionally integrated to obtain the noise SD. The noise SD and its spectral power were measured in several thousand MIS structures. It is possible to separate all the investigated structures into two types. The first type (>90% of all elements in linear and two-dimensional arrays) is characterized by a white noise of the MIS structure at nonequilibrium pulsed depletion (normal elements). The typical noise spectrum for the collection time of 1.3 ms is shown in Fig. 11. Among the studied elements, there are those in which the noise SD is several times (sometimes an order of magnitude or more) higher than the noise of normal elements. Such elements are referred to as anomalous. The noise spectrum of such an element is shown in Fig. 12. The spectrum is characterized by $1/f$ noise, which raises the low-fre-

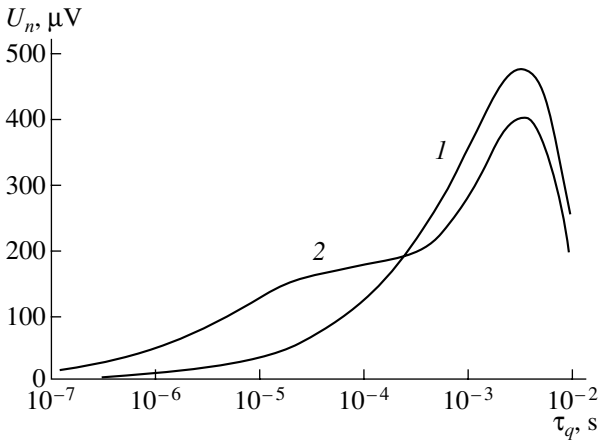


Fig. 13. Theoretical dependences of the noise SD on the collection time in the dark: (*1*) band-to-band tunneling is absent, (*2*) tunneling is taken into account. The computing parameters: $T = 150$ K, $S = 3.3 \times 10^{-5}$ cm^2, $V_i = 12$ V, $C_n = 50$ pF, $N_d = 3 \times 10^{15}$ cm^{-3}, the diffusion length of the minority carriers in the InAs quasi-neutral bulk $L_p = 5$ μm, and the generation lifetime of the minority carriers $\tau_g = 3 \times 10^{-8}$ s.

quency part of the spectrum. The origin of $1/f$ noise is not quite clear. However, we established two reasons by indirect evidence. They are local leakages in MIS structures and the deterioration of contact between the structure gate and indium pillar (for actual linear and two-dimensional arrays assembled by the flip-chip method).

It was mentioned above that an additional component related to the band-to-band tunneling appears as the injection-pulse amplitude increases. Theoretical noise SD dependences on the collection time are shown in Fig. 13. The plateau related to an excess noise occurs in the curve at small collection times (10–100 ms) provided that tunneling is taken into account (curve *2*). At large collection times, the tunnel current becomes lower than the thermal generation currents in SCR and the mechanism of the current formation is conventional, i.e., thermal.

CONCLUSION

Physicotechnological studies of InAs-based MIS structures made it possible to form a nearly perfect InAs–insulator interface. The lowest possible density of surface states ($<2 \times 10^{10}$ cm^{-2} eV^{-1}) enabled us to fabricate multielement FD linear and two-dimensional arrays of PDs which are characterized by virtually white noise in the absence of a background and by the number of the noise charge carriers (about 75–100 electrons) in the potential well of an element. On the basis of the structures formed, hybrid PD microcircuits with the detectivity of 10^{12} cm Hz$^{1/2}$ W^{-1} were fabricated [24–27]. It should be noted that the InAs-based MIS detector arrays of IR range operating under nonequilibrim depletion have been fabricated for the first time.

REFERENCES

1. A. V. Rzhanov, *Electronic Processes on Semiconductor Surface* (Nauka, Moscow, 1971).
2. W. E. Spicer, P. Pianetta, I. Lindau, and P. W. Chye, J. Vac. Sci. Technol. **14**, 885 (1977).
3. I. N. Sorokin and V. I. Kozlov, Izv. Akad. Nauk SSSR, Neorg. Mater. **15** (3), 53 (1979).
4. E. von Kauer and L. Rabenau, Z. Naturforsch. A **13** (7), 531 (1958).
5. R. L. Weiher and R. P. Ley, J. Appl. Phys. **37**, 299 (1966).
6. W. P. Doyle, J. Phys. Chem. Solids **4**, 144 (1958).
7. T. P. Smirnova, N. F. Zakharchuk, A. N. Golubenko, and V. I. Belyĭ, in *New Materials for Electronics* (Nauka, Novosibirsk, 1990), p. 62.
8. V. I. Belyĭ, in *Problems in Electronics Material Science* (Nauka, Novosibirsk, 1986), p. 29.
9. G. P. Schwartz, Thin Solid Films **103** (1/2), 3 (1983).
10. G. P. Schwartz, W. A. Sunder, and I. E. Griffths, Thin Solid Films **94**, 205 (1982).
11. C. W. Wilmsen, L. G. Meiners, and D. A. Collins, Thin Solid Films **46** (3), 331 (1977).
12. V. N. Davydov, E. A. Loskutova, and I. I. Fefelova, Mikroélektronika **15** (5), 455 (1986).
13. R. K. Ahrenkiel, L. L. Kazmerski, O. Jamjoum, *et al.*, Thin Solid Films **95** (4), 327 (1982).
14. A. P. Solov'ev, N. A. Valisheva, I. I. Marakhovka, *et al.*, USSR Inventor's Certificate No. 1604097 (1990).
15. N. A. Kornyushkin, N. A. Valisheva, A. P. Kovchavtsev, and G. L. Kuryshev, Fiz. Tekh. Poluprovodn. (St. Petersburg) **30** (5), 914 (1996) [Semiconductors **30**, 487 (1996)].
16. N. A. Valisheva, T. A. Levtsova, L. M. Logvinskiĭ, *et al.*, Poverkhnost, No. 11, 53 (1999).
17. G. Lukovski and R. S. Bauer, J. Vac. Sci. Technol. **17** (5), 946 (1980).
18. W. A. Harrison, Phys. Rev. **123** (1), 85 (1961).
19. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1963; Pergamon, New York, 1977).
20. W. W. Anderson, Infrared Phys. **17**, 147 (1977).
21. H. C. Gard, Solid State Commun. **31**, 877 (1979).
22. J. C. Kim, IEEE J. Solid-State Circuits **SC-13** (1), 158 (1978).
23. N. I. Khaliullin, Mikroélektronika **16** (5), 463 (1987).
24. G. L. Kurishev, A. P. Kovchavtzev, V. M. Bazovkin, *et al.*, Proc. SPIE **2746**, 268 (1996).
25. A. P. Kovchavtzev, E. S. Kogan, G. L. Kurishev, and G. L. Logvinski, in *OPTO-92, Paris, 1992*, p. 620.
26. G. L. Kuryshev, A. P. Kovchavtsev, B. G. Vaĭner, *et al.*, Avtometriya, No. 4, 5 (1998).
27. G. L. Kuryshev, A. P. Kovchavtsev, V. M. Bazovkin, *et al.*, Avtometriya, No. 4, 13 (1998).

Translated by I. Kucherenko

Semiconductors, Vol. 35, No. 9, 2001, pp. 1072–1080. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1120–1127.

SEMICONDUCTOR STRUCTURES, INTERFACES, AND SURFACES

Recombination of Point Defects and Their Interaction with the Surface in the Course of the Clusterization of these Defects in Si

L. I. Fedina

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

e-mail: fedina@thermo.isp.nsc.ru

Submitted February 14, 2001; accepted for publication March 1, 2001

Abstract—The growth kinetics of the {113} interstitial defects in *n*- and *p*-Si under *in situ* electron irradiation in a JEOL-1250 high-voltage electron microscope was analyzed in detail. The competing effects of the recombination of point defects and their interaction with the surface on the point-defect clusterization were considered. It is shown that the prevailing interaction of vacancies with the crystal surface is a consequence of the emergence of an energy barrier for the recombination of point defects in lightly doped Si crystals. Under these conditions, the drain of vacancies to the surface compensates for an increase in the crystal energy caused by the separation of the Frenkel pairs. It is assumed that a large difference between the growth rates of the {113} defects in *n*-Si and *p*-Si observed experimentally under the conditions when recombination is dominant is caused by a difference in the energy barriers for recombination. The barrier heights obtained correlate with the lattice-relaxation energy of a vacancy in various charge states. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The recombination of point defects and their interaction with the surface are the basic reactions that control the steady-state concentrations of point defects in Si. It is assumed that the competition between these reactions specifies not only the dominant type of point defects in the course of crystal growth but also serves as the key to controlling the formation of secondary microdefects [1, 2]. However, gaining insight into the mechanisms of the above reactions is hampered by the diversity of contradictory theoretical and experimental data, including those related to the basic point-defect parameters such as the activation volumes, diffusion coefficients, and equilibrium concentrations. It follows from calculations (reported, for example, in [3, 4]) that the vacancy-relaxation volume is large and negative; this volume not only compensates for the positive volume of an interstitial atom but also ultimately gives rise to the Frenkel pair formation having a negative volume, which should have resulted in spontaneous Frenkel pair generation in Si. This contradicts the data on the large threshold energy for the formation of a Frenkel pair in silicon (15 eV per atom [5]); on the low probability of the pair separation (10^{-3}–10^{-4} [6]); and, correspondingly, on the low steady-state concentrations of point defects under irradiation [7]. It is believed that the high rate of recombination in a closely spaced Frenkel pair is caused by the charge state of the point defects. However, even after the Frenkel pair separation, the point defects in Si recombine efficiently [7]; as a result, the growth kinetics of interstitial clusters in the surface region of the crystal is limited by an annihilation reaction of the second order and is controlled by the coefficient $k_{iv} = 4\pi r_{iv}/a$, where r_{iv} is the radius of a sphere in which the annihilation of point defects occurs and a is the interatomic distance. This is hardly consistent with the well-known theoretical notions about the presence of an energy barrier with a height of 1 eV for the interaction of point defects, which supposedly exist in Si in extended (split) configurations [3, 4, 8, 9]. These and many other contradictions do not allow the properties of point defects generated thermally or by irradiation to be generalized, nor do they allow a method to be devised for controlling and improving the structural quality of Si crystals. The objective of this study was to consider the competing effects of the recombination of point defects and their interaction with a surface on clusterization of these defects in Si. This consideration is based on a detailed *in situ* analysis of the defect-growth kinetics in the {113} plane under irradiation with 1-MeV electrons in a high-voltage electron microscope (HVEM).

The fact that a focused beam of electrons with an above-threshold energy is used in a HVEM makes it

1063-7826/01/3509-1072$21.00

possible to attain an irradiation intensity as high as 6×10^{19} electron/(cm^2 s), which is comparable with light-ion implantation in the generation rate of point defects. In combination with using high-resolution electron microscopy (HREM) to visualize the crystal structure, this method becomes very informative in studying the reactions of point defects. With this method, much insight into how the point defects interact with each other and with various sinks in the crystal has been gained [7, 10–14]. In this study, a HVEM is used to show that there is no basic discrepancy between the experimental data and theoretical concepts and that a barrier to the recombination of separated Frenkel pairs does exist. However, the height of this barrier depends on the type and concentration of the dopant and, supposedly, on the structural configuration of the vacancy and the magnitude of its lattice relaxation.

The notion about a surface as an efficient recombination center is widely accepted in scientific publications and implies an equiprobable drain of point defects to the surface. In fact, the *in situ* experiments with irradiation of Si with electrons in a HVEM indicate that the surface layers acting as sinks for vacancies and interstitial atoms are relatively thick and extend to 500 and 200–300 nm, respectively, from the surface if the native-oxide thickness does not exceed 1 nm [7]. However, as the oxide-film thickness increases to 15–30 nm, the thickness of the sink layer for vacancies changes only slightly, whereas that for interstitial atoms decreases to 10 nm. These data formed the basis for determining the coefficients of the interaction of point defects with the surface. It has been shown that the coefficient of interaction between vacancies and the oxidized Si surface (k_{vs} = 0.1–1) exceeds the corresponding coefficient for interstitial atoms (k_{is} = 10^{-3}–10^{-2}) by almost two orders of magnitude [7]. This gives rise to a buildup of interstitial atoms in the surface layer of the crystal irradiated in a HVEM, which are clusterized at $T < 600$°C in the {113} plane in the form of metastable configurations similar to configurations of atoms in the dislocation core. Detailed data about the interstitial-atom configurations in the Si and Ge crystals in relation to the irradiation temperature can be found elsewhere [7]. At the same time, during ion implantation, the point defects annihilate so effectively that only a single interstitial atom per each implanted ion remains in the crystal and is involved in the formation of the interstitial-type defects and in impurity diffusion (one of the latest reviews in this field can be found in the publication by Chason *et al.* [15]). In this so-called (+1) model, the drain of vacancies to the surface is ignored and, in fact, excludes vacancies from involvement in the implanted-impurity diffusion. An increase in the implanted-ion dose leads to an increase in the number of interstitial atoms accumulated in associations to 4.5 per each ion; i.e., the (+1) model is transformed into a (+4.5) model. This may be assumed as indicative of both the clusterization of vacancies in the implanted layer and the origination of a drain of vacancies to the surface. The latter assumption is supported by the results obtained from *in situ* irradiation with Si^+ ions in ultrahigh vacuum in a scanning tunneling microscope [16]. In this study, it is shown that the drain of vacancies to the atomically clean surface at T = 20–500°C precedes the drain of interstitial atoms. It may be assumed that the drain of vacancies to the surface in the course of growing the silicon ingots also ensures supersaturation with respect to interstitial atoms in the growing crystal if the crystallization-front velocity is low. Using the recent experimental [14] and theoretical [17] data on relaxation volumes of point defects, we will show that, when a barrier for the point-defect recombination emerges, the drain of vacancies to the surface may compensate for an increase in the crystal energy caused by the Frenkel pair separation.

2. EXPERIMENTAL

In this study, the wafers of P-doped *n*-Si grown by the floating-zone method (FZ-Si) and of P- and B-doped *n*- and *p*-Si grown by the Czochralski method (CZ-Si), with the (110), (001), and (111) orientations were used; the resistivities ranged from 4000 to 0.005 Ω cm. The oxygen, carbon, and phosphorus concentrations in FZ-Si were 10^{16}, 3×10^{15}, and 3×10^{12} cm^{-3}, respectively. The concentrations of oxygen and carbon in CZ-Si were 10^{18} and 10^{17} cm^{-3}, respectively. Disks about 2.8 mm in diameter were then cut from the wafers. These disks were thinned by etching in a HF : HNO_3 = 1 : 5 solution. The etching was continued until a hole appeared; variable-thickness regions transparent for electrons were formed around this hole. The thickness of the crystal was determined from the number of extinction contours formed in the image of a tapered crystal around the hole under the conditions of dynamic diffraction of electrons. In order to ensure that the conditions for the interaction of point defects with the surface were reproducible, all the thinned samples were coated with 15-nm-thick SiO_2 or Si_3N_4 layers (a detailed description of experiments in a HVEM can be found in [7]). The samples were irradiated with 1-MeV electrons with intensities of 3×10^{18}–5×10^{19} electron/(cm^2 s) in the temperature range 20–800°C; a JEOL-1250 HVEM was used.[1] Temperature was measured using a thermocouple embedded in the sample holder.

[1] Experiments with a JEOL-1250 were conducted by me at the Belgian National Center for Electron Microscopy (the RUCA-EMAT University, Antwerp) in 1996–1997.

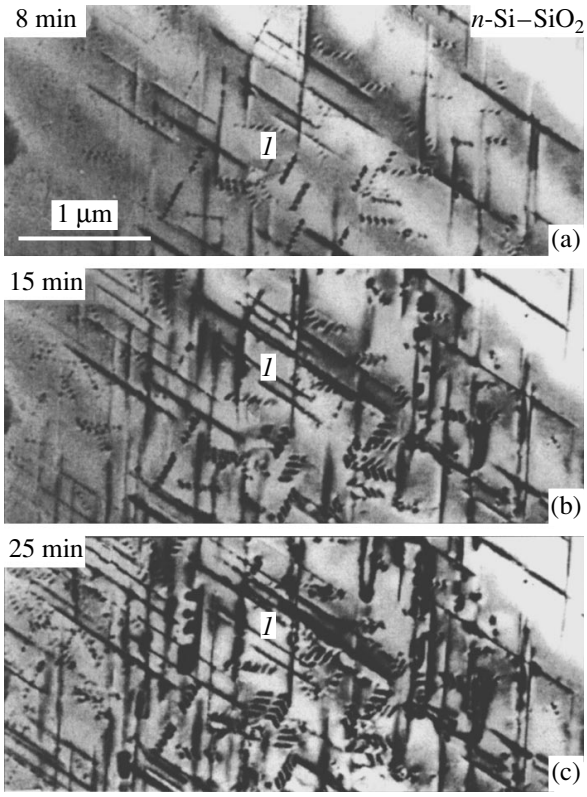


Fig. 1. Successive growth stages for the {113} defects in n-Si–SiO_2 (with the Si thickness h = 0.1–0.4 μm and with the phosphorus concentration $C_P = 10^{17}$ cm^{-3}) in the course of irradiation with electrons in a high-voltage electron microscope. The SiO_2 thickness is 15 nm. The irradiation intensity $I = 5 \times 10^{18}$ electron/(cm^2 s); T = 500°C. The irradiation duration is indicated in the microphotographs. The numeral *1* in Figs. 1a–1c indicates the same slanting defect.

3. RESULTS

3.1. Clusterization of Point Defects in Si at T = 20–350°C

The process of clusterization of point defects in this temperature range, including the *in situ* irradiation in a high-resolution electron microscope (HREM), has been described elsewhere [10–12]. At T = 20°C, this process involves a joint clusterization of point defects, which results in the formation of amorphous regions in the {113} family of planes. Depending on the local concentration of interstitial atoms, these regions are transformed either into the well-known interstitial-type {113} defects with hexagonal structure or into the perfect crystal. This made it possible to detect a barrier to the recombination of interstitial atoms with extended vacancy clusters (the barrier height is about 1 eV); owing to this barrier, metastable configurations of interstitial atoms can form within a vacancy defect [12].

3.2. Growth Kinetics of the {113} Defects in Lightly Doped Si Crystals

The increase in the irradiation temperature to 500°C gives rise to steady-state stages of growth of the {113} defects; these stages set in after a short period of enhanced growth of defects in the ⟨110⟩ direction. The initial stage of the defect growth is related to the pipe diffusion of interstitial atoms and is considered in detail by Aseev *et al.* [7]. In this stage, the size of the defects in the ⟨110⟩ directions may be as large as several micrometers at T = 500°C; that is why these defects were originally referred to as rodlike defects. The steady-state stage of defect growth is accompanied by an increase in their sizes in the ⟨332⟩ direction perpendicular to the ⟨110⟩ defect axis; as a result, the defects acquire a two-dimensional shape with the {113} crystal habit. Long steady-state growth stages are observed in the n-Si–SiO_2 and p-Si–SiO_2 structures subjected to low-intensity electron irradiation ($I \lesssim 6 \times 10^{18}$ cm^{-2} s^{-1}) if the crystal thickness $h \lesssim 0.5$ μm. A series of electron microscopy images of the {113} defects in n-Si–SiO_2 is shown in Figs. 1a–1c; it can be seen that the growth with the rate of V_{332} = 0.1 nm/s in the ⟨332⟩ direction is time-independent. The results of measuring the steady-state growth rates V_{332} in relation to the irradiation intensity I in the n-Si–SiO_2 and p-Si–SiO_2 structures with the Si thickness $h < 0.5$ μm are shown in Fig. 2a. It can be seen that, for n-Si, the dependence $V_{332}(I)$ is nearly linear ($V_{332} \propto I$), whereas this dependence in p-Si deviates from linear for $I > 6 \times 10^{18}$ electron/(cm^2 s) and starts to follow a parabolic law ($V_{332} \propto I^{1/2}$).

More pronounced distinctions in the growth kinetics of the {113} defects and in their spatial arrangement between n-Si–SiO_2 and p-Si–SiO_2 are observed if the Si thickness $h \gtrsim 1.5$–2 μm (see Fig. 2b). All defects in n-Si originate and grow in a thin (0.1–0.2 μm) surface layer and do not penetrate into the crystal bulk (Figs. 3a–3c). It follows from the electron microscopy images that the {113} defects inclined to the surface virtually do not penetrate into the bulk; these defects are displayed in the form of thin streaky defects in a thin n-Si layer (Fig. 1, defect *1*). In the initial stage, some of the defects penetrate into the bulk but immediately disappear. These defects are completely unobserved in ultrahigh-purity FZ-Si crystals with a phosphorus concentration of 10^{12} cm^{-3}. The high growth rate of the {113} defects in n-Si results in the fact that they rapidly attain a critical size of 200 nm in the ⟨332⟩ direction and are immediately transformed into perfect dislocation loops and dipoles; these extended defects climb with different kinetics controlled by changes in the configurations of interstitial atoms in the plane of the loops. At the same time, very long and slanting {113} defects penetrating deep into the crystal bulk are observed in p-Si (Figs. 4a–4c). The results of measuring the dependence $V_{332}(I)$ in n- and p-Si–SiO_2 and in

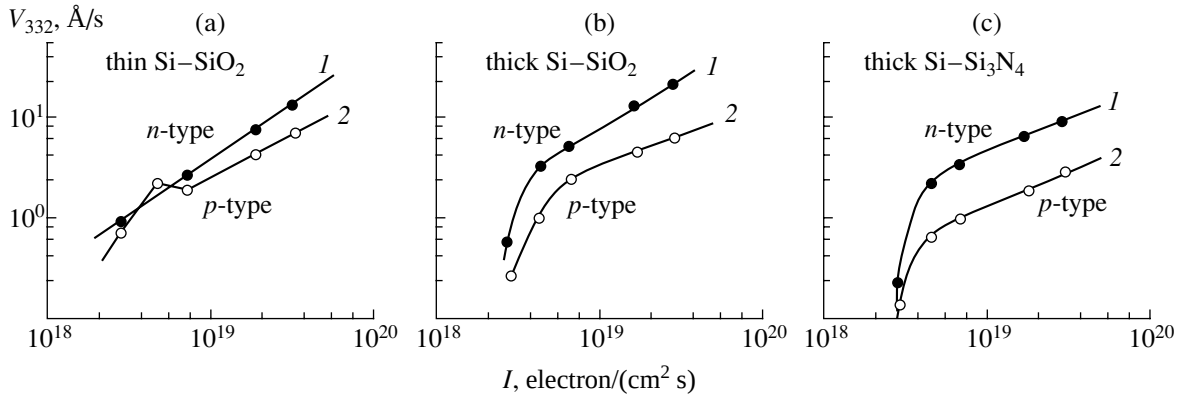


Fig. 2. Dependences of the steady-state growth rates of the {113} defects V_{332} on the irradiation intensity I in lightly doped structures: (a) and (b) correspond to the Si–SiO_2 structures and (c) is for a Si–Si_3N_4 structure; the Si thickness was h = (a) 0.1–04 and [(b), (c)] 1.5–2 μm. The phosphorus and boron concentrations were equal to 10^{16} cm^{-3}.

n- and *p*-Si–Si_3N_4 are shown in Figs. 2b and 2c for the Si thickness of h = 1.5–2 μm. It can be seen that the linear dependence $V_{332} \propto I$ is observed for *n*-Si–SiO_2 (Fig. 2b, curve *1*), whereas the dependence $V_{332} \propto I^{1/2}$ is observed in *n*-Si–Si_3N_4 (Fig. 2c, curve *1*). Deviations from these dependences for small values of I are related to the fact that only the pipe diffusion of interstitial atoms along ⟨110⟩ occurs owing to the low generation rate of point defects (about 10^{-7} s^{-1}). In *p*-Si, the dependence $V_{332} \propto I^{1/2}$ is observed irrespective of the type of coating deposited on the surface of irradiated samples (Figs. 2b, 2c, curves *2*). It also follows from Fig. 2c that the steady-state growth rates for the {113} defects are threefold higher in *n*-Si than in *p*-Si.

3.3. Growth Kinetics of the {113} Defects in Heavily Doped Si Crystals

In silicon with the boron and phosphorus concentrations $C_{B,P} \gtrsim 10^{19}$ cm^{-3}, the growth kinetics of the {113} defects changes drastically. The dependences $V_{332}(I)$ in *n*-Si–Si_3N_4 with the Si thickness of h = 2 μm and C_P = 10^{16}–10^{19} cm^{-3} are shown in Fig. 5a. It can be seen that the dependence $V_{332}(I)$ changes only slightly when C_P is varied in a wide range. However, it is worth noting that, in these crystals, dislocation loops are formed and slanting {113} defects penetrating into the bulk appear. The dependences $V_{332}(I)$ for comparatively thick *p*-Si–Si_3N_4 crystals with $C_B = 5 \times 10^{15}$–2.5×10^{19} cm^{-3} are shown in Fig. 5b. The growth rate V_{332} is seen to decrease drastically as the concentration C_B increases. Symbol *4* in Fig. 5b corresponds to the growth rate for the {113} defects in very thin crystals, whereas the {113} defects are not observed at all in thick *p*-Si crystals with $C_B = 2.5 \times 10^{19}$ cm^{-3}. The process of the clusterization of point defects in heavily doped Si crystals was outlined in detail in [7, 18].

4. DISCUSSION

4.1. Recombination of Point Defects in Si

Clusterization of point defects during *in situ* irradiation in a HVEM is adequately described in terms of

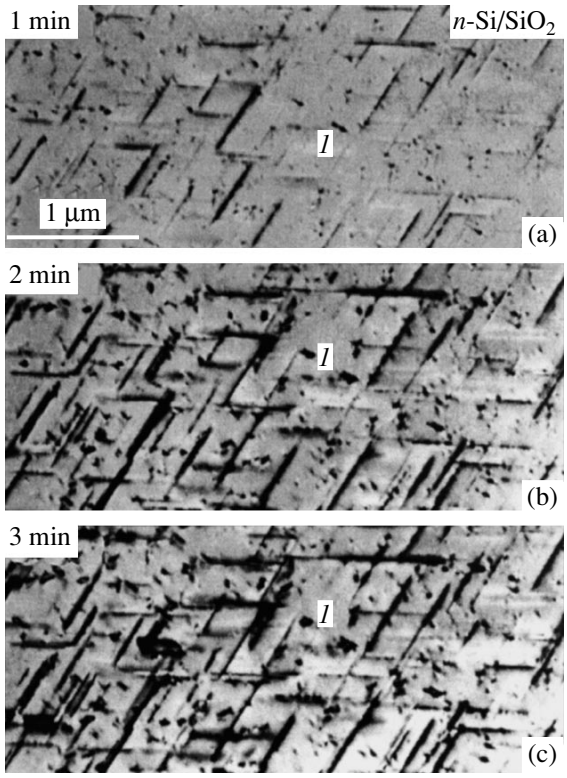


Fig. 3. Successive growth stages for the {113} defects in *n*-Si–SiO_2 (with a Si thickness of h = 1.5–2 μm and a phosphorus concentration of $C_P = 10^{17}$ cm^{-3}) in the course of irradiation with electrons in a high-voltage electron microscope. The SiO_2 thickness is equal to 15 nm. The irradiation intensity $I = 3 \times 10^{19}$ electron/(cm^2 s); T = 500°C. The irradiation duration is indicated in the microphotographs. The numeral *1* in Figs. 3a–3c indicates the same slanting defect.

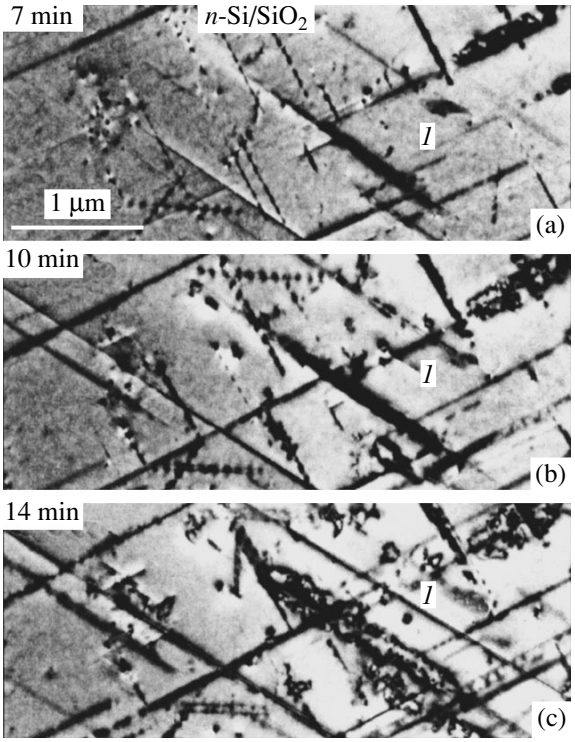


Fig. 4. Successive growth stages for the {113} defects in p-Si–SiO_2 (with the Si thickness h = 1.5–2 μm and the boron concentration $C_B = 10^{17}$ cm^{-3}) in the course of irradiation with electrons in a high-voltage electron microscope. The SiO_2 thickness is equal to 15 nm. The irradiation intensity $I = 3 \times 10^{19}$ electron/(cm^2 s); T = 500°C. The duration of irradiation is printed in the micrographs. The numeral *1* in Figs. 4a–4c indicates the same slanting defect.

the theory of the rates of quasi-chemical reactions; the applicability of this theory to Si and Ge has been considered in detail by Aseev *et al.* [7]. According to this theory, the growth rate of the interstitial-type {113} defects in the steady-state stage of irradiation can be given by

$$V_{332} \approx a k_{id} C_i D_i. \quad (1)$$

Here, a is the interatomic distance, k_{id} = 1–2 is the coefficient of interaction between interstitial atoms and the {113} defect, C_i is the interstitial-atom (i) concentration during irradiation in reference to the concentration of atoms in the crystal, and $D_i = \nu_0 \exp(-E_{im}/kT)$ is the coefficient of interstitial-atom migration. Depending on which reaction (recombination or drain of vacancies to the surface) controls the establishment of quasi-equilibrium in the crystal, one can obtain simple expressions for the point-defect concentrations. If recombination prevails, we have

$$C_i D_i \approx (P D_v / k_{iv})^{1/2}, \quad (2)$$

whereas, if the drain of vacancies to the surface is dominant, we obtain

$$C_i D_i \approx P / c_s k_{vs}. \quad (3)$$

Here,

$$P = k_a z \sigma I \quad (4)$$

is the generation rate of point defects; k_a is the coefficient accounting for direct recombination of Frenkel pairs with closely spaced components ($k_a = 10^{-3}$–10^{-4} [6]); z is the number of defects generated by a single electron (z = 1); σ is the cross section for displacement of an atom as a result of elastic collision with an electron ($\sigma = 5 \times 10^{-23}$ cm^2); c_s is the surface-sink density; and k_{vs} is the coefficient accounting for the interaction of vacancies with the surface (k_{vs} = 0.1–1 [7]). According to [19], the coefficient accounting for the point-defect recombination that follows the second-order reaction is written as

$$k_{iv} = 4\pi r_{iv} / a, \quad (5)$$

and $r_{iv} = 4a$. It follows from formulas (1)–(3) that $V_{332} \propto I$ if the drain of vacancies to the surface is prevalent, whereas $V_{332} \propto I^{1/2}$ if the point-defect recombination is dominant.

Experimental dependences $V_{332} \propto I$ in n-Si–SiO_2 (Figs. 2a, 2b) indicate that the growth kinetics of the {113} defects in n-Si–SiO_2 during irradiation is controlled by the drain of vacancies to the surface irrespective of the crystal thickness. This also indicates that the recombination of point defects in n-Si is retarded. At the same time, in both relatively thin and relatively thick p-Si–SiO_2 structures, the dependence $V_{332} \propto I^{1/2}$ prevails; i.e., the growth of the {113} defects is controlled by the reaction of point-defect recombination. The value of r_{iv} in p-Si can be estimated from the change in the dependence $V_{332}(I)$ in a thin p-Si–SiO_2 structure for $I = 6 \times 10^{18}$ electron/(cm^2 s) (see Fig. 2a); this change is indicative of the onset of recombination. By using (4), we obtain the generation rate $P = 3 \times 10^{-8}$–3×10^{-7} s^{-1}, which corresponds to the concentration of generated point defects approximately equal to 10^{15}–10^{16} cm^{-3}; thus, we obtain $r_{iv}^p \approx$ 50–100 nm. The latter value is much larger than $r_{iv} \approx 15a \approx 7$ nm estimated [6] on the assumption of the Coulomb interaction between the point defects and suggests that the interaction of point defects is elastic in the course of their recombination. The fact that the dependence $V_{332} \propto I^{1/2}$ is observed in n-Si–Si_3N_4 and p-Si–Si_3N_4 and that the growth rates of point defects in n-Si and in p-Si differ by a factor of 3 corresponds approximately to the tenfold difference between the recombination rates of point defects; these rates are accounted for by the coefficients k_{iv} in formula (2).

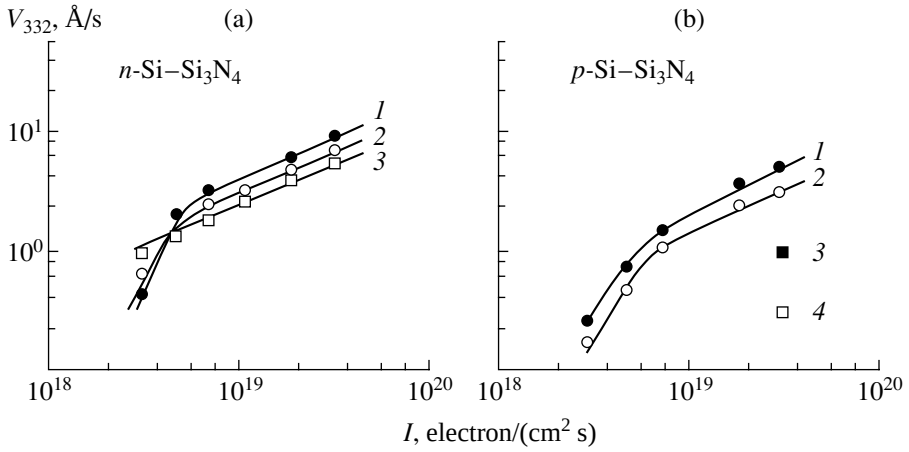


Fig. 5. Dependences of the steady-state growth rates for the {113} defects V_{332} on the irradiation intensity I in (a) n-Si–Si_3N_4 and (b) p-Si–Si_3N_4. The Si thickness h = 1.5–2 μm. In Fig. 5a, the phosphorus concentration C_P = (*1*) 10^{16}, (*2*) 10^{17}, and (*3*) 10^{19} cm^{-3}. In Fig. 5b, the boron concentration C_B = (*1*) 5×10^{15}, (*2*) 10^{17}, (*3*) 10^{18}, and (*4*) 2.5×10^{19} cm^{-3}.

Obviously, both the charge states of defects and their lattice relaxation can play an important role in the recombination of point defects. However, it is difficult to suggest with confidence which types of point defects are dominant in *n*- and *p*-Si at elevated temperatures and are involved in recombination. It is believed that the lattice relaxation of point defects is well understood. According to [3, 14], an interstitial atom features a large positive value of the relaxation volume $V_i^{\text{rel}} = (0.9–0.94)\Omega$, where Ω is the volume occupied by a Si atom in the equilibrium Si lattice. For a vacancy, the corresponding volume V_v^{rel} is negative and varies from $(–0.25)\Omega$ to $(–0.97)\Omega$ [3, 14, 17, 20]. Recent calculations performed *ab initio* [17] using a large cluster (incorporating 200 atoms) have shown that the volume of lattice relaxation of a vacancy and the relaxation energy depend on the vacancy's charge state. The relaxation energies are equal to 0.33 and 0.45 eV for positively charged vacancies (V^+ and V^{++}) and to 0.90 eV for a neutral vacancy (V^0). Unfortunately, the data on negatively charged vacancies are lacking in [17]. However, it is clear that a large magnitude of lattice relaxation for a vacancy is caused by the capture of electrons by this defect and is a consequence of the Jahn–Teller effect. The data obtained by Watkins from experiments with electron spin resonance also indicate that the Jahn–Teller stabilization energy varies from 0.4 eV for V^+ to about 1 eV for a vacancy–phosphorus complex [21]. The strong dependence of the vacancy lattice relaxation on the charge state and the lack of similar data for interstitial atoms suggest that the recombination of point defects is largely controlled by the structural configuration of the vacancy, rather than by this configuration of an interstitial atom. Taking into account that an interstitial atom can also exist in the split configuration [3, 4, 8, 9, 13, 14], we may assume that the probability of its recombination should decrease appreciably as the vacancy becomes more extended. This is indirectly supported by the presence of a barrier with a height of about 1 eV for the recombination of interstitial atoms with extended vacancy clusters in Si; these clusters were observed *in situ* in Si irradiated with electrons in a HREM [10–14].

Assuming that there is an energy barrier to the recombination of separated Frenkel pairs, we may write the expressions for the recombination coefficients in *n*- and *p*-Si as

$$k_{iv}^{p,n} = (k_{iv}^{p,n})^0 \exp(S_{iv}^{p,n}/k)\exp(-E_{iv}^{p,n}/kT), \qquad (6)$$

where $S_{iv}^{p,n}$ is the entropy and $E_{iv}^{p,n}$ is the relaxation energy.

It follows from the aforementioned estimate $r_{iv}^p \approx$ 50–100 nm for *p*-Si and from expression (5) that $k_{iv}^p \approx$ 1000–2000. Assuming that, in expression (6) for k_{iv}^p, the number of possible lattice sites around the vacancy $(k_{iv}^p)^0 = 4$ and that the barrier height in *p*-Si corresponds to one of the values for the lattice-relaxation energy of the positively charged vacancies (V^+ and V^{++}) $E_{iv}^p \approx 0.33–0.45$ eV, we obtain $S_{iv}^p \approx (10–11)k$ for the entropy. Such a large value of S_{iv}^p is typically related to the extended configurations of point defects and testifies that the point defects interact elastically while recombining. It follows from the experimentally determined ratio between the growth rates of the {113} defects in *n*- and *p*-Si (Fig. 2c) that the ratio of the recombination coefficients is given by $k_{iv}^p/k_{iv}^n \approx 10$. On the assumption that the values of $(k_{iv}^{p;n})^0$ and $S_{iv}^{p,n}$ are

the same for *n*- and *p*-Si, we then obtain the barrier height for recombination in *n*-Si equal to $E^n_{iv} \approx$ 0.5–0.6 eV. Obviously, a higher energy barrier to the recombination in *n*-Si gives rise to higher steady-state concentrations of point defects in this material and to a shift of the growth zone for the {113} defects towards the surface. The value obtained for E^n_{iv} is close to $E^n_{iv} \approx$ 0.6–0.8 eV determined from the nucleation kinetics of the point-defect clusters in heavily doped *n*-Si at T = 200°C [7, 18]. The fact that the values obtained for E^n_{iv} are nearly the same and that the growth kinetics of the {113} defects depends only slightly on the phosphorus concentration (Fig. 5a) is indicative of a unified mechanism of the defect recombination in *n*-Si. According to the data reported by Watkins [21], this implies that a single type of strongly relaxed vacancies is dominant in *n*-Si. At the same time, the heavy dependence of the growth kinetics of the {113} defects in *p*-Si on the boron concentration (see Fig. 5b) indicates that the bimolecular mechanism for the point-defect recombination is replaced by recombination at the boron atoms; i.e.,

$$\mathrm{Si}_i + \mathrm{B}_s = \mathrm{B}_i, \quad \mathrm{B}_i + V = \mathrm{B}_s.$$

According to the data reported previously [7, 18] and to the results of this study, it may be concluded that only the reaction $\mathrm{B}_i + V = \mathrm{B}_s$ causes the drain of vacancies to the surface to be suppressed; as a result, clusterization of interstitial atoms in heavily doped *p*-Si ceases.

4.2. The Drain of Point Defects to the Surface

According to [7], the predominant drain of vacancies to the oxidized Si surface in the absence of a high boron concentration is related to the fact that $k_{vs} \gg k_{is}$. The above analysis shows that the growth of the interstitial-type {113} defects is caused by the drain of vacancies to the surface and manifests itself under the conditions of the point-defect recombination with an energy barrier involved. On the basis of the recent experimental and theoretical data on the relaxation volumes of point defects [14, 17], it can be explained why it is the drain of vacancies, rather than interstitial atoms that occurs under the conditions of intense irradiation. Without going into the details of the theoretical analysis suggested previously [22] to calculate the activation volumes of point defects, the total energy of the crystal can be written as

$$E = (1/2)\sum \varphi_{ij} + \sum \varphi_{jk} + ac_1 + bc_1^2, \qquad (7)$$

where $(1/2)\sum \varphi_{ij}$ is the crystal energy related to the lattice potential of interaction between atoms, $\sum \varphi_{jk}$ is the energy of the near-boundary region, ac_1 is the contribution of the work expended on maintaining the ideal lattice in equilibrium, and bc_1^2 is the contribution of displacement of atoms in the crystal to the elastic energy of the crystal. Eshelby has shown [23] that an increase in the volume ΔV, which is related to the lattice configuration in a finite sample and is a linear function of the elastic variable c_1, is given by

$$\Delta V = 4\pi a_1 c_1, \qquad (8)$$

where a_1 is the force correction coefficient.

As an extremely simplified variant, let us consider the variation in the energy E related only to a change in the crystal volume due to the generation of the Frenkel pairs. According to the data reported by Fedina [14], the volume corresponding to the formation of a separated Frenkel pair $V^f_{i,v}$ is equal to the sum of relaxation volumes of an interstitial atom and a vacancy; i.e.,

$$V^f_{i,v} = V^{\mathrm{rel}}_i + V^{\mathrm{rel}}_v = 0.94\Omega + (-0.5\Omega) = 0.44\Omega,$$

where Ω is the atom's volume in the equilibrium silicon lattice. If vacancies do not reach the surface, then

$$\Delta V = N_s\Omega(1 + 0.44C_{i,v}),$$

where N_s is the number of atoms in the crystal and $C_{i,v}$ is the number of separated Frenkel pairs referenced to the number of atoms in the crystal. As follows from (7) and (8), this should result in an increase in the crystal energy. If a vacancy reaches the surface, the volume corresponding to the formation of a Frenkel pair becomes formally equal to the volume of formation of an interstitial atom; i.e., $V^f_i = (-\Omega) + 0.94\Omega = -0.06\Omega$. The change in the crystal volume is then given by

$$\Delta V = N_s\Omega(1 - 0.06C_v),$$

where C_v is the concentration of vacancies that reached the surface; this concentration is divided by the number of atoms in the crystal. Such a change in the volume is much more favorable. The corresponding energy gain is compensated by the energy of interstitial atoms clusterized in the form of the {113} defects; the latter energy is small and does not exceed 0.7 eV per interstitial atom [24]. If an interstitial atom reaches the surface, the volume corresponding to the formation of a Frenkel pair becomes equal to the vacancy-formation volume; i.e., we have $V^f_v = \Omega + (-0.5\Omega) = 0.5\Omega$. Correspondingly, this results in an increase in the crystal energy, as in the case of separation of a Frenkel pair in the crystal bulk. It is only if both a vacancy and an interstitial atom reach the surface that we have $V^f_{i,v} = \Omega + (-\Omega) = 0$, which is equivalent to their recombination at the surface. The basic inequality $V^f_i < V^f_v$ should have yielded a calculated energy of formation of an interstitial atom lower than that of a vacancy. However, an analysis of published data shows that variations in these energies are related to the type of potential used in the

calculations, so that the relation between V_i^{rel} and V_v^{rel} and, correspondingly, between V_i^f and V_v^f, as well as between the formation energies of point defects, may become reversed [3, 4]. Nevertheless, it follows from the comparison of the data obtained from the first principles in terms of the local approximation for the electron-density functional (LDA) that the interstitial-atom energy in Si is 3.35 eV, whereas this energy amounts to 3.65 eV for a vacancy [24, 25].

Thus, if there are thin SiO_2 or Si_3N_4 layers at the surface of irradiated crystals (in which case, interstitial atoms cannot reach the surface), the drain of vacancies to the surface ensures the compensation for an increase in the crystal energy caused by the separation of the Frenkel pairs and by the accumulation of point defects. It should be noted that the drain of vacancies to the surface is suppressed not only by the high concentration of boron atoms but also (see, for example, [14]) by large compression stresses (larger than 0.3%) in pseudomorphic Si_xGe_{1-x} layers grown over Si. In these crystals, recombination of point defects occurs even if the layer thickness is about 10 nm. The data obtained in this study (Fig. 2b, 2c) also testify that the drain of vacancies to the surface diminishes. Due to the origination of weak compression stresses at the $Si–Si_3N_4$ interface, the drain of vacancies diminishes, which ensures the manifestation of recombination in the growth kinetics of the {113} defects in *n*-Si.

5. CONCLUSION

The roles of recombination and the interaction of point defects with the surface in the course of the clusterization of point defects in Si were considered on the basis of a detailed analysis of steady-state growth kinetics of the {113} interstitial-type defects in relation to the crystal thickness, the type of surface coating (SiO_2 or Si_3N_4), the doping impurity (boron or phosphorus) concentration, and the intensity of 1-MeV electron irradiation in a HVEM. It is shown that the higher growth rates of the {113} defects in *n*-Si compared to *p*-Si under the conditions of the predominance of recombination are caused by a difference in the heights of energy barriers for recombination in these crystals: 0.6–0.8 and 0.33–0.45 eV, respectively. The energy-barrier heights obtained in *n*- and *p*-Si correlate with the relaxation energies of vacancies in different charge states. The fact that the growth kinetics of the {113} defects in *n*-Si is independent of the phosphorus concentration implies the invariance of the point-defect recombination mechanisms in these crystals. At the same time, the strong dependence of the defect-growth kinetics in *p*-Si on the boron concentration is related to a change in the point-defect recombination mechanism; as a result, the drain of vacancies to the surface and the clusterization of interstitial atoms are suppressed. Using the data on the relaxation volumes of point defects, it is shown that the drain of vacancies to the surface under the conditions of the existence of a barrier to the recombination of point defects ensures the compensation for an increase in the crystal energy caused by the separation of the Frenkel pairs.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 98-02-17798.

I thank Professor J. van Landuyt, head of the RUCA-EMAT Department of the Antwerp University (Belgium), for letting me conduct the experiments using a JEOL-1250 electron microscope and also Professor A.L. Aseev for his helpful participation in discussions and for his critical comments advanced after reading this paper.

REFERENCES

1. V. V. Voronkov, J. Cryst. Growth **59**, 625 (1982).
2. R. Falster and V. V. Voronkov, MRS Bull. **25** (6), 28 (2000).
3. M. Tang, L. Colombo, J. Zhu, and T. D. de la Rubia, Phys. Rev. B **55**, 14279 (1997).
4. D. Maroudas and R. A. Brown, Phys. Rev. B **47**, 15562 (1993).
5. L. I. Fedina, A. L. Aseev, S. G. Denisenko, and L. S. Smirnov, Fiz. Tekh. Poluprovodn. **21**, 2037 (1987) [Sov. Phys. Semicond. **21**, 363 (1987)].
6. A. L. Assev, S. G. Denisenko, and L. I. Fedina, Fiz. Tekh. Poluprovodn. (Leningrad) **25**, 582 (1991) [Sov. Phys. Semicond. **25**, 352 (1991)].
7. A. L. Aseev, L. I. Fedina, D. Hoehl, and H. Barsch, *Clusters of Interstitial Atoms in Silicon and Germanium* (Nauka, Novosibirsk, 1991; Akademie Verlag, Berlin, 1994).
8. U. Gösele and T. Y. Tan, in *Diffusion in Solids, Unsolved Problems* (Trans. Tech. Publ., Zurich, 1992), p. 189.
9. J. W. Corbett and J. C. Bourgoin, in *Point Defects in Solids*, Vol. 2: *Semiconducand Molecular Crystals*, Ed. by J. H. Grawford, Jr. and L. M. Slifkin (Plenum, New York, 1975; Mir, Moscow, 1979).
10. L. Fedina, A. Gutakovskii, A. Aseev, J. van Landuyt, and J. Vanhellemont, *In situ Electron Microscopy in Material Research* (Kluwer, Dordrecht, 1997), Chap. 4, p. 63.
11. L. Fedina, A. Gutakovskii, A. Aseev, *et al.*, Philos. Mag. A **77**, 423 (1998).
12. L. Fedina, A. Gutakovskii, A. Aseev, *et al.*, Phys. Status Solidi A **171**, 147 (1999).
13. L. Fedina, A. Gutakovskii, and A. Aseev, Cryst. Res. Technol. **35**, 775 (2000).
14. L. Fedina, O. Lebedev, G. van Tendeloo, *et al.*, Phys. Rev. B **61**, 10336 (2000).
15. E. Chason, S. T. Picraux, J. M. Poate, *et al.*, J. Appl. Phys. **81**, 6513 (1997).

16. P. J. Bedrossian, M.-J. Caturla, and T. Diaz de la Rubia, Appl. Phys. Lett. **70**, 176 (1997).
17. S. Öğüt, H. Kim, and J. R. Chelikowsky, Phys. Rev. B **56**, R11353 (1997).
18. L. I. Fedina and A. L. Aseev, Fiz. Tverd. Tela (Leningrad) **32**, 60 (1990) [Sov. Phys. Solid State **32**, 33 (1990)].
19. A. C. Damask and G. J. Dienes, in *Point Defects in Metals* (Gordon and Breach, New York, 1963; Mir, Moscow, 1966), p. 291.
20. A. Antonelli and J. Bernholc, Phys. Rev. B **40**, 10643 (1989).
21. G. D. Watkins, in *Radiation Effects in Semiconductors* (Plenum, New York, 1968), p. 67.
22. Yu. M. Plishkin, in *Defects in Crystals and Their Computer Simulation* (Nauka, Leningrad, 1980), p. 100.
23. J. D. Eshelby, J. Appl. Phys. **25**, 255 (1954).
24. J. Kim, F. Kirchhoff, J. W. Wilkins, and F. S. Khan, Phys. Rev. Lett. **84**, 503 (2000).
25. J. Zhu, T. Diaz de la Rubia, L. H. Yang, *et al.*, Phys. Rev. B **54**, 4741 (1996).

Translated by A. Spitsyn

Semiconductors, Vol. 35, No. 9, 2001, pp. 1081–1087. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1128–1134.

LOW-DIMENSIONAL SYSTEMS

Spin Response of 2D Electrons to a Lateral Electric Field

L. I. Magarill, A. V. Chaplik, and M. V. Éntin

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

Submitted February 14, 2001; accepted for publication February 14, 2001

Abstract—The spin orientation of two-dimensional (2D) electrons by a lateral electric field is considered. The electron dispersion law is assumed to contain linear terms due to the spin–orbit band splitting in an asymmetric quantum well. The coefficient of spin orientation in a DC electric field is found. The mean electron spin is oriented in the sample plane perpendicularly to the electric field. The interaction of an AC electric field with spins of 2D electrons is studied. It is shown that transitions between different spin states give rise to a narrow absorption band. These states are mixed with 2D plasmons in the frequency range related to these transitions, with the result being that the plasmon spectrum is modified and a new type of oscillations arises (spin-plasmon polariton). The problem of the generation of spin-plasmon polaritons by an external field is solved. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The recent revival of interest in spin nanoelectronics makes studying the spin effects in low-dimensional systems an urgent problem. First of all, this is related to the possible application of the spin degrees of freedom to create quantum bits. To use the spin degrees of freedom in semiconductor devices, it is desirable to control electron spins electrically since other means of control (magnetic field, light) are not local enough and, therefore, cannot ensure selective action on a comparatively small modern semiconductor device. Moreover, the magnetic field needs to be sufficiently strong to cause an essential redistribution of electrons over spin states, and such a field cannot be changed fast enough. The effect of electric field on the spin degree of freedom can be provided by the spin–orbit coupling. It is well known that the spin–orbit coupling of two-dimensional (2D) electrons on an oriented surface[1] can be described using the Hamiltonian proposed by Vas'ko [1] and Rashba [2]:

$$\hat{\mathcal{H}}_0 = \frac{\mathbf{p}^2}{2m} + \hat{\mathcal{H}}_{so}, \quad \hat{\mathcal{H}}_{so} = \alpha(\boldsymbol{\sigma}[\mathbf{pn}]) \equiv (\boldsymbol{\sigma}\boldsymbol{\Omega}). \quad (1)$$

Here, $\mathbf{p} = (p_x, p_y)$ is the 2D electron momentum, the z-axis is perpendicular to the system plane, α is the spin–orbit coupling constant, and σ_i are the Pauli matrices; hereafter, $\hbar = 1$.

The Hamiltonian (1) yields the energy spectrum

$$\varepsilon_\mu(p) = \frac{p^2}{2m} + \mu|\alpha|p, \quad (2)$$

where $\mu = \pm 1$ enumerates two branches of the spin-split spectrum of the 2D electron gas. The splitting of the branches is usually small, constituting about 10^{-2} of the Fermi energy. That is why thermodynamic manifestations of the spin–orbit coupling are hardly of any interest; however, the spin degrees of freedom may be essential for some kinetic processes. In this paper, we consider the effect of the DC and AC electric field with the vector lying in the plane of the system on the 2D Fermi gas and take into account the spin–orbit coupling. A DC electric field leads to spin polarization of electrons. In a high-frequency field, a narrow resonance emerges in the absorption spectrum, related to transitions between spin–orbit-split subbands (in the absence of a magnetic field!). The interplay of this transition with 2D plasmons gives rise to mixed plasmon-spin waves. The spectrum of these oscillations will be found, and the possibility of their generation will be considered.

2. SPIN ORIENTATION OF 2D ELECTRONS BY AN ELECTRIC FIELD

In this section, we consider the emergence of a mean spin density **S** in a 2D system under the action of a lateral electric field **E**. The phenomenology of the effect in a system isotropic in the (x, y) plane (z-axis is parallel to the normal **n** to the surface) is described by the equation

$$\mathbf{S} = \gamma[\mathbf{n} \times \mathbf{E}]. \quad (3)$$

Here, γ is the spin orientation coefficient.

[1] A physical implementation of an oriented surface is the 2D electron system with different upper and lower boundaries, e.g., a heterojunction or an inversion channel.

1063-7826/01/3509-1081$21.00 © 2001 MAIK "Nauka/Interperiodica"

The spin orientation coefficient can be written in terms of the Kubo formalism [3]:

$$\gamma = \frac{-e}{A}\int_0^\infty dt e^{-\delta t} \times \int_0^{1/T} d\lambda \langle \mathrm{Sp}\{ f(\hat{\mathcal{H}})\hat{\upsilon}_x(-i\lambda)(1 - f(\hat{\mathcal{H}}))\hat{s}_y(t)\}\rangle. \quad (4)$$

Here, A is the system area, T is temperature, $-e$ is the electron charge, $\hat{\mathcal{H}}$ is the total Hamiltonian of the system, $f(\varepsilon) = 1/(1 + \exp(\varepsilon - \zeta)/T)$ is the Fermi distribution (ζ is the chemical potential), $\delta \longrightarrow +0$, and $\hat{\mathbf{v}}(t)$ and $\hat{\mathbf{s}}(t)$ are the operators of electron velocity and spin in the Heisenberg representation. The angle brackets denote averaging over the impurity distribution. For the sake of definiteness, the electric field is chosen to be aligned with the x-axis, with the resulting spin directed along the y-axis.

Averaging over scatterers in formula (4) leads to the inclusion of the relaxation mechanism. The formula can be disentangled using the Edwards graph technique [4]. It can also be interpreted directly without averaging, if δ implies the increment of the electric field (turning-on by the exponential law ~expδt) or a phenomenological inverse relaxation time.

One more way to use this formula is to directly calculate the spin density generation rate $(\dot{\mathbf{S}})_g$ instead of the spin density response. In the case of an electric field exponentially rising with time, we have

$$(\dot{\mathbf{S}})_g = \delta\mathbf{S} = \delta\gamma[\mathbf{n}\times\mathbf{E}].$$

It can be seen from what follows that, in the limit of slow field turn-on and weak scattering, the quantity $\delta\gamma$ contains neither the rate at which the field rises nor any relaxation parameters.

In the eigenfunction basis of the Rashba Hamiltonian,

$$\psi_\mu = \frac{\exp(i\mathbf{pr})}{\sqrt{2S}}\begin{pmatrix} i\mu\exp(-i\varphi_\mathbf{p})\,\mathrm{sgn}(\alpha) \\ 1 \end{pmatrix}, \quad (5)$$

where $\varphi_\mathbf{p}$ is the polar angle of the $\mathbf{p}$ vector, and the matrix elements of the electron velocity and spin operators are as follows:

$$\mathbf{v}_{\mu\mu}(\mathbf{p}) = \frac{\mathbf{p}}{pm}(p + m|\alpha|\mu), \quad \mathbf{v}_{\mu,-\mu}(\mathbf{p}) = i|\alpha|\mu\frac{[\mathbf{p}\times\mathbf{n}]}{p}, \quad (6)$$

$$\mathbf{s}_{\mu\mu}(\mathbf{p}) = \mu\frac{\alpha[\mathbf{p}\times\mathbf{n}]}{2|\alpha|p}, \quad \mathbf{s}_{\mu,-\mu}(\mathbf{p}) = i\mu\frac{\alpha\mathbf{p}}{2|\alpha|p}. \quad (7)$$

With scattering disregarded, formula (4) can be written in the form

$$\gamma = -\frac{e}{A}\sum_{\mathbf{p};\,\mu\mu'}\frac{f[\varepsilon_\mu(p)] - f[\varepsilon_{\mu'}(p)]}{\varepsilon_\mu(p) - \varepsilon_{\mu'}(p)} \times \{\mathbf{n}[\mathbf{v}_{\mu\mu'}(\mathbf{p})\times\mathbf{s}_{\mu'\mu}(\mathbf{p})]\}\frac{i}{\varepsilon_\mu(p) - \varepsilon_{\mu'}(p) + i\delta}. \quad (8)$$

As a result, in the collisionless approximation, we have

$$\gamma = \frac{me\alpha}{4\pi\delta}. \quad (9)$$

Using formula (9), we obtain for the rate of spin generation by static electric field

$$(\dot{\mathbf{S}})_g = \frac{me\alpha}{4\pi}[\mathbf{n}\times\mathbf{E}]. \quad (10)$$

The relaxation can be accounted for by the spin density balance equation

$$(\dot{\mathbf{S}})_g + (\dot{\mathbf{S}})_r = 0,$$

where the spin relaxation rate is determined by the expression $(\dot{\mathbf{S}})_r = -\mathbf{S}/\tau_s$ in terms of the spin relaxation time τ_s. This approach is justified if τ_s is independent of the electron energy or if only monoenergetic electrons (those on the Fermi surface) are involved. As a result, $1/\tau_s$ should be substituted for the quantity δ in (9).

In order to verify and generalize the phenomenological result (9), we use the quantum kinetic equation approach. The electron kinetics is described by a one-particle spin density matrix $\hat{\rho}(\mathbf{p})$. The quantum kinetic equation linearized in the external electric field reads

$$\dot{\rho} = i[\hat{\mathcal{H}}_{so}, \hat{\rho}] + \hat{\mathcal{F}} = \mathrm{St}(\hat{\rho}), \quad (11)$$

where $\mathcal{F}$ is the field term and $\mathrm{St}(\hat{\rho})$ is the collision term.

It is convenient to use a basis of states with a fixed spin quantization axis. The field term then reads

$$\hat{\mathcal{F}} = -\frac{e}{4m\varepsilon}\mathbf{E}\{2\mathbf{p}(2\varepsilon + \hat{\mathcal{H}}_{so})f'(\varepsilon + \hat{\mathcal{H}}_{so}) + i[\mathbf{p},\mathbf{n}](\boldsymbol{\sigma}\mathbf{n})[f(\varepsilon + \hat{\mathcal{H}}_{so}) - f(\varepsilon - \hat{\mathcal{H}}_{so})]\}, \quad (12)$$

where $\varepsilon = p^2/2m$.

Considering the relaxation, we take into account only collisions with impurities. The collision integral relying upon the smallness of the parameter $\lambda = \alpha/\upsilon_F$, which allows one to take into account only scattering with a conserved spin, can be transformed to

$$\mathrm{St}(\hat{\rho}) = 2\pi N_i\sum_{\mathbf{p}'}|V_{\mathbf{p}'-\mathbf{p}}|^2\delta(\varepsilon - \varepsilon')[\hat{\rho}(\mathbf{p}') - \hat{\rho}(p)]. \quad (13)$$

Here, $V_{\mathbf{p}'-\mathbf{p}}$ is the Fourier component of the impurity potential (with ignored spin–orbit corrections to the Hamiltonian of interaction with impurities).

To the zero and first orders in the spin–orbit coupling constant, the quantity $\hat{\mathscr{F}} = \hat{\mathscr{F}}^{(0)} + \hat{\mathscr{F}}^{(1)}$ reads

$$\mathscr{F}^{(0)} = -e\frac{\mathbf{E}\mathbf{p}}{m}f'(\varepsilon), \quad f'(\varepsilon) = \frac{\partial f}{\partial \varepsilon}, \quad \varepsilon = \frac{p^2}{2m}, \quad (14)$$

$$\mathscr{F}^{(1)} = -e\mathbf{E}\left\{\frac{\mathbf{p}}{p^2}(f'(\varepsilon) + 2\varepsilon f''(\varepsilon))\mathscr{H}_{so} - \alpha f'(\varepsilon)\frac{[\mathbf{p}\times\mathbf{n}]}{p^2}(\boldsymbol{\sigma}\mathbf{p})\right\}. \quad (15)$$

The spin density is determined by the trace of the density matrix with the spin operator. Solving the quantum kinetic equation to a lower order in the constant α, we find

$$\gamma = -\frac{e}{8\pi\alpha}\int_0^\infty \frac{d\varepsilon}{\varepsilon\tau_1}[f'(\varepsilon) + \varepsilon f''(\varepsilon)]. \quad (16)$$

Here,

$$\tau_n^{-1} = 2\pi N_s\int\frac{d^2p'}{(2\pi)^2}|V(\mathbf{p}-\mathbf{p}')|^2\delta[\varepsilon(p)-\varepsilon(p')] \times\{1-\cos[n(\phi-\phi')]\} \quad (17)$$

is the relaxation time of the *n*th angular moment of the distribution function (τ_1 is the conventional transport time of momentum relaxation).

In the low-temperature limit, the coefficient γ is determined by the contribution from the Fermi surface. As a result, we obtain

$$\gamma = \frac{e}{8\pi\alpha}\left[\frac{1}{\tau_1(\varepsilon_F)\varepsilon_F} - \frac{d}{d\varepsilon_F}\left(\frac{1}{\tau_1(\varepsilon_F)}\right)\right]. \quad (18)$$

The relaxation time for scattering on charged impurities in the system plane is proportional (with screening disregarded) to the electron energy: $\tau_1 \propto \varepsilon$. In the limit of a degenerate Fermi gas with Fermi energy ε_F, we have for γ from formula (18)

$$\gamma = \frac{e}{4\pi\alpha\varepsilon_F\tau_1(\varepsilon_F)}, \quad (19)$$

which corresponds to formula (9) with $\tau_s^{-1} = \pi\alpha^2 n_s\tau_1$ coinciding with the inverse time of spin relaxation via scattering on impurities by the D'yakonov–Perel mechanism [5, 6]. In the case of scattering on neutral impurities (also lying in the system plane), τ_1 is independent of energy, with the second term in formula (18) vanishing.

As follows from formula (19), γ grows with decreasing α. This seemingly paradoxical statement is a consequence of the spin relaxation slowing down at $\alpha \longrightarrow 0$. Let us explain this in more detail.

A very important feature of the spin degree of freedom is that, in the limit under consideration $\Omega\tau_1 \ll 1$, the spin relaxes relatively slowly in comparison with the momentum, following the D'yakonov–Perel mechanism. Owing to this slowed-down relaxation, electrons that are transferred from one spin subband to another are accumulated there, which leads to an increase in responses dependent on the electron distribution over the spin subbands. In particular, the D'yakonov–Perel [5] mechanism gives an inverse spin relaxation time proportional to the squared splitting of the spin subbands. At the same time, as seen from formula (9), the rate of spin generation by a static lateral electric field is proportional to the first power of the spin–orbit splitting. As a result, the mean spin induced by the lateral field increases rather than decreases with a decreasing spin–orbit splitting.

In the above reasoning, we disregarded a weaker Elliott–Jafet mechanism of spin relaxation, governed by the spin–orbit coupling with the impurity potential, rather than by the spin–orbit splitting. Taking this mechanism into account changes the dependence of γ on α in such a way that the coefficient γ vanishes at $\alpha \longrightarrow 0$.

According to formula (16), the optimal condition for the effect to be observed is relatively low mobility of electrons. In addition to an increase in γ, this enables a higher electric field to be applied without electron or sample heating.

In conclusion, we estimate the effect for the typical case of GaAs/GaAlAs heterostructure with an electron density of 5×10^{11} cm^{-2} and a mobility of 10^4 cm/(V s), $\alpha = 4\times10^5$ cm/s [2]. Substituting these values into formula (19), we obtain $\gamma = 3\times10^7$ (cm V)$^{-1}$; i.e., 3×10^8 spins per cm^2 are oriented when a field of 10 V/cm is applied.

Let us briefly discuss the possibility of observing the considered effect. The spin polarization of 2D electrons could be directly detected on the basis of changes in the magnetic moment of the 2D system in the electric field. An infinite plane with magnetic moment lying in the plane does not create a magnetic field.

Let us consider a 2D strip $-L < y < L$ directed along the electric field. Spins in this strip will be oriented along the surface and perpendicular to the current direction (*x*-axis). Their magnetic moments induce a magnetic field $B_y^{(S)}(z=0)$ in the plane of the system described by the interpolation formula

$$B_y^{(S)} = 2g\mu_B SL((y^2-L^2)^2 + 4d^2L^2)^{-1/2},$$

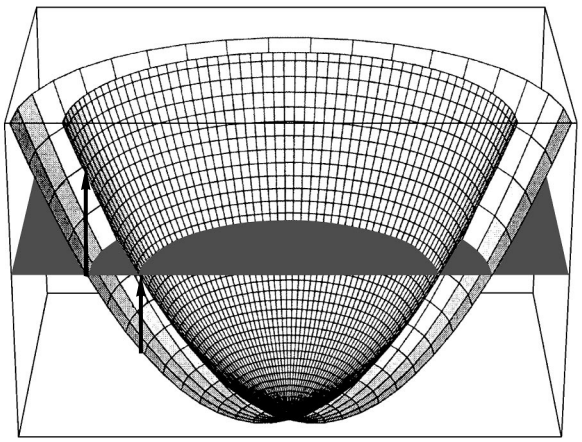


Fig. 1. Spin–orbit-split energy spectrum of the 2D system. Arrows show transitions between spin subbands with the threshold frequencies $2(|\alpha|p_F \pm m\alpha^2)$.

taking into account the finite thickness d of the 2D layer (g is the Landé splitting factor, and μ_B is the Bohr magneton).

This field is added to the magnetic field of the current j flowing along the strip. Just above and below the plane, this field has a magnitude of $B_y^{(j)} = 2\pi j/c$. Outside the strip, in the plane of the system, the field has only a z-component of about the same order of magnitude. For the above parameters, we estimate the maximum field magnitudes to be $B_y^{(S)} \approx 3 \times 10^{-3}$ G and $B_z^{(j)} \approx 10^{-3}$ G.

Precision measurements of the magnetic field are possible with the use of a SQUID formed in the sample plane. Its typical size is several micrometers. Therefore, the maximum magnetization field is three orders of magnitude less at the SQUID. At the same time, SQUID is mainly sensitive to the horizontal component of the magnetic field. According to [7], the SQUID sensitivity is good to 10^{-10} G, which is quite sufficient for detecting the magnetization.

It is noteworthy that the effect considered in this study is to some extent an inverse effect with respect to the appearance of a current under the action of spin-polarized electrons [6]. The primary spin polarization induced by optical transitions under the action of circularly polarized light, leading in the end to a steady-state current, was considered in [6]. In our case, the current induced by the DC electric field leads to the orientation of electron spins.

3. SPIN-PLASMON OSCILLATIONS IN A 2D ELECTRON GAS

Let us study the effect of an ac electric field on the 2D Fermi-gas, taking into account the spin–orbit coupling, and first calculate the dynamic conductivity in the collisionless approximation. The Kubo formula [3] for the dynamic conductivity reads

$$\sigma_{ij}(\omega) = \frac{\pi G_0}{S}\int_0^\infty dt e^{-\delta t}\int_0^{1/T} d\lambda \times \langle \mathrm{Sp}\{ f(\hat{\mathcal{H}})\hat{\upsilon}_j(-i\lambda)[1 - f(\hat{\mathcal{H}})]\hat{\upsilon}_i(t)\}\rangle. \tag{20}$$

Here, $G_0 = e^2/\pi$ is the conductance quantum.

In the collisionless limit, we obtain from formula (20)

$$\sigma_{ij}(\omega) = \frac{G_0}{4}\int d\mathbf{p}\sum_\mu \left\{ \frac{\partial f_{\mu p}}{\partial \zeta}\frac{\upsilon^j_{\mu\mu}(\mathbf{p})\upsilon^i_{\mu\mu}(\mathbf{p})}{\delta - i\omega} \times \frac{f_{\bar{\mu}p} - f_{\mu p}}{\varepsilon_\mu(p) - \varepsilon_{\bar{\mu}}(p)}\frac{\upsilon^j_{\mu\bar{\mu}}(\mathbf{p})\upsilon^i_{\bar{\mu}\mu}(\mathbf{p})}{\delta - i\omega + \varepsilon_\mu(p) - \varepsilon_{\bar{\mu}}(p)} \right\}. \tag{21}$$

Here, $\bar{\mu} = -\mu$, $f_{\mu p} \equiv f[\varepsilon_\mu(p)]$.

In what follows, we assume that the electron gas is degenerate in terms of the smallness of temperature in comparison with the Fermi energy ε_F. At the same time, the temperature may be comparable with the splitting between the spin subbands.

Substitution of formulas (2) and (6) into (21) gives ($\sigma_{ij}(\omega) = \delta_{ij}\sigma(\omega)$):

$$\sigma(\omega) = \sigma_D(\omega) + \sigma_s(\omega),$$
$$\sigma_D(\omega) = \frac{n_s e^2}{m(\delta - i\omega)} = iG_0\frac{\varepsilon_F}{\omega + i\delta}, \tag{22}$$

$$\sigma_s(\omega) = \frac{G_0|\alpha|}{8}\int_0^\infty dp \times [f(p^2/2m - |\alpha|p) - f(p^2/2m + |\alpha|p)] \times \left[\frac{1}{\delta - i\omega - 2i|\alpha|p} + \frac{1}{\delta - i\omega + 2i|\alpha|p}\right]. \tag{23}$$

Here, n_s is the 2D electron density. The quantity σ_D is the contribution to the conductivity, which is associated with the velocity matrix elements diagonal in terms of μ. We write it in the leading order in the parameter $(\alpha/\upsilon_F)^2$ (υ_F is the Fermi velocity), in which this contribution coincides with the Drude conductivity. The quantity σ_s is related to transitions between the spin subbands.

We are interested in the frequency range close to ω_0 ($|\omega - \omega_0| \ll \omega_0$). In this range, expression (23) for σ_s simplifies to

$$\mathrm{Re}\sigma_s = \frac{iG_0}{16}\left[\frac{1}{\exp\left(\frac{\eta - 1}{\theta}\right) + 1} - \frac{1}{\exp\left(\frac{\eta + 1}{\theta}\right) + 1}\right],$$

$$\mathrm{Im}\sigma_s = \frac{iG_0}{16}\mathcal{P}\int_{-\infty}^{\infty}\frac{dy}{\eta - y + i\nu_s} \tag{24}$$

$$\times\left[\frac{1}{\exp\left(\frac{y-1}{\theta}\right)+1} - \frac{1}{\exp\left(\frac{y+1}{\theta}\right)+1}\right],$$

where the following dimensionless parameters are introduced: temperature $\theta = T/(\alpha p_F)$ and detuning relative to the center of the spin absorption band, $\eta = (\omega - \omega_0)/(2m\alpha^2)$. The symbol $\mathcal{P}$ stands for the principal part of the integral.

In the low-temperature limit $\theta \ll 1$, we obtain from formula (24)

$$\sigma_s = \frac{iG_0}{16}\left[\ln\left|\frac{\eta+1}{\eta-1}\right| + \pi\theta(1-\eta^2)\right]. \tag{25}$$

At $T = 0$, $\mathrm{Re}\sigma(\omega)$ is a step function of ω and is nonzero within the range

$$\omega_0 - 2m\alpha^2 < \omega < \omega_0 + 2m\alpha^2.$$

This reflects the energy conservation law for vertical transitions between spin subbands (see Fig. 1). The imaginary part of the conductivity has logarithmic singularities at the ends of this interval. The low-temperature limit is reached at temperatures lower than the subband splitting $2|\alpha|p_F$ near the Fermi momentum. At higher temperatures, the absorption peak is broadened by $2Tm\alpha/p_F$ and the absorption amplitude decreases because of the equalization of band occupancies.

The electron scattering can be taken into account by means of the quantum kinetic equation method formulated in Section 2. In order for the plasmon attenuation to be small near the resonance $\omega \approx \omega_0$, the inequality $\Omega\tau_1 \gg 1$ must be valid. After rather cumbersome algebra, we obtain formula (29) in the collisionless limit and, in the low-temperature limit $\theta \ll 1$, the expression

$$\sigma_s = \frac{iG_0}{16}\ln\frac{\eta + 1 + i(2m\alpha^2\tau_s)^{-1}}{\eta - 1 + i(2m\alpha^2\tau_s)^{-1}}. \tag{26}$$

In the limit under consideration, the collision-related broadening of spin transitions, $\tau_s^{-1} = (2\tau_1^{-1} + \tau_2^{-1})/4$ has the same order of magnitude as τ_1 governing the mobility.

The value of τ_s is easily found in various particular cases. For example, $\tau_s = (4/3)\tau_1$ for neutral impurities. If scattering is determined by unscreened charged impurities lying in the 2D layer plane, then $\tau_s = \tau_1$. Finally, for small-angle scattering (charged impurities, thick spacer), $\tau_s = (2/3)\tau_1$.

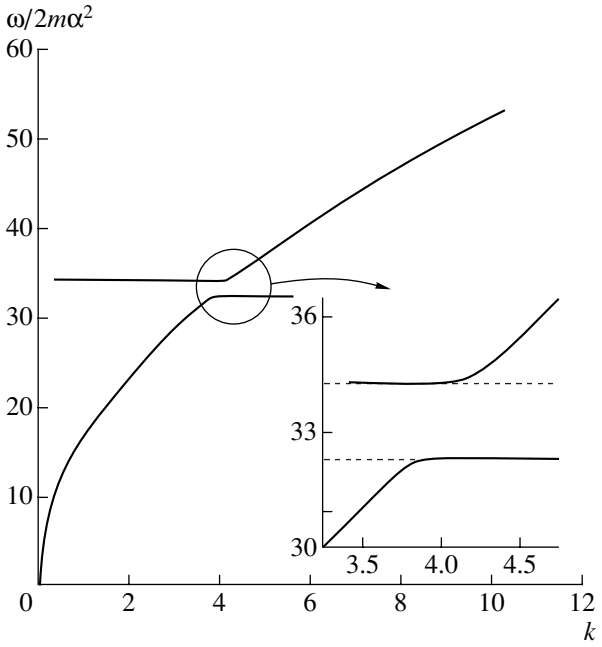


Fig. 2. Spin-plasmon spectrum in the absence of wave attenuation for $\lambda = 0.0259$, corresponding to an InAs/GaSb heterostructure with carrier density 7.5×10^{11} cm^{-2}. $2m\alpha^2 = 0.117$ meV. The spectrum shows the square-root behavior $\omega \propto \sqrt{k}$ far away from the intersection with spin transitions. Inset: the intersection region. Dotted lines show limiting frequencies for spin transitions, which are asymptotes of the spectrum; the spectrum branches exponentially approach these straight lines.

4. PLASMON SPECTRUM

Let us consider the problem of oscillations of a 2D electron gas, taking into account transitions between spin subbands in the absence of a field electrode and ignoring displacement currents.

The continuity equation $\dot{\rho} + \nabla\mathbf{j} = 0$ for the surface charge density ρ and the current density $\mathbf{j}$, the Poisson equation $\Delta\phi = 4\pi\rho\delta(z)$ for the potential ϕ, and the constitutive equation $\mathbf{j} = -\sigma\nabla\phi$ are to be solved together in order to find the oscillation spectrum.

After a Fourier transform in the sample plane, the dispersion equation can be obtained similarly to [8] (see also [9]). In the simplest case, the dispersion equation for the plasmon reads

$$\frac{2\pi i\sigma(\omega)q}{\omega} = -\kappa. \tag{27}$$

Here, q is the wave vector; κ is the effective static dielectric permittivity (if the 2D plane is confined between two insulators with permittivities κ_1 and κ_2, then $\kappa = (\kappa_1 + \kappa_2)/2$). Thus, knowledge about the frequency dependence of the conductivity implicitly determines the spin-plasmon spectrum $\omega(k)$.

In the collisionless limit, $\delta_D = \delta_s = 0$ and the spectrum at zero temperature is determined by the equation

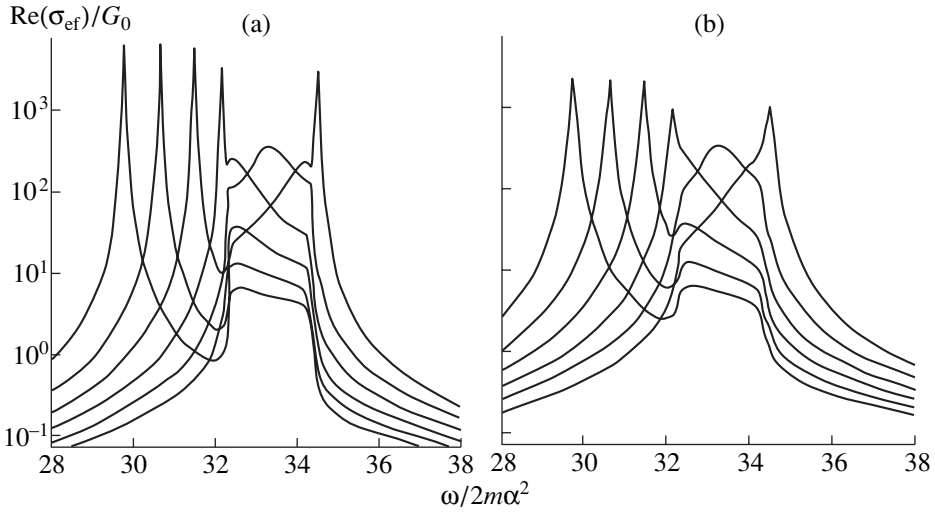


Fig. 3. Frequency dependence of the absorption coefficient near the spin-plasma resonance for the system presented in Fig. 2. The dimensionless wave vector takes the values k = 3.2, 3.4, 3.6, 3.8, 4.0, and 4.2. To these values corresponds the shift of the absorption peak from the left to the right; mobility μ: (a) 10^6 and (b) 3×10^6 cm^2/(V s).

$$\frac{1}{k} = \frac{1}{16}\frac{1}{1+\eta\lambda}\left[\frac{4}{1+\eta\lambda} + \lambda\ln\left(\frac{\eta+1}{\eta-1}\right)\right], \qquad (28)$$

where the dimensionless wave vector $k = q\pi G_0/\kappa m\alpha^2$ and the parameter $\lambda = \alpha/\upsilon_F$ are introduced.

There are undamped plasma waves at $\eta > 1$ and $\eta < -1$, which correspond to the frequencies $\omega < 2\alpha p_F - 2m\alpha^2$ and $\omega > 2\alpha p_F + 2m\alpha^2$ (see Fig. 2). Within the frequency range $2\alpha p_F - 2m\alpha^2 < \omega < 2\alpha p_F + 2m\alpha^2$, the plasmon attenuation is due to collisionless energy transfer to spin excitations. However, the wave attenuation is rather weak in this frequency range: $\mathrm{Im}k/\mathrm{Re}k \approx \lambda$ and the obtained spectrum can be considered exact.

In the same limit in the wave interaction region $\eta \ll 1/\lambda$, Eq. (28) can additionally be simplified by introducing the dimensionless variable $\xi = (k-4)/8\lambda$:

$$\ln\left(\frac{\eta-1}{\eta+1}\right) = 8(\xi - \eta). \qquad (29)$$

Equation (29) shows that the characteristic range of frequencies and momenta in which the plasmon spectrum is modified is determined, in order of magnitude, by the intersection of the plasmon dispersion range with that of spin transitions, being equal to $|\omega - 2\alpha p_F| \approx 2m\alpha^2$ and $(k-4) \approx 8\lambda$, respectively.

5. GENERATION OF PLASMONS

As is well known, the plasmon spectrum can be determined from the absorption of electromagnetic wave incident on the 2D structure. Since the wavelength in the free space at appropriate frequencies is large, the electric field of the wave is commonly spatially modulated with a grating structure [10] (see also review [11] and references therein). The power density absorbed by a 2D system is determined by the Fourier harmonics of the electromagnetic field in the 2D system plane $\mathbf{E}(\mathbf{q}, z = 0)$. For a 1D grating with grooves in the direction of the y-axis and spaced by d, $\mathbf{q}_n = (0, 2\pi n/d)$. These harmonics are linearly related to the zero-field harmonic $\mathbf{E}(\mathbf{q} = 0, z = 0)$ by the coefficients $c_\mathbf{q}$ determined solely by the grating properties:

$$\mathbf{E}(\mathbf{q}, z = 0) = \mathbf{E}(\mathbf{q} = 0, z = 0)c_\mathbf{q}. \qquad (30)$$

According to [11], the absorbed power can be written as

$$W = \frac{1}{2}|E_x(\mathbf{q} = 0, z = 0)|^2 \mathrm{Re}[\sigma_{ef}(\omega)], \qquad (31)$$

where

$$\sigma_{ef}(\omega) = \sum_\mathbf{q} \sigma(\mathbf{q}, \omega)|c_\mathbf{q}|^2, \qquad (32)$$

$$\sigma(\mathbf{q}, \omega) = \sigma(\omega)\left[1 + \frac{2\pi i q\sigma(\omega)}{\omega\kappa}\right]^{-1}. \qquad (33)$$

The zeros in the denominator determine the oscillation spectrum $\omega(\mathbf{q})$.

The frequency dependence of the absorption coefficient at different wave vectors is presented in Fig. 3. We use the following InAs/GaSb heterostructure parameters: $m = 0.055m_0$ and $\alpha = 9 \times 10^{-10}$ eV cm [12] with an electron mobility of $\mu = 10^6$ cm^2/(V s) (Fig. 3a) and $\mu = 3 \times 10^6$ cm^2/(V s) (Fig. 3b) and electron density $n_s = 7.5 \times 10^{11}$ cm^{-2}.

According to Fig. 3, there are two absorption peaks: narrow plasma-related and broad spin-related, with the spin resonance being significantly (by several orders of magnitude) weaker than the plasma-related one when

the resonances have radically different frequencies. The frequency of the absorption peak "follows" the plasmon spectrum $\omega(q)$. For clarity, only plots for the k values related to plasma resonances lying to the left of or within the spin transition range are given in the figure. If the wave vector determined by the diffraction grating exceeds $k_c = 4$, then plasma peaks are located symmetrically to the right of the range of spin-flip transitions.

When approaching the range of spin transitions, the plasma resonance significantly falls in magnitude and the intensity of the spin resonance increases. The oscillator strength swapping occurs because of the external field screening by the polarization of the medium, which is induced by spin transitions. When the resonance frequency falls within the range of spin transitions, the narrow plasma resonance tends to the boundary of this region in accord with the spin-plasmon spectrum in the collisionless limit (Fig. 2).

For the above InAs/GaSb heterostructure parameters, we have $\hbar\omega_0 = 3.91$ meV and $q_c = 2.44 \times 10^4$ cm^{-1}. It is noteworthy that an essential increase in the absorption coefficient due to spin transitions near the resonance with plasma oscillations makes the experimental observation of this effect easier.

ACKNOWLEDGMENTS

This study supported in part by the Russian Foundation for Basic Research (project nos. 99-02-17127 and 00-02-17658) and by the Russian Scientific Program "Physics of Solid-State Nanostructures."

REFERENCES

1. F. T. Vas'ko, Pis'ma Zh. Éksp. Teor. Fiz. **30**, 574 (1979) [JETP Lett. **30**, 541 (1979)].
2. Yu. A. Bychkov and É. I. Rashba, Pis'ma Zh. Éksp. Teor. Fiz. **39**, 66 (1984) [JETP Lett. **39**, 78 (1984)]; E. I. Rashba and V. I. Sheka, in *Landau Level Spectroscopy*, Ed. by G. Landwehr and E. I. Rashba (Elsevier, Amsterdam, 1991), p. 178.
3. R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
4. S. F. Edwards, Philos. Mag. **3**, 1020 (1958).
5. M. I. D'yakonov and V. I. Perel', Zh. Éksp. Teor. Fiz. **60**, 1954 (1971) [Sov. Phys. JETP **33**, 1053 (1971)].
6. E. L. Ivchenko, Yu. B. Lyanda-Geller, and G. E. Pikus, Zh. Éksp. Teor. Fiz. **98**, 989 (1990) [Sov. Phys. JETP **71**, 550 (1990)].
7. I. Meinel, D. Grundler, S. Bargstadt-Franke, *et al.*, Appl. Phys. Lett. **70**, 3305 (1997).
8. A. V. Chaplik, Zh. Éksp. Teor. Fiz. **62**, 746 (1972) [Sov. Phys. JETP **35**, 395 (1972)].
9. R. H. Ritchie, Phys. Rev. **106**, 874 (1957); R. A. Ferrel, Phys. Rev. **111**, 1214 (1958).
10. S. S. Allen, Jr., D. C. Tsui, and R. A. Logan, Phys. Rev. Lett. **38**, 980 (1977).
11. T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
12. J. Luo, H. Munekata, F. F. Fang, and P. J. Stiles, Phys. Rev. B **38**, 10 142 (1988); **41**, 7685 (1990).

Translated by S. Ktitorov

Semiconductors, Vol. 35, No. 9, 2001, pp. 1088–1094. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1135–1142.

LOW-DIMENSIONAL SYSTEMS

Germanium Quantum Dots in an Unstrained GaAs/ZnSe/Ge/ZnSe Heterosystem

I. G. Neizvestnyĭ, S. P. Suprun, A. B. Talochkin, V. N. Shumsky, and A. V. Efanov

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—Arrays of Ge quantum dots in unstrained GaAs/ZnSe/Ge heterostructures were obtained by molecular-beam epitaxy for the first time. Their spatial parameters are examined by scanning tunneling microscopy, and their electronic structure is studied by Raman spectroscopy. *© 2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Investigations of self-organized arrays of quantum dots (QDs) have attracted much attention in recent years. Experimental studies in this field are mostly carried out on the basis of heterostructures with a large lattice mismatch, such as GaAs/AlGaAs/InAs and Si/Ge [1–8]. In this case, spontaneously ordered nanostructures (equilibrium arrays of three-dimensional (3D) coherently strained islands [1] or atomic clusters [2]) can be obtained by molecular-beam epitaxy (MBE). Mechanisms of the Ge island formation were discussed previously [4].

Self-organization of the QD ensemble can take place both under equilibrium and nonequilibrium conditions. The thermodynamic approach is used to describe the first case, and the kinetic approach, the second. It is commonly believed that the self-organization process is driven by the elastic strain fields that appear due to the lattice mismatch between the substrate and the epilayer. The underlying assumption is that the formation of 3D coherently strained islands leads to a decrease in the elastic energy of the system.

Elastic strains in mismatched heterosystems affect the energy spectrum of the QDs, which may be distorted considerably. It is of interest to study the electronic structure of a system whose properties are not complicated by this effect. In this case, the spread in the spatial parameters and in the energy characteristics of the QD array can be compared directly. It also becomes possible to investigate the properties of the QDs composed of the same material but grown in different heterosystems to clarify the effects, first, of the energy structure of the crystalline matrix and, second, of the elastic strain itself on the QD energy spectrum.

In this paper, we report on the fabrication and investigation of the QD arrays in the GaAs/ZnSe/Ge/ZnSe heterosystem. Unlike QD arrays in the GaAs/AlGaAs/InAs and Si/Ge systems, the lattice mismatch is small in this case (less than 0.3%) and the self-organization mechanisms discussed above become inefficient. It should also be noted that undoped ZnSe layers possess a high resistivity, which is necessary in order to fabricate zero-dimensional (0D) objects electrically isolated from the substrate; thus, one may expect that this system will find applications in nanoelectronics.

1. Ge QUANTUM DOTS IN A GaAs/ZnSe/Ge HETEROSYSTEM: FABRICATION AND CHARACTERIZATION

1.1. Growth Technology

The structures under study were grown in an MBE unit with a residual pressure of 10^{-8} Pa; it is equipped with a high-energy electron diffractometer for the control of the surface structure in the reflection configuration (RHEED), a quartz gauge for measuring the thickness of the deposited layers, and a manipulator with a thermocouple to monitor the temperature directly at the substrate surface. A GaAs wafer with orientations (110) and (100) were used as substrates. After conventional chemical treatment, they were mounted onto a molybdenum holder using an indium–gallium eutectic. Thermal evaporators were used as Ge and ZnSe sources; the ZnSe compound was loaded into the source for growing the ZnSe layers [9]. Technological approaches to the growth of continuous submicrometer ZnSe and Ge layers and Ge/ZnSe superlattices on GaAs substrates, as well as their properties, were reported previously [10–12].

After cleaning the GaAs substrate (an atomically clean surface is obtained), the temperature was reduced to 230°C and a ZnSe epilayer with Se-stabilized 2×1 surface superstructure was grown. Subsequently, an array of Ge QDs was formed on top of the ZnSe layer; the following two procedures were used.

In the first case, the substrate heating is switched off for 2 h and the surface temperature decreases to 30°C. After that, a Ge layer of thickness from 1 to 7 nm is deposited. A RHEED pattern typical of the amorphous

1063-7826/01/3509-1088$21.00

material is observed at this stage. Next, the substrate temperature is increased at a certain rate. As the temperature increases, the surface Ge layer transforms from the amorphous to the polycrystalline phase; then, at $T \approx 200$°C, bulk reflections appear and, at $T \approx 300$°C, stretch into streaks. To obtain an array of islands, the heating should be switched off at the appearance of the bulk reflections to allow the sample to cool to room temperature. If the heating is continued, a 2×2 surface structure is finally formed at $T \approx 320$–350°C.

The time needed for the transformation from one type of surface structure to another depends on the average thickness of the Ge layer on ZnSe. By varying the angle of incidence of the diffractometer electron beam on the sample surface, it was found that the solid-state recrystallization of the submicrometer Ge layer starts from the heterointerface.

In the second case, the QD array is formed at the same time as Ge deposition at a temperature of 100°C and below. In both methods, 3–4 Ge layers of different thicknesses can be grown on a single substrate with the use of a movable shutter.

The grown samples were studied by Raman spectroscopy and scanning tunneling microscopy (STM); the STM images, obtained with a RIBER–OMICRON microscope, were processed using computer graphics and statistical analysis programs. To prevent oxidation, the only surface of those samples prepared for the Raman measurements was covered by a protective ZnSe layer immediately after the QD formation. The structures prepared for STM were grown on a thin tunneling-transparent ZnSe layer and were selected by their current–voltage characteristics.

1.2. STM Analysis

In Fig. 1a, an STM image of a QD array on a (110)-oriented substrate is shown. The islands were formed during Ge deposition; the average Ge layer thickness, measured during the growth by a calibrated quartz gauge, equals 2 nm. One can see that the surface contains an array of islands; the height profile along the direction indicated by the straight line in Fig. 1a is shown in Fig. 1b.

It is well known that the zero level in the profile obtained by STM is taken to be at the depth corresponding to the maximum deflection of the tip. For a system characterized by a low island density and a low degree of the surface coverage by the epilayer material, there is a high probability that the tip will pass over the substrate surface and, consequently, that the zero level thus determined is valid. In the case under investigation, the island density is $\sim 2.5 \times 10^{11}$ cm^{-2} (see Fig. 1) and the validity of the zero-level determination should be considered more carefully. To test the correctness of the zero level position, one can compare the average thickness of the deposited layer calculated from the STM profile with that measured experimentally. If the ratio

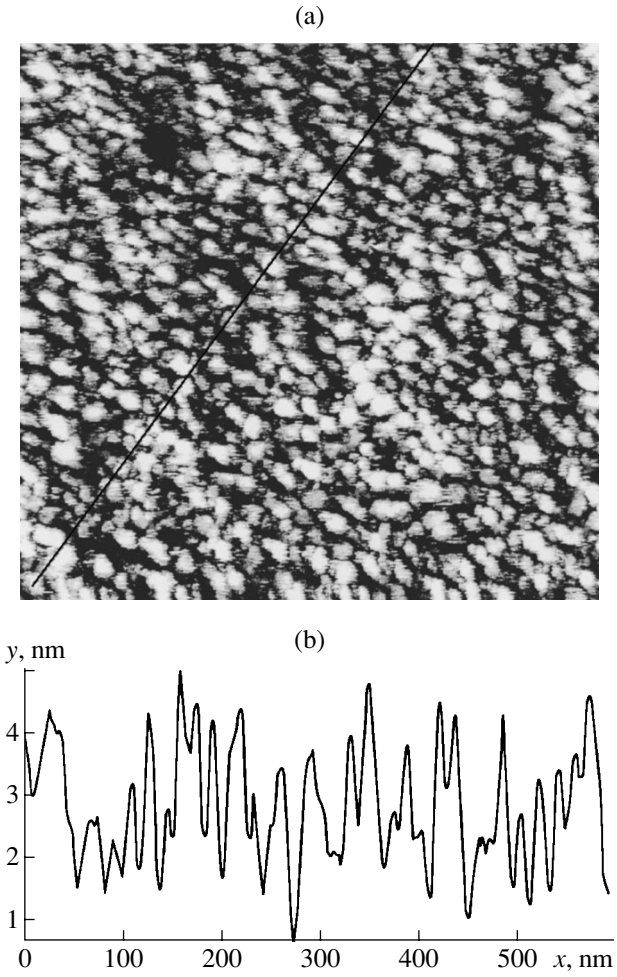


Fig. 1. (a) An STM image of a quantum dot array for the (110)-oriented sample with the average Ge layer thickness of 2 nm. (b) The island height distribution along the line shown in Fig. 1a.

of these two values is lower than unity, this means that there is a continuous sublayer and the zero level is shifted to coincide with its surface. If this ratio is higher than unity, the continuous sublayer does not exist and the most probable reason for such a discrepancy is the following. In the STM measurements, the height of a small protrusion over a nearly flat surface is determined with maximum accuracy, but the depth of a small pit between the islands is determined with a significant error caused by the relatively large diameter of the tip and the appearance of peripheral parasitic currents over the tip circumference.

The statistical analysis of the surface profiles taken along different directions yielded an average layer thickness of 2.5–3.0 nm, well in excess of the value determined experimentally using the quartz gauge (2.0 nm); at the same time, the minimum height in the obtained profiles is not equal to zero. This means that there exists an array of separate islands, and the depth of the valleys between them is not determined accu-

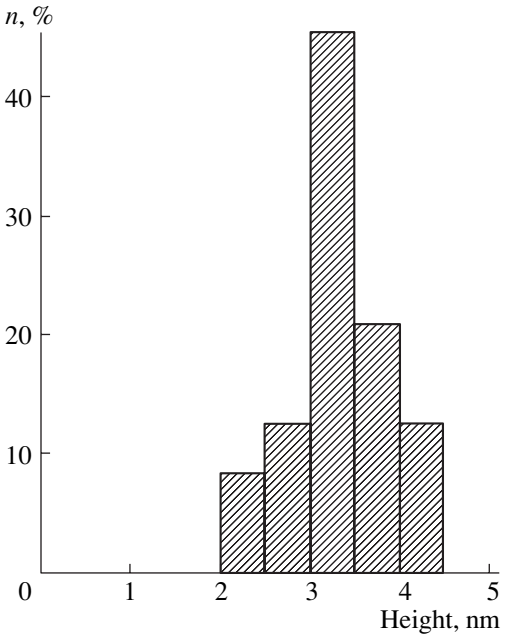


Fig. 2. Distribution of the island heights in a quantum dot array for the sample with the average Ge layer thickness of 2 nm.

rately by the STM technique due to the above reasons. This conclusion is corroborated by the Raman scattering data: the spectra typical of isolated 0D objects (i.e., QDs) are obtained for this sample.

To estimate the size uniformity of the QDs in the ensemble, the distributions of the QDs over their height and area were determined. The island height was determined by the maximum values obtained by profiling (by the reasons explained above, the height of the peaks is determined by the STM technique with the minimum error).

The height distribution for the QD array with the average Ge layer thickness of 2.0 nm is shown in Fig. 2. One can see that the height of nearly half of the islands falls within the range 3.0–3.5 nm, which is evidence of their high uniformity. Similar distributions were obtained for other coverage values; the average height was observed to increase with the coverage.

The distribution of the QDs over their areas was also obtained from the QD array STM images. The distribution is asymmetric: it has a sharp edge from the side of low values and a smooth tail from the side of large values. It was established that the average area of the island pedestal increases and the distribution broadens with increasing coverage. According to the STM data, the average density of the islands varies from 2.5×10^{11} to 2.0×10^{11} cm^{-2} for the mean Ge epilayer thickness increasing from 2 to 4 nm. In the latter case, at the reduced island density, the calculated mean thickness of the epilayer (3.8–4.3 nm) agrees well with the correct value. This coverage corresponds to the onset of the island merging, which is confirmed by the Raman

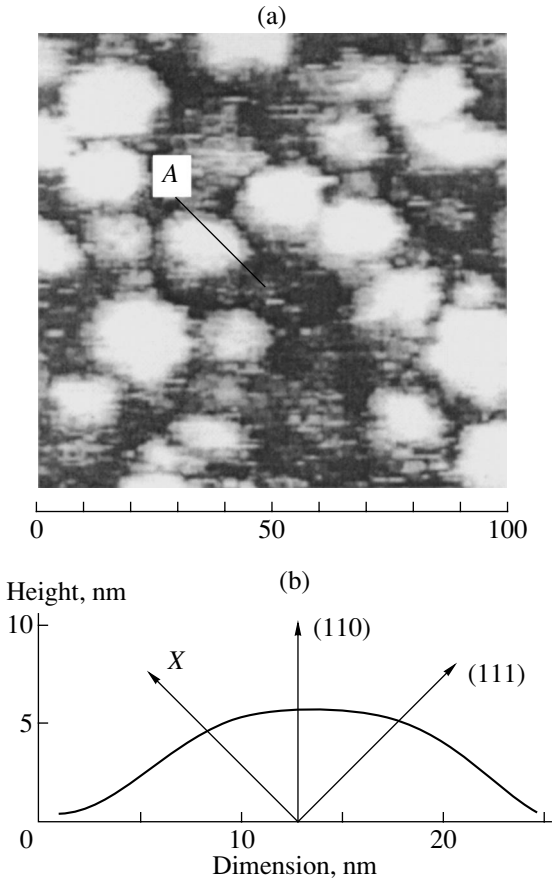


Fig. 3. (a) An STM image of a 100×100 nm^2 area on the surface of a sample with Ge quantum dots. (b) The profile obtained along the line A.

scattering data. For a mean epilayer thickness of 4 nm, the predominant orientation in the spatial arrangement and the shape of the nuclei can be clearly seen: they are elongated in the [110] direction.

STM images of the Ge layers deposited on the (100)-oriented substrate were obtained for the average layer thicknesses of 1.7, 3.4, and 5.1 nm. In this case, a somewhat different picture was observed. The main difference is that, for the same coverages, the measured island density for (100) orientation is higher than that for the (110) orientation and lies in the range from 7×10^{11} to 9×10^{-11} cm^{-2}. The typical island size is smaller than that for the (110) substrate orientation and increases slightly with increasing coverage; viewed from the top, the islands are round.

The shape of the QDs was determined by the STM. Fig. 3a represents an STM image with a 100×100 nm^2 area with an effective Ge layer thickness of 34 Å, obtained at the tunneling current of 0.2 nA. In Fig. 3b, the profile recorded along the line A is plotted. One can see that the QDs have a droplike shape, and there is no faceting at the surfaces. The typical QD size in the growth plane amounts to 20 nm, and the height is ~5 nm.

1.3. Electronic Structure of Ge Quantum Dots in ZnSe Matrix

Optical-phonon Raman scattering spectra of the grown GaAs/ZnSe/Ge/ZnSe structures were studied at temperatures of 300 and 77 K. The spectra were excited by discrete lines of an Ar laser and recorded with a DFS-52 spectrometer. The lines of longitudinal (LO) and transverse (TO) optical phonons originating from the GaAs substrate (LO at 294 cm^{-1} and TO at 268 cm^{-1}) and the ZnSe matrix (LO at 250 cm^{-1} and TO at 226 cm^{-1}) were observed as well as the Ge line (LO + TO) at 300 cm^{-1} scattering. The position of the Ge line and the scattering intensities measured in different polarization configurations coincide with those of bulk Ge. This indicates that there is no mechanical strain in the Ge layer, and its misorientation with respect to the crystallographic directions of the substrate is insignificant. A considerable difference in the Ge and ZnSe phonon frequencies results in a strong localization of phonons inside a QD. This makes it possible to study electron states in the dots by resonance optical-phonon Raman scattering.

The resonance Raman scattering by Ge optical phonons in the structures under consideration was studied in the excitation photon energy range from 2.4 to 2.7 eV. The dependences obtained at $T = 300$ K for three different samples are shown in Fig. 4 (curves *a*, *b*, *c*). Experimental points corresponding to the different lines of Ar laser are connected by straight lines. Note that the scattering intensity is normalized to the scattering volume in each sample. Samples *a* and *b* are grown on the (100)-oriented surface, in which case the diameter of the Ge QDs in the growth plane is 7–10 nm and QD height is 2.3–2.8 nm. The two structures differ in the Ge layer effective thickness (3.4 nm for sample *a* and 2 nm for sample *b*). Sample *c* is grown on the surface with orientation (110), and the size of QDs here is larger by a factor of 2; the effective Ge layer thickness is 3.4 nm. For comparison, the resonance dependence obtained in the region of E_1, $E_1 + \Delta_1$ transitions in the bulk Ge [13] is also plotted in Fig. 4; the peak of this curve is also normalized to the scattering volume. The energies of the transitions E_1 and $E_1 + \Delta_1$ in Ge are shown by vertical arrows. One can see that resonance position in the QDs is shifted to higher energies with respect to the bulk, and its amplitude for sample *a* is three times higher. It should be noted that, with increasing effective thickness *d* of the Ge layer, the QDs merge into a continuous layer at $d = 5$–7 nm. In this case, the intensity of Raman scattering in the region of excitation photon energies 2.4–2.7 eV drops sharply, and the shape of the resonance dependence coincides with that of the bulk Ge. These observations are related to the quantization of the spectrum of the QD electron and hole states that are coupled by direct optical transitions under the conditions of E_1, $E_1 + \Delta_1$ resonance.

Let us consider the nature of the resonance in the Ge band structure shown in Fig. 5. The dispersion curves

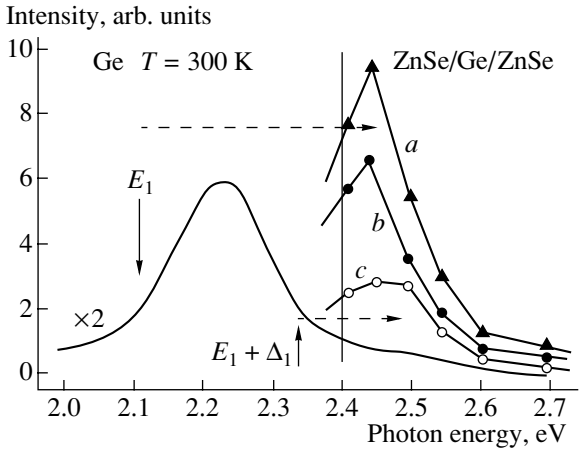


Fig. 4. Intensity of the resonance Raman scattering by Ge optical phonons in GaAs/ZnSe/Ge/ZnSe samples with Ge quantum dots in the excitation-photon energy range 2.4–2.7 eV. Samples *a* and *b* are grown on (100)-oriented substrates, and sample *c* is grown on a (110)-oriented substrate. A curve in the left-hand part of the figure shows the optical-phonon Raman scattering resonance in bulk Ge [13]. The curve labels *a*, *b*, and *c* correspond to the sample notation.

along the direction $k_z \parallel (111)$ for the conduction band (states Λ_1) and for both branches of the valence band (states $\Lambda_{4,5}$ and Λ_6) are parallel over most of the Brillouin zone. In the plane (k_x, k_y), perpendicular to (111), the reduced effective mass for the transitions E_1 and $E_1 + \Delta_1$ (which are shown by vertical arrows in Fig. 5) is $m = (1/m_e + 1/m_h)^{-1} = 0.045m_0$ [14] (here, m_0 is the free electron mass and m_e and m_h are electron and hole effective masses, respectively). For such a band structure, there is a two-dimensional (2D) critical point in the interband density of states [15], which is represented in the optical spectra. Thus, optical-phonon Raman scattering is enhanced for the excitation photon energy in the range of the transitions E_1, $E_1 + \Delta_1$ (see Fig. 4). The main contribution to the observed resonance comes from the three-band processes, where the hole states $\Lambda_{4,5}$ and Λ_6 are mixed by the optical-phonon deformation potential [13]. The two-band contributions, related to the electron–hole transitions with phonon emission within a single band, are much smaller due to the low value of the corresponding deformation potential [13]. The shape of the resonance dependence is given by the expression $|\varepsilon_+ - \varepsilon_-|^2$, where ε_+ and ε_- are the contributions of the transitions E_1, $E_1 + \Delta_1$ to the dielectric constant and the resonance peak occurs between the two transition energies (see Fig. 4). The contribution of the two-band processes to the Raman scattering is proportional to $\sim d\varepsilon/d\omega$ for each of the transitions [13], and the resonance profile contains two peaks at energies E_1 and $E_1 + \Delta_1$.

One can see from Fig. 5 that most of the Ge conduction and valence band states with $k \parallel (111)$ fall within the ZnSe band gap (the energies of ZnSe conduction

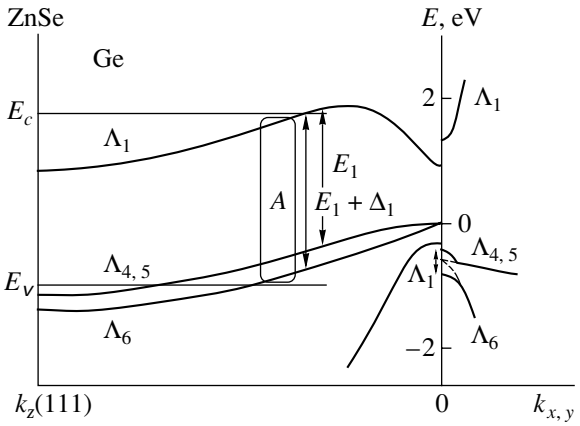


Fig. 5. Calculated band spectrum of bulk Ge along the ⟨111⟩ direction [10]. The arrows indicate the direct optical transitions related to the E_1, $E_1 + \Delta_1$ Raman resonance. The dispersion curves for electron and hole bands in the plane (k_x, k_y) (perpendicular to (111)) in the energy range of E_1, $E_1 + \Delta_1$ transitions are shown to the right of the vertical axis.

and valence band edges, E_c and E_v, respectively, are shown by horizontal lines). For these Ge states, a localizing potential barrier exists, whose height is equal to the energy spacing between these levels and the corresponding ZnSe band edge. This results in the quantization of the electron spectrum in the Ge QDs. Quantization of the carrier motion along the (111) direction does not lead to the shift in the interband transition energies, since the electron and hole dispersion curves are parallel in the region of the transitions E_1, $E_1 + \Delta_1$ (see Fig. 5). Quantization in the perpendicular plane (k_x, k_y) leads to an increase in the energies of the lowest electron and hole states. The energies of the transitions E_1 and $E_1 + \Delta_1$ increase, with the shift being equal to the sum of electron and hole contributions. In bulk Ge, the interband density of states, which determines the resonance profile, is spread in energy over the interval Δ_1; in the QDs, it transforms into a δ-function for all k_z. It follows from the shape of the ε(ω) dependence for the case of the discrete spectrum [15] that the two-band term becomes dominant and exceeds the contribution from the three-band processes by a factor of $\sim(\Delta_1/\Gamma)^4 \approx 10^4$, where $\Gamma \approx 20$ meV is the broadening of the transition in bulk Ge. As a result, the Raman resonances at E_1 and $E_1 + \Delta_1$ should manifest themselves as separate peaks. We attribute the resonance observed in sample *c* (see Fig. 4) to the transition $E_1 + \Delta_1$ within the QDs. It is shifted by 0.13 eV from the corresponding bulk position; the shift is indicated by a horizontal arrow in Fig. 4. In Fig. 5, the region of the *k*-space related to the $E_1 + \Delta_1$ transitions in the QDs is indicated. With a decrease in the QD size, the electron (hole) states are pushed out into the ZnSe continuous spectrum, which leads to a sharp reduction in the Raman scattering intensity. When all of the states from region *A* (Fig. 5) are in the continuous spectrum, the resonance $E_1 + \Delta_1$ no longer manifests itself. Thus, for the QDs in samples *a* and *b*, whose size is half of that in sample *c*, the resonance in the energy range considered is related to the E_1 transition (see Fig. 4, curves *a*, *b*). Its shift, marked by a dashed arrow, is equal to 0.34 eV. The intensity of the E_1 resonance exceeds that of the $E_1 + \Delta_1$ resonance by a factor of three due to larger *k*-space volume (see Fig. 5). Thus, the modification of the interband density of states caused by the quantization of the energy spectrum results in a change in the optical-phonon Raman scattering mechanism and in an increase in the scattering intensity in the QDs in comparison to the bulk material.

Next, let us consider the energies of electron and hole states in the Ge QDs. The cross section of a Ge island by a plane perpendicular to the base and containing the ⟨111⟩ direction is shown in Fig. 3b. The directions of the wave vectors $k_z \parallel (111)$ and $k_x \perp (111)$ in the plane of the drawing are also shown, and k_y is perpendicular to this plane. For the case of the quadratic dispersion law, it is conventional to consider the quantization of the electron spectrum using the envelope wave function approximation. In the case under consideration, the dispersion along $z \parallel (111)$ is described by a nearly linear function of the momentum, and this approximation cannot be used. Qualitatively, the effect of the QD size on the motion along the direction $z \parallel (111)$ is that the momentum component of electron (hole) states in this direction is quantized. The spectrum consists of a number of discrete levels with $k_z = \pi n_1/\sqrt{2}H$, where n_1 is an integer, H is the island height, and $\sqrt{2}H$ is the island dimension along the (111) direction. The spacing between the levels can be estimated as $\Delta E_w = (dE/dk_z)k_z \approx 0.1$ eV. The direct transitions between the electron and hole states with the same k_z (i.e., the same n_1) contribute to the observed resonance, and the quantization along (111) does not change the resonance spectral position. The quantization of electron and hole motion in the (x, y) plane can be described using the envelope wave functions, since the energy spectrum is quadratic. One can see from Fig. 3b that the size of the QD cross section by the plane (x, y) varies insignificantly near the top due to the droplike shape of the island. Let us approximate the shape of this cross section by a rectangle with the sides $\sim\sqrt{2}H$ and $\sim d/2$, where d is the size of the QD base. The energy of a particle of mass m in a 2D potential well of infinite depth is given by

$$E = (\pi^2\hbar^2/2m)(n_2^2/2H^2 + 4n_3^2/d^2), \qquad (1)$$

where n_2 and n_3 are integers. Thus, for each state with a given $k_z(n_1)$, a system of levels appears due to the 2D quantization. The energy shift of the ground state ($n_2 = n_3 = 1$) is given by

$$\Delta E = (\pi^s\hbar^2/2m)(1/2H^2 + 4/d^2). \qquad (2)$$

Substituting $m = (1/m_e + 1/m_h)^{-1}$ into this expression, we obtain the shift of the E_1, $E_1 + \Delta_1$ resonance energy; both electron and hole energy shift due to quantum confinement are taken into account. For the QDs in sample *c* ($H \approx 5$ nm), the calculated value $\Delta E = 0.163$ eV is in agreement with the observed shift of the $E_1 + \Delta_1$ resonance energy (Fig. 4, curve *c*). For samples *a* and *b*, the average QD height $H \approx 2.5$ nm. The energy shift defined by Eq. (2) is $\Delta E = 0.66$ eV, while the observed shift of the E_1 resonance is 0.34 eV (Fig. 4, curve *a*). This discrepancy is explained by special features of the hole dispersion in Ge. The dispersion of the valence subbands and the conduction band in the plane (k_x, k_y) (perpendicular to (111)) in the energy region of E_1, $E_1 + \Delta_1$ transitions is shown in Fig. 5. The dotted lines represent the dispersion of the hole subbands calculated by the (**kp**) method with the spin–orbit coupling neglected [16]. Consideration of this coupling in a 2D Kane-type model [17] leads to the lifting of the degeneracy at $k = 0$, with both resulting subbands in the coupling region being characterized by the effective mass $m_h = 2m_e$ (where m_e is the effective mass in the conduction band (Λ_1) in the plane (k_x, k_y)). It can be seen from Fig. 5 that the dispersion curve of the upper valence band bends at the energy $\Delta_1/2 = 0.115$ eV from its top in such a way that the effective mass increases; the reduced effective mass $m = (2/3)m_e$ corresponding to the transition E_1 increases by a factor of 1.5, and the quantum confinement energy is now determined by the electron mass m_e. Taking these considerations into account, we obtain $\Delta E = 0.47$ eV. Thus, using a very simple confinement model and taking into account the features of the bulk Ge electronic spectrum, it is possible to explain the observed energy shifts of the E_1 and $E_1 + \Delta_1$ transitions in the Ge QDs.

2. EQUILIBRIUM STATE IN A SYSTEM OF UNSTRAINED 3D ISLANDS

Let us consider a simple model demonstrating that the formation of an island array is possible in unstrained heterostructures. According to [1], the terms contributing to the total energy of the system are the energy of the formation of the new phase (an island), the energy of the elastic strains between the deposited material and the substrate, and the energy of the island faceting. Interaction between the elastic fields of the islands results in their repulsion, which leads to self-organization. The existence of a minimum in the total energy upon the transition of the film from a continuous layer to an array of islands is a necessary condition for the formation of such a structure; the abruptness of this minimum determines, in general, the spread in the island sizes.

Similar to [1], the suggested model implies that the material from a certain portion of the continuous film of area $L \times L$ cm^2 is collected into an island of area $d \times d$ cm^2. Next, it is assumed that the elastic relaxation energy, faceting energy, and the island interaction energy can be neglected. Thus, only the term related to the change in the surface energy is retained in the expression that describes the variation of the total energy of the film upon the transition from the continuous layer to the island array. Although this assumption is definitely incorrect in the general case, experimental data confirm its validity for the case under consideration. The elastic relaxation energy is very low due to the small value of the lattice mismatch. The islands observed were not faceted; consequently, the corresponding energy term can be dropped. The absence of faceting is, apparently, related to the growth conditions, specifically, to the low substrate temperature. Under these conditions, the adatom kinetic energy is insufficient to form an ordered facet plane. Finally, the high density of the nuclei indicates that the island repulsion is small, which means that the interaction energy can also be neglected. Thus, it was assumed that the variation in the film energy upon the transition from the continuous layer to the island array originates only from the surface changes in the system.

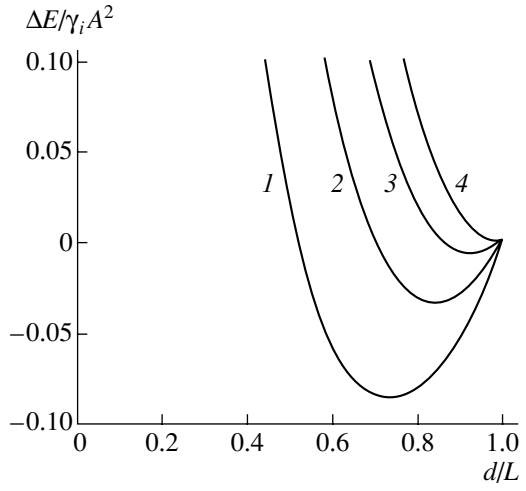


Fig. 6. Dependence of the change in the heterosystem specific surface energy on the island linear dimensions. It is assumed that the material is collected into an island from an area $L \times L = 20 \times 20$ nm^2 and $\beta = (\gamma_S - \gamma_{i-S})/\gamma_i = 0.5$. The average Ge layer thickness δ is (*1*) 2, (*2*) 3, (*3*) 4, and (*4*) 5 nm.

Then, the surface energy can be written as

$$E_{\text{surf}} = \gamma_i S_i + \gamma_{i-S} S_{i-S} + \gamma_S S_S.$$

Here, the subscripts *i*, *i–S*, and *S* refer to the film, the film–substrate interface, and the substrate, respectively; γ is the specific energy for the corresponding surface; and *S* is the surface area.

For a film of total area $A \times A$ cm^2, the energy difference between the island array and the continuous layer is given by

$$\Delta E' = \frac{\Delta E}{\gamma_i A^2}$$
$$= \left(\frac{d^2}{L^2} - 1\right)\left(1 - \frac{\gamma_S - \gamma_{i-S}}{\gamma_i}\right) + 4\delta\left(\frac{1}{d} - \frac{1}{A}\right),$$

where δ is the average film thickness.

Obviously, the specific energy difference per unit area is obtained by substituting $A = 1$. Numerical calculations show that, for certain values of the parameters in the above expression, the dependence of the specific energy difference ΔE on the island size has a minimum, and $\Delta E_{\min}$ is negative (Fig. 6). This means that the island structure is preferable.

CONCLUSION

In this study, Ge QD arrays are obtained for the first time in an elastically unstrained GaAs/ZnSe/Ge heterosystem. Their spatial parameters and electronic spectra are investigated. It is shown that the structures are characterized by high island density and lack island faceting. The mechanism for the formation of such arrays is governed, first, by the technological conditions (actually, by the low temperature of the epitaxy, which causes low surface diffusion coefficient) and, second, by the special features of the materials used—specifically, by the low energy of the Ge–ZnSe bond in comparison to the surface energy of Ge. The latter conclusion is corroborated by the possibility of forming the QD array by the recrystallization of amorphous Ge under the conditions of increasing substrate temperature.

The Raman scattering by Ge QD optical phonons near the E_1, $E_1 + \Delta_1$ resonance was investigated. It is shown that the modification of the interband density of states in the QDs leads to a change in the phonon scattering mechanism and to an increase in the resonance amplitude as compared to the bulk Ge. Application of a very simple model of quantum confinement taking into account the features of the Ge electronic spectrum makes it possible to explain the observed energy shifts of the E_1, $E_1 + \Delta_1$ resonance in the QDs.

ACKNOWLEDGMENTS

We are grateful to I.G. Kozhemyako for the STM images of the surfaces of the samples. This study was supported by the Russian–Ukrainian Program "Nanophysics and Nanoelectronics" and, in part, by the Russian Foundation for Basic Research (project no. 98-02-17938).

REFERENCES

1. N. N. Ledentsov, V. M. Ustinov, V. A. Shchukin, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 385 (1998) [Semiconductors **32**, 343 (1998)].
2. A. I. Yakimov, V. A. Markov, A. V. Dvurechenskii, and O. P. Pchelyakov, Philos. Mag. B **65**, 701 (1992).
3. A. V. Dvurechenskiĭ and A. I. Yakimov, Izv. Akad. Nauk, Ser. Fiz. **63**, 306 (2000).
4. A. V. Rzhanov and S. I. Stenin, in *Growth of Semiconductor Crystals and Films* (Nauka, Novosibirsk, 1984), Part 1, p. 5.
5. M. Grundmann, J. Christen, N. N. Ledentsov, *et al.*, Phys. Rev. Lett. **74**, 4043 (1995).
6. I. L. Krestnikov, M. V. Maximov, A. V. Sakharov, *et al.*, J. Cryst. Growth **184/185**, 545 (1998).
7. N. N. Ledentsov, V. A. Shchukin, M. Grundmann, *et al.*, Phys. Rev. B **54**, 8743 (1996).
8. M. Iwamatsu and Y. Okabe, J. Appl. Phys. **86**, 5541 (1999).
9. I. G. Neizvestnyĭ, S. P. Suprun, and V. N. Shumskiĭ, Avtometriya, No. 4, 34 (1994).
10. K. S. Zhuravlev, A. É. Klimov, I. G. Neizvestnyĭ, *et al.*, in *Proceedings of the All-Russia Scientific and Technical Conference on Microelectronics and Nanoelectronics, Zvenigorod, 1999*, Abs. P2-18.
11. S. P. Suprun, I. G. Neizvestny, V. N. Sherstyakova, and V. N. Shumsky, Phys. Low-Dimens. Struct. **1**, 67 (1995).
12. S. P. Suprun, A. B. Talochkin, A. K. Gutakovsky, and V. N. Shumsky, Phys. Low-Dimens. Struct. **1**, 59 (1995).
13. F. Cerdeira, W. Dreybrodt, and M. Cardona, Solid State Commun. **10**, 591 (1972).
14. K. L. Teo, S. H. Kwok, P. Y. Yu, and Soumyendu Guha, Phys. Rev. B **62**, 1584 (2000).
15. M. Cardona and G. Gunterodt, in *Light Scattering in Solids*, Ed. by M. Cardona and G. Gunterodt (Springer-Verlag, Berlin, 1982; Mir, Moscow, 1984), Vol. II.
16. F. Pollak and M. Cardona, Phys. Rev. **142**, 530 (1966).
17. G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Nauka, Moscow, 1972; Wiley, New York, 1975).

Translated by M. Skorikov

Semiconductors, Vol. 35, No. 9, 2001, pp. 1095–1105. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1143–1153.

LOW-DIMENSIONAL SYSTEMS

Type-II Ge/Si Quantum Dots

A. V. Dvurechenskii* and A. I. Yakimov

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: dvurech@isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—The electronic structure of spatially indirect excitons, multiparticle excitonic complexes, and negative photoconductivity in arrays of Ge/Si type-II quantum dots (QDs) are considered. A comparison is made with the well-known results for type-II III–V and II–VI QD heterostructures. The following fundamental physical phenomena are observed in the structures under study: an increase in the exciton binding energy in QDs as compared with that for free excitons in homogeneous bulk materials, a blue shift in the excitonic transitions during the generation of multiparticle complexes (charged excitons, biexcitons), and the capture of equilibrium carriers to localized states induced by the electric field of charged QDs. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The relative position of energy bands on both sides of a heterointerface in a heterosystem is determined by the structure and composition of the constituent semiconductor materials [1]. If the lowest energy states of electrons and holes in the band diagram are located in one of the two materials forming the system, the system is referred to as a type-I heterostructure. If the lowest energy state for electrons is associated with one material, and that for holes, with the other, the system is of type II. Therefore, in type-II heterostructures with two heterojunctions, a potential well can be formed only for one type of carrier, electron or hole with a barrier existing for the other type of carrier. In the literature, the band diagram of such a structure is often staggered, implying similar energy steps in conduction or valence band discontinuities: ascending or descending energy steps at the heterointerface for each of the bands.

In clusters of a material introduced into the bulk of another material, the carrier motion is confined in all three directions. If the cluster size is comparable with the de Broglie wavelength of an electron or hole, or with the Bohr radius of an exciton, then the inclusions are referred to as quantum dots (QDs) [2], and semiconductor structures with such clusters are called QD heterostructures [1]. In contrast to quantum-well (QW) and quantum-wire heterostructures (two-dimensional and one-dimensional systems), the properties of electrons and holes in QD heterostructures cannot be described as a gas of quasiparticles. A fruitful concept in this case is that of localized states. The electron or hole localization radius in a nanocluster is comparable with the cluster size, and it frequently exceeds the Bohr radius of single impurity atoms with shallow levels in homogeneous bulk semiconductors. On the other hand, the energy level in a QD may be deep, and this is one more feature of QDs as deep-level impurity centers. Owing to specific features mentioned above, the study of QD heterostructures now constitutes a separate branch of condensed-matter physics.

Charge localization in a QD modifies the potentials in the neighboring space. This gives rise to a potential well for carriers of opposite sign around a QD and leads to the formation of bound states in this well. In type-II heterostructures, localized states for holes and electrons emerge in self-consistent potential wells on different sides of the heterointerface. Electrons and holes in these states are spatially separated, and the transition between these states is spatially indirect.

Finding the energy spectrum parameters, analyzing the kinetics of transitions between electron states and the interaction of elementary excitations, and revealing the correlation effects constitute the main part of the contemporary fundamental studies of QDs. The utmost attention is attracted by $In_xGa_{1-x}As/Al_yGa_{1-y}As$ systems owing to the prospects for their application in semiconductor lasers (see, e.g., reviews [1, 3] and references therein). Type-II QDs formed in (In,Ga,Al)Sb/GaAs, (Al,In,As)/InP, (Ga,In,As,Sb)/GaSb, Ge/Si, and ZnTe/ZnSe heterostructures are less understood [4–11]. Heterostructures with compound semiconductor QDs are of particular importance for optoelectronics. Ge/Si heterosystems with QDs open up new prospects for micro- and nanoelectronics [12–15].

The electronic processes in QD systems have mainly been studied by optical methods [1, 3, 5, 6]. Our investigations were the first to apply electrical methods to reveal the discrete energy spectrum in an array of self-organized QDs and the Coulomb interaction in charge transport and to determine the carrier capture cross sections of localized states [4, 13, 15, 16]. The existence of a wealth of data concerning the surface and

1063-7826/01/3509-1095$21.00

the phase boundaries for Ge and Si [17], as well as the possible application of the developed techniques in modern silicon technology of discrete devices and circuits, makes investigating the Ge/Si system a priority. In this paper, we present data obtained in studying electronic processes, mainly in the Ge/Si system, by means of a set of optical and electrical techniques. The results are compared with the available data for other QD systems.

2. OBJECTS AND METHODS OF STUDY

Presently, the most promising method to produce an array of QDs is based on the self-organization of semiconductor nanostructures in heteroepitaxial systems [1, 5, 14]. Elastic strains in an epitaxial film and islands on its surface are crucial both for a morphological transition from a flat film to island growth (Stranski–Krastanow mode) and for further changes in the size, shape, and spatial distribution of the islands. An important stage in the sequence of the occurring kinetic transitions is the formation of coherent (defect-free) three-dimensional (3D) islands with a uniform size, which gives nanometer islands with quantum confinement energy of about 100 meV [13, 15]. This value noticeably exceeds the room-temperature thermal energy of particles (26 meV); therefore, the thermal redistribution of carriers over localized states within an energy window on the order of k_0T (k_0 is the Boltzmann constant, T, the temperature) can be disregarded. These systems enable the room-temperature operation of QD structures exhibiting temperature-insensitive device characteristics over a wide temperature range [18].

We have found the conditions of Ge heteroepitaxy onto (110) Si that provide a 3×10^{11} cm^{-2} sheet concentration of clusters [14]. The size distribution of germanium islands was studied by STM. The average size of islands (pyramid base) was 15 nm, and the pyramid height was 1.5 nm, with deviations being no more than 17%. Typical conditions of Ge/Si structure formation included homoepitaxy at 800°C onto (100) Si at a rate of 1–2 monolayers per second (ML/s); Ge heteroepitaxy at 300°C, 0.2 ML/s; and Si epitaxy over the Ge islands at 500°C. The effective thickness of the Ge layer was varied within d_{eff} = 0–20 ML. In Ge/Si QD heterostructures, holes are localized in Ge nanoclusters. Nonuniform structural strains (resulting from a 4% lattice mismatch between Ge and Si) and the positive charge created in Ge by holes can induce a potential well for electrons in the conduction band of silicon in the vicinity of the heterojunction.

3. SPATIAL DISTRIBUTION OF ELASTIC STRAINS

Nonuniform elastic strains in heterostructures can cause significant changes in the electrical and optical properties as a result of the energy-spectrum modification by about 0.1 eV [19, 20]. Furthermore, nonuniform strains favor the spatial ordering of nanoclusters during the formation of multilayer structures [21]. Therefore, determining the fields of elastic strains is a necessary step in both calculating the band structure of self-arranged QDs and modeling the epitaxy on strained surfaces. The elastic strain fields in Ge nanoclusters and in their environment were calculated by means of an original method relying upon Green's tensor of the elastic problem [22]. The sizes of the QDs studied are so small that the continuum approximation is inapplicable to a description of the elastic properties of the system. The strain was calculated using the Keating [23] potential, taking into account the atomic structure of the substance. To reduce the calculation error introduced by the finite crystal volume, we sought the deformational field as a convolution of an auxiliary function with the Green's function (Green's tensor) of the elastic atomistic problem. This method yields the distribution of strain at the atomic level for a system containing inclusions of one material in the matrix of another. The crystal anisotropy and the different elastic properties of the medium with inclusions of another phase are taken into account.

The problem was solved for the objects studied in our experiments: germanium QDs had the form of tetrahedral pyramids with an 11° slope on their lateral faces, a base size of 15 nm (in the growth plane), and a height of 1.5 nm. A QD sheet located on top of a 0.7-nm-thick wetting Ge layer was embedded in bulk crystalline silicon. The solution was obtained for a single QD, which means that the superposition of elastic strain fields from the array of surrounding QDs was disregarded.

It was established that the strain inside a Ge nanocluster is tensile in the direction of structure growth (z direction) and compressive in the lateral direction (xy plane) (Fig. 1). The highest strain exists along the outline of the pyramid base in the growth plane, and the most relaxed is in the vicinity of the apex. The highest strain in the silicon environment of a QD is located in the vicinity of the pyramid apex. Calculations show that the strain in the central part of the pyramid is practically independent of the Ge nanocluster size for the pyramid base within 6–15 nm. Near the edge of the pyramid base, the strain grows as a logarithm of the base size.

Recently, results obtained in calculating the strain in the Ge/Si QD system with the use of two empirical potentials—Keating and Stillinger–Weber—have been compared [24]. Both methods yield similar results for the lateral tensor components, with a quantitative difference found for the normal tensor components. As a result, the authors of [24] recommended that the Stillinger–Weber potential should be used for Ge nanoclusters with a pyramid base smaller than 10 nm.

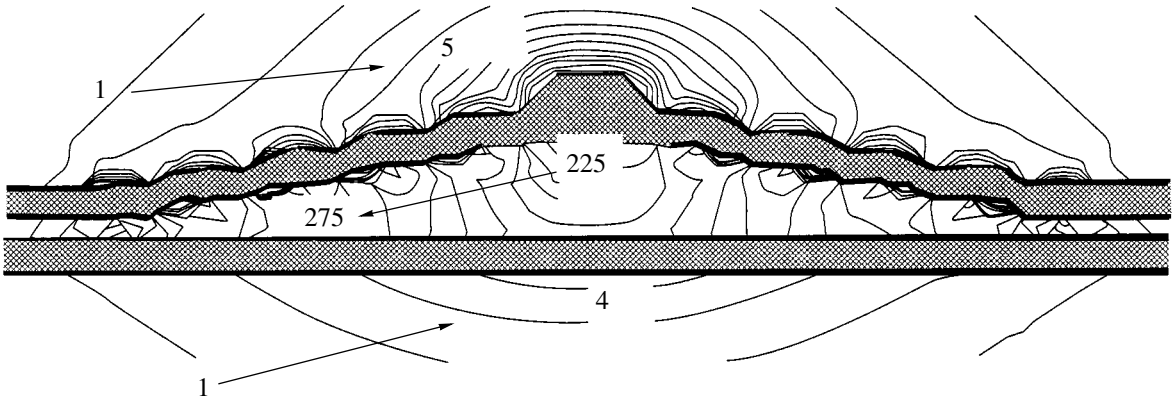


Fig. 1. Distribution of elastic energy in a QD and its neighborhood in the (100) plane passing through the pyramid axis. The figures indicate the energies per atom in units of 10^{-4} eV. Arrows show the directions of increasing energy [22].

4. ELECTRONIC STRUCTURE OF A SPATIALLY INDIRECT EXCITON

For homogeneous bulk semiconductors, the term "indirect exciton" is applied to an excited electron state formed upon indirect optical transitions [25]. In type-II QDs, there exist optical transitions that are indirect in real space. The same transitions may either be indirect in the *k*-space, where *k* are the wave vector components (e.g., in Ge/Si and GaAs/AlAs, the latter system being type II for a QD size less than 56 Å and type I for QDs larger 56 Å [8]) or direct (e.g., in the InAs/GaSb system with InAs QDs, belonging to type II for QD size less than 87 Å [8], and in GaSb/GaAs [6]). The excited states formed upon optical transitions, with the electron and hole localized on different sides of the heterointerface, are called spatially indirect excitons.

To obtain the absorption spectrum of a nanocluster containing N electron–hole pairs, it is necessary to solve the Schrödinger equation with a Hamiltonian that includes the kinetic and potential energy of a noninteracting electron and hole and the energy of their interaction [26].

Wave functions and the energy spectrum of electrons and holes in spatially indirect excitons were simulated numerically for a Ge pyramid in Si, with dimensions specified in Section 3. The band offsets at the Ge/Si heterojunction were calculated on the basis of the obtained spatial distribution of elastic strains both inside and outside the pyramidal QD as well as on the known strain potentials for Si and Ge [27]. The elastic strains in Si lift the sixfold degeneracy of the Δ valleys, with four and twofold-degenerate valleys being formed. The twofold degenerate valleys are oriented along the [001] and $[00\bar{1}]$ directions and lie lower in energy than states in the conduction band of Ge (the minimum of the conduction band). The maximum of the valence band is formed by heavy holes from Ge nanoclusters.

The states of spatially indirect excitons and excitonic complexes are described in terms of the effective mass method. A system of 3D Schrödinger equations was solved: two equations for a single exciton, three equations for an exciton–hole complex, and four equations for two excitons in a QD.

The interaction between charged particles was simulated by a static screened Coulomb potential. The conduction band offset between the corresponding Δ minima of unstrained Ge and Si was 340 meV, and the valence band offset was 610 meV. The electron effective masses in the Si conduction band were $m_{xy} = 0.19m_0$ in the QD growth plane, and $m_z = 0.92m_0$; the hole effective masses in the valence band were $m_{xy} = 0.39m_0$ and $m_z = 0.2m_0$. Only heavy holes were considered, because the light-hole states were close to the valence band edge of Si. The exciton wavefunctions were taken in the Hartree approximation as a product of the electron and hole wavefunctions.

For the case of a single exciton, the calculations demonstrated that the electron is localized in the region of the maximal strain in Si (in the vicinity of the Ge pyramid apex), and the hole is confined near the pyramid base. In a two-exciton complex, the repulsion of electrons causes their spatial separation; therefore, the second electron is bound near the interface between Si and the continuous Ge layer, on which the pyramids lie (Fig. 2). In this region, there exists a second local minimum of the conduction band induced by elastic strains originating from the lattice mismatch between Ge and Si. In the vicinity of the pyramid apex, this local extremum is shallower than the first one.

The calculations also demonstrated that the electron binding energy in exciton formation is $E_t = 38$ meV. The main contribution to the electron localization energy near the Ge/Si heterointerface comes from the electron–hole Coulomb interaction (the exciton binding energy is 29 meV). The remaining part (9 meV) is

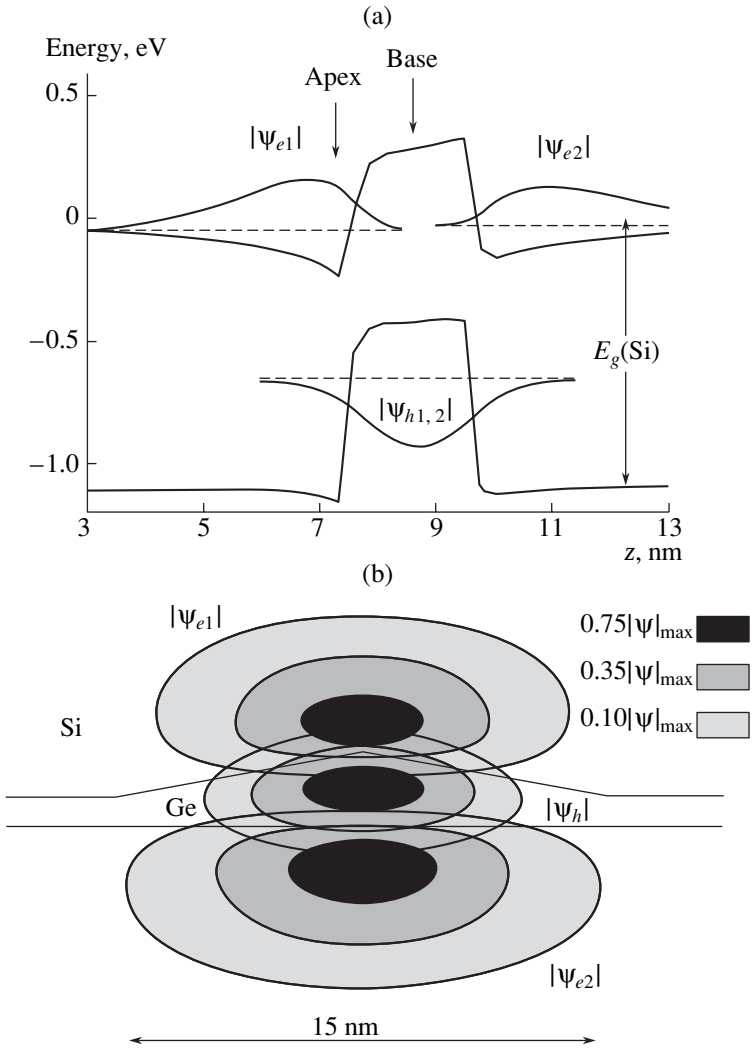


Fig. 2. (a) Calculated potential profile (along the *z*-axis passing through the apex of a Ge pyramid) in which electrons and holes move, constituting an excitonic complex. (b) 2D image of the modulus of electron and hole wavefunctions ($|\psi_e|$ and $|\psi_h|$) in the cross section of a QD and its neighborhood. The extents of black denote regions at whose boundaries the wavefunctions decay to 75, 35, and 10% of the maximum value $|\psi|_{max}$ [26, 28].

due to the contribution from the nonuniform distribution of strains to the potential well formation at the Ge/Si interface. In fact, the latter value is the energy of the electron state in the potential well of a neutral Ge nanocluster.

5. EXCITONIC ABSORPTION

The absorption peak related to the electron transition from the Ge QD valence band to the Si conduction band with the formation of an exciton is observed at 760–770 meV [26, 28]. This transition yields the ground state of a spatially indirect exciton (a hole is formed in the *H*0 ground state in a Ge nanocluster, and an electron passes into the *E*0 ground state in Si at the heterojunction). A weaker absorption peak in the range 850–860 meV is attributed to an excited exciton, with an electron and a hole in both the *H*1 and *E*1 excited states. The band width of 50–70 meV is presumably due to fluctuations of the shape and size of Ge clusters.

5.1. Single Exciton

The oscillator strength *f* of the excitonic transition in Ge/Si was determined from the optical absorption data. The obtained value of 0.5 is about 20 times less than the oscillator strength for direct (in real and *k* space) excitons in InAs/GaAs QD structures having $f = 10.9$ [29, 30]. The oscillator strength of an excitonic transition is proportional to the squared overlap integral of the electron and hole wavefunctions. In InAs/GaAs QD structures, the electron and hole are localized within the same nanocluster. Therefore, the overlap of their wavefunctions is relatively large, about 80% [31]. The solu-

tion of a self-consistent problem for the Ge/Si system yields a 15% overlap of the electron and hole wavefunctions [28]. The smaller overlap in Ge/Si (compared with that in InAs/GaAs) is a direct consequence of the spatial separation of the electron and hole in type-II heterostructures. At the same time, the overlap in type-II QD structures depends on the height of the barrier separating the electron and the hole as well as on the QD size. Evidently, an infinite barrier corresponds to complete separation of wavefunctions and zero overlap. This situation virtually does not occur in the discussed structures with finite barrier height.

Of more interest is the dependence of the overlap factor on the QD size at a constant (finite) barrier. The wavefunction is "squeezed out" of the QD volume with a decreasing QD size or barrier height. This enhances the overlap of the electron and hole wavefunctions and makes the correlation of their motion stronger. According to calculations made in [9], the overlap factor starts to grow dramatically when the ratio of the QD size to the Bohr radius of the hole becomes less than four. The overlap factor decreases with an increasing QD size (in Bohr radius units). The conclusion that the oscillator strength may become appreciable in type-II QDs was also made on analyzing the kinetics of the photoluminescence decay in an array of pyramidal GaSb/GaAs QDs [7].

The measured absorption probability ($\alpha = 1.6 \times 10^{-4}$) allows the evaluation of the effective interband absorption cross section in Ge QDs to be $\sigma_{ph} = \alpha/2\sigma_{qd} = 2.5 \times 10^{-16}$ cm^2, where σ_{qd} is the density of QDs. This value exceeds the typical photoionization cross sections of deep levels in Si (~10^{-17} cm^2 [32]) by more than an order of magnitude. This fact is an indication that the absorption band at 750–850 meV is not associated with defects or impurities in Si. Another argument in favor of the developed concept is furnished by experiments on annealing Ge/Si QD structures. The absorption band at 750–850 meV is not affected by a 30-min annealing at 500°C, whereas many point defects in Si are annealed at this temperature, being transformed into more intricate complexes.

5.2. Multiparticle Excitonic Complexes

The existing methods for studying QD structures make it possible to create conditions under which several charged excitons are formed in a single QD. For example, holes are accumulated at the top of the valence band in Ge nanoclusters when holes are injected into the Ge/Si QD structures [28]. An 11-meV blue shift of the line of main excitonic absorption (*H*0–*E*0) was observed in single-charged Ge QDs. Figure 3 shows the shift of the excitonic absorption peak in relation to the average number of holes per QD. The transition energy increases sharply when an extra hole appears in the QD ground state, thus enabling the formation of a charged (⟨2 holes⟩–electron) excitonic complex on photon absorption. The observed blue shift is accounted for by the dependence of the excitonic transition energy on the QD charge state: the generation of an exciton in a hole-containing a QD requires more energy than that in a neutral QD. The positive sign of this effect is a specific feature of type-II QDs, which means that the energy of the hole–hole Coulomb interaction in a QD—E_{hh} (repulsion)—exceeds the hole–electron interaction energy E_{eh} (exciton binding energy).

This result contradicts the previous data for spatially direct excitons in InAs/GaAs QD arrays, when the energy of the excitonic transition decreases upon the formation of excitonic complexes [29, 30]. In the case of direct excitons, the electron–hole interaction dominates and the absorption line of a charged exciton is redshifted [29]. For type-II QDs, it would be natural to expect that $E_{hh} > E_{eh}$, owing to the spatial separation of an electron and a hole. This will cause a shift of the excitonic line to shorter wavelengths upon the formation of a charged complex.

The additional energy of the charged exciton as compared with a neutral one is determined by the difference $E_{ex\text{-}h} = E_{hh} - E_{eh}$, equal to the experimentally found 11-meV shift of the optical transition. Taking into account that $E_{hh} = 36$ meV [33], we obtain the electron–hole interaction energy $E_{eh} = 25$ meV, which correlates with the solution of the self-consistent problem. The obtained E_{eh} exceeds the free exciton binding energy for Ge by nearly an order of magnitude and is twice that for Si.

The optical absorption peak of the excited exciton state is shifted to a shorter wavelength to a lesser extent upon the appearance of a hole in the QD ground state (Fig. 3). This is due to a weaker electron–hole interaction because of the smaller overlap of their wavefunctions in states with large localization radii.

An alternative way to produce excitonic complexes in a single QD consists in the additional illumination of a structure with interband light when measuring the absorption spectra [28]. Additional illumination leads to the filling of hole levels in a Ge QD and the electron levels in Si in the vicinity of the heterojunction with nonequilibrium carriers. Thus, optical absorption of radiation causing *H*0–*E*0 excitonic transition in a structure containing a single exciton gives rise to a biexciton, whereas the injection of holes in optical absorption yields an exciton–hole complex.

A stronger blue shift of the excitonic absorption line was observed under additional illumination as compared with that in the case of injection of holes into a QD. If we assume, during the formation of an exciton–exciton complex, that a pair of holes is in the ground state of the Ge QD and a pair of electrons is in one QW at the phase boundary, then the energy of exciton–exciton interaction increases by

$$\Delta E_{ex\text{-}ex} = E_{hh} + E_{ee} - 2E_{eh},$$

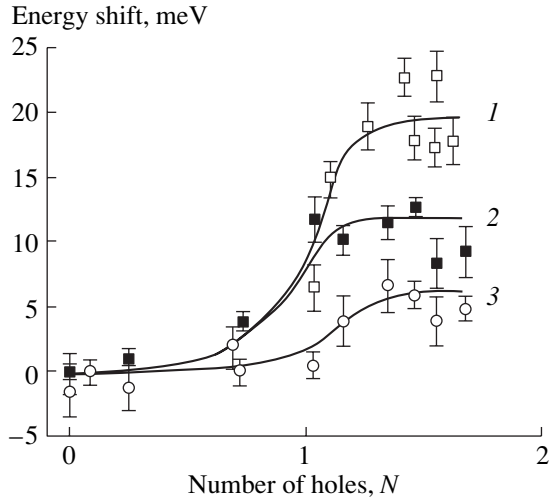


Fig. 3. Shift of excitonic absorption peak vs. the average number of holes per QD in a Ge/Si structure (*1*) under illumination and with injection of holes for (*2*) the ground state and (*3*) excited state of exciton.

where E_{ee} is the energy of electron–electron interaction in the QD. Using the experimental values $\Delta E_{\text{ex-ex}}$ = 20 meV (the blue shift of the excitonic absorption line under additional illumination), E_{hh} = 36 meV, and E_{eh} = 25 meV, we obtain E_{ee} = 34 meV, i.e., $E_{ee} \approx E_{hh}$.

However, this result contradicts rather obvious concepts. The localization radius of a hole (localized within a Ge nanocluster) must be considerably smaller than the radius of the electron localization at a QD. Since the interaction energy is inversely proportional to the characteristic radius of the wavefunction, the relation $E_{ee} < E_{hh}$ must always be valid. The estimated value E_{ee} = 34 meV indicates that the above assumption where two electrons are present in the same QW is wrong. If the second electron is in another potential well at the heterointerface, additional single-particle energies E_1, E_2 for each electron in its own potential well must appear in the relation for $\Delta E_{\text{ex-ex}}$, and the corresponding potential wells are not equivalent (see Sec. 4). In this case, the expression for $\Delta E_{\text{ex-ex}}$ must include one more positive term $(E_2 - E_1)$ and the E_{ee} value obtained using this relation may be substantially smaller than E_{hh}. As shown by calculations for the electronic configuration of the exciton (Sec. 4), the first of the two electrons is localized in Si at the apex of the Ge pyramid; the second, under the pyramid base.

The calculated interaction energy of two electrons E_{ee} = 19 meV. Then, using the known interaction energies E_{hh} = 36 meV and E_{eh} = 25 meV, we obtain $\Delta E_{\text{ex-ex}}$ = 5 meV. This means that, if the electrons had equal single-particle energies, the excitonic absorption line would shift by only 5 meV upon the formation of an exciton–exciton complex. The experimental value of this shift is 20 meV. Therefore, the "blue" shift observed upon the formation of two excitons in a single QD is related to the higher quantum-confinement energy of the second electron as compared to that of the first, with the difference in energies arising from the localization of two electrons in nonequivalent potential wells.

6. NEGATIVE PHOTOCONDUCTIVITY

Along with the spatial separation of the exciton ground state in Ge/Si QDs, solving the self-consistent problem to determine the electron and hole energy spectrum yielded results that allowed us to predict and experimentally find the effect of negative interband photoconductivity [34, 35] consisting in that the conductivity of the QD layer decreased under additional illumination. For most data on the photoconductivity in semiconductors, the conductivity of a material increases under interband illumination. It was found that a QD containing N holes can trap $N + 1$ electrons, with the energy of the extra electron dependent on the number of holes. Figure 4 shows the calculated binding energy of the $(N + 1)$-th electron in the vicinity of a Ge island containing N holes in relation to the number of holes in a QD. Since interband illumination generates electrons and holes in pairs (i.e., at equal concentrations), such an additional (and, consequently, unfilled) state would be a trap for the equilibrium conduction-band electrons.

Let us consider *n*-Si with undoped Ge nanoclusters introduced into it (Fig. 5). The dark conductivity of the system is determined by free electrons excited into the Si conduction band via the thermal ionization of donors. As mentioned above, even in the absence of holes in the QDs, a shallow electron level with the binding energy $E_t \approx 9$ meV exists at the Si/Ge heterointerface owing to nonuniform strains that create a potential well for electrons (Fig. 4). Therefore, the equilibrium concentration of electrons in the conduction band is lowered because of the electron capture to this level. In the absorption of light, which induces interband transitions to form electron–hole pairs, holes are accumulated in Ge QDs, charging them positively. Consequently, the potential wells for electrons appear in Si at the Ge/Si heterointerfaces, with photoelectrons accumulated in these wells (Fig. 5b). With an increasing number of holes in the islands (on raising the illumination intensity), the energy depth of the "excess" electron level increases (at $N < 3$) (Fig. 4). Since the electron filling of the level increases as its energy becomes lower, the electron concentration in the conduction band must decrease with the conductivity of the system.

Negative photoconductivity in semiconductors was first observed by A.F. Ioffe and A.V. Ioffe. They demonstrated that, under conditions of strong surface absorption of light, such conductivity is possible when the majority carriers recombine in the surface layer and the inward diffusion of minority carriers enhances the recombination in the bulk, i.e., suppresses the bulk con-

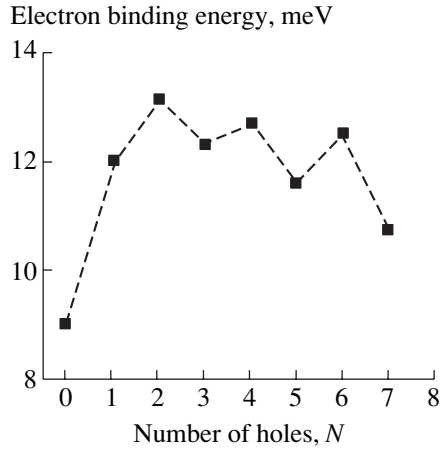


Fig. 4. Binding energy of the (N + 1)-th (excess) electron at the Ge/Si interface vs. the number of holes in Ge nanocluster, N [28].

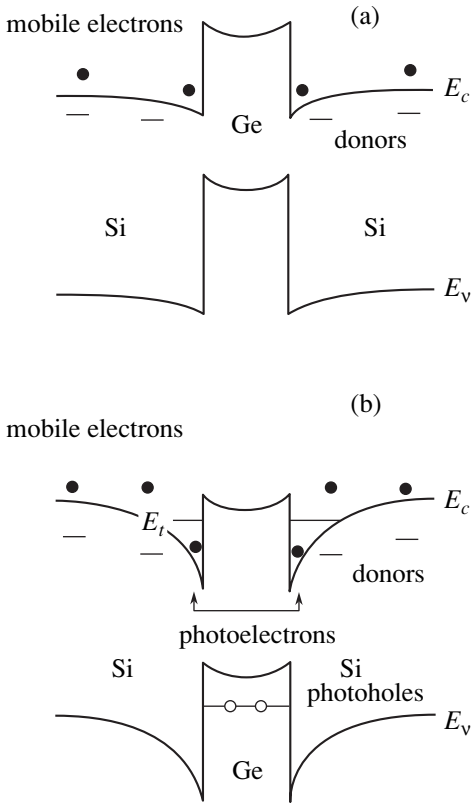


Fig. 5. Band diagram of Ge/*n*-Si heterostructure with Ge QDs (a) in the dark and (b) under illumination. Solid circles, electrons; open circles, holes.

ductivity [36]. Since, in our case, the recombination time in the surface layer with QDs is longer than that in the bulk [35], this mechanism is inoperative.

Another known mechanism of the negative photoconductivity observed earlier in bulk *p*-Ge samples is the modulation of the mobility of hot holes upon carrier excitation of carriers from a heavy- to a light-hole band [37, 38]. However, this situation only occurs under intraband illumination (photon energy ≈120 meV).

Negative photoconductivity associated with the recharging of scattering centers by light was also observed in layered GaSe and InSe crystals [39].

It should also be noted that a decrease in the carrier concentration in 2D subbands and an increase in the structure resistance under illumination with $h\nu > E_g$ photons has been observed in 2D systems based on $In_{0.25}Ga_{0.75}Sb/InAs$ [40], $InAs/Al_{0.5}Ga_{0.5}Sb$ [41], and $ZnS_{1-x}Se_x/Zn_{1-y}Cd_ySe$ heterostructures [42]. The effect was ascribed to the capture of photoexcited electrons from the quantum valley to electronic traps within the barrier. The presence of electronic traps accounted for the negative photoconductivity in PbTe [43] and $Pb_{1-x}Sn_xTe\langle In\rangle$ films [44].

The nature of a trapping center is specific to each particular material. For example, the traps in GaAs structures are, presumably, *DX*-centers [45]. We were the first to demonstrate [34] that it is possible for photoinduced traps in type-II QD arrays of one type of carrier to be unrelated to structural defects but to arise from fundamental quantum effects.

To verify the proposed concepts, we studied the photoconductivity of epitaxial Ge/Si structures grown on (001) Si substrate with phosphorus concentration $N_s \approx 10^{15}$ cm^{-3}. The substrate thickness was L_s = 300 μm. The structures comprised ten layers of Ge islands interspersed with 30-nm-thick Si layers. The overall thickness of the epitaxial layer was $L_{epi} \approx$ 0.3 μm. The donor (Sb) concentration in the epitaxial Si was $N_{epi} \approx 2.5 \times 10^{16}$ and 8×10^{16} cm^{-3}. The sheet nanocluster density was $\sigma_{qd} = 3 \times 10^{11}$ cm^{-2}.

A GaAs LED with peak emission at ~0.9 μm was used as a light source. The LED emission intensity was modulated at a frequency of 2 kHz. The lateral photoconductivity was recorded at the modulation frequency in the linear (Ohmic) region of the current–voltage characteristic (at electric field strengths no more than 0.5 V/cm). Electric contacts were fabricated by Al sputtering, with subsequent heating of the structure to 450°C in an N_2 atmosphere. To preclude the illumination of contacts, which could induce spurious negative photoconductivity [46], the contacts and the adjacent region were screened with an opaque coating.

Figure 6 presents the relative photoconductivity $\Delta G/G$ as a function of the illumination intensity (P) for Ge/*n*-Si QD structures at different temperatures (solid and dotted lines) and for a sample having no epilayer with QD-islands at 77 K (dashed line). In this sample, the photoconductivity is positive and increases practically linearly with P. A negative photoconductivity is observed in the structures with Ge nanoclusters at illumination intensities $P < 100$ mW/cm^2.

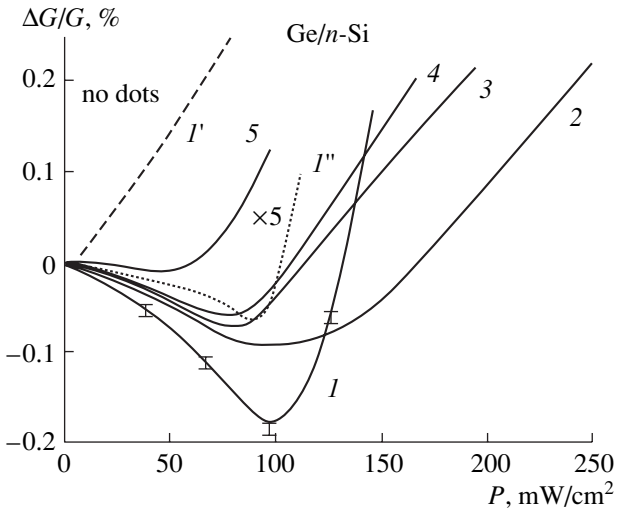


Fig. 6. Relative photoconductivity vs. the interband illumination intensity in *n*-Si with QDs, at temperatures *T*: (*1*, *1*', *1*") 77, (*2*) 87, (*3*) 99, (*4*) 102, and (*5*) 180 K. Antimony concentration: (*1*–*5*) 2.5×10^{16} and (*1*") 8×10^{16} cm^{-3}. (*1*') Structure without QDs.

The negative photoconductivity is not observed in *p*-layers with Ge QDs. Figure 7 shows the photoconductivity of a Ge/*p*-Si structure as a function of the illumination intensity. The conditions of sample growth were similar to those for the Ge/*n*-Si structures, with the exception of the conduction type of the substrate and the top epitaxial Si layer with Ge islands (boron acceptor concentration ~10^{15} cm^{-3}). As can be seen, there is no region of negative photoconductivity in these curves. Instead, an extended region with a small positive photoconductivity is observed, after which there occurs a sharp rise in the photocurrent. Such behavior at small illumination intensities is associated with the capture of both types of nonequilibrium carriers to bound states in the vicinity of Ge islands (holes tied to states inside the islands; electrons tied to states at the Ge/Si heterointerface). The rise in photoconductivity at high illumination intensities is due to the filling of electronic levels and the generation of free nonequilibrium electrons. This result also indicates that the observed negative photoconductivity is not related to the carrier mobility modulation under illumination.

The described mechanism of negative photoconductivity is only operative in type-II QDs, since both electrons and holes are localized in one semiconductor in type-I heterostructures irrespective of the QD charge state.

7. COMPARISON OF PARAMETERS OF PHOTOSENSITIVE QD STRUCTURES

It is of interest to compare the available data on the photon absorption cross sections and detectivity for structures with both types of QDs. For the Ge/Si system, the wavelength range 6–20 μm corresponds to interlevel transitions [47, 48], and the range 1.7–3 μm, to indirect (in real and *k*-space) interband transitions. As can be seen, the larger cross section corresponds to transitions between states related to the vertical confinement of particle motion (in the *z*-direction). The exceptions are the InAs/InAlAs and Ge/Si systems, in which large cross sections of photon absorption have also been observed in laterally polarized transitions (light beam polarization in the *xy*-plane of the QD sheets). This property allows us to regard these systems

Parameters of photosensitive structures with QDs and superlattices

QD systems (QD/matrix)	*T*, K	λ_m, μm	*n*	σ_{ph}, 10^{-15} cm^2	*D**, 10^8 cm Hz$^{1/2}$/W	Source
InAs/GaAs	300	6	*z*	3		[51]
InAs/GaAs	90	7	*z*	3		[52]
InAs/GaAs	120	11	*xy*	0.16		[53]
InAs/GaAs	120	6–8	*xy*	0.25		[53]
InAs/InAlAs	300	13.8	*xy*	15		[54]
InAs/GaAs	300	10.6	*xy*		0.3	[55]
InGaAs/InGaP	77	5.5	*xy*		0.47	[56]
Ge/Si	77	6	*xy*	200		[50]
Ge/Si	300	10–20	*xy*	0.8	0.7–1.7	[57]
Ge/Si	300	1.7–3	*xy*	1		[57]
2D systems						
$Si_{1-x}Ge_x$/Si	77	9	*z*		10	[58]
InAs/GaSb	77	10.3	*z*		13	[59]

Note: *T* is the temperature of the experiment, λ_m is the wavelength in the spectral sensitivity peak, *n* is the radiation polarization, σ_{ph} is the cross section of photonic absorption, and *D** is the detectivity.

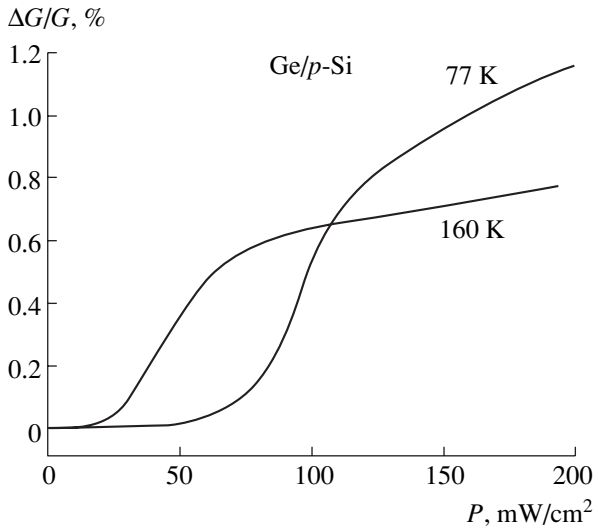


Fig. 7. Relative photoconductivity vs. the interband illumination intensity for a Ge/*p*-Si structure with Ge QDs.

as good candidates for the fabrication of photodetectors ensuring the absorption of radiation under normal incidence of light onto a structure [49, 50] (see table). A similar conclusion for the Ge/Si heterostructures follows from a comparative analysis of the detectivities D^* of photodetectors with different kinds of QDs, presented in the table. It is seen that the room-temperature D^* for the Ge/Si system is more than twice as large as the value obtained for InAs/GaAs QDs (type-I heterosystem).

A comparison of the detectivities of photodetectors based on structures with 2D electron gas and QD structures (see table) shows that the D^* for 2D systems is about an order of magnitude higher than the values that have been achieved in QD structures. We believe that this is due to the higher carrier concentration in 2D superlattices ($\sim 10^{12}$ cm^{-2}). The photosensitive structures in this study contain no more than ten QD layers. Simple evaluations show that raising the number of Ge QD layers must bridge the detectivity gap.

8. CONCLUSION

The performed investigations of type-II QD systems were aimed at determining the electronic spectra of excitations and their interactions and finding the physical parameters characterizing the class of zero-dimensional structures under study. The fundamental physical phenomena specific to type-II QDs are as follows:

(1) The exciton binding energy is higher than the free exciton binding energy in homogeneous bulk semiconductors, which results from the confinement of an electron and a hole, and a higher contribution of the overlap of their wavefunctions, resulting from their mutual penetration into a potential barrier of finite height.

(2) A blue shift of the excitonic transition occurs in the formation of exciton–hole and exciton–exciton complexes; spatial separation of electrons occurs in a biexciton in the Ge/Si structure.

(3) Negative photoconductivity under interband illumination occurs due to the capture of equilibrium carriers to localized states formed by the field of charged QDs.

Further, there exist a number of phenomena that have been discovered in type-II QDs but which may also be characteristic of type-I QDs:

(1) the larger cross section of photon absorption as compared with that associated with point defects in bulk semiconductors;

(2) the blue shift of the interlevel absorption line in QDs at a higher carrier concentration (lateral depolarization effect) [26, 47, 48, 57];

(3) the larger cross section of carrier trapping into QDs [60];

(4) the oscillatory behavior of conductivity in an array of tunnel-coupled QDs when varying the extent their filling [61, 62];

(5) the universal character of the preexponential factor in the temperature dependence of conductivity along the tunnel-coupled QD sheets equal to e^2/h [63].

We should also mention the most important (as regards application in semiconductor devices) results the studies performed:

(1) the "single-electron" FET with a channel of tunnel-coupled QDs [61, 62];

(2) absorption of light in its normal incidence onto a QD structure [49];

(3) IR photodetectors operating on optical transitions between the quantum confinement levels and on interband transitions; control by means of external bias over the spectral range of light absorption [49].

Important technological features associated with the development of this line of research consist of self-organization in low-temperature epitaxy, the investigations of controlling the nanocluster formation by means of irradiation with low-energy ions during molecular-beam heteroepitaxy [64], and the application of silicon-on-insulator structures.

ACKNOWLEDGMENTS

The elastic strain fields were calculated, and the self-consistent problem of the energy spectrum was solved by A.V. Nenashev.

This study was supported by the Russian–Ukrainian Scientific and Technological Program "Nanophysics and Nanophotonics" (project no. 2000-F2), Program "Physics of Solid-State Structures" (project no. 98-1100), Program "Universities of Russia—Basic Research" (project no. 015.01.01.34), and the Russian Foundation for Basic Research (project no. 00-02-17885).

REFERENCES

1. Zh. I. Alferov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 3 (1998) [Semiconductors **32**, 1 (1998)].
2. S. V. Gaponenko, Fiz. Tekh. Poluprovodn. (St. Petersburg) **30**, 577 (1996) [Semiconductors **30**, 315 (1996)].
3. N. N. Ledentsov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33**, 1039 (1999) [Semiconductors **33**, 946 (1999)].
4. A. I. Yakimov, V. A. Markov, A. V. Dvurechenskii, and O. P. Pchelyakov, Philos. Mag. B **65**, 701 (1992).
5. N. N. Ledentsov, V. M. Ustinov, V. A. Shchukin, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 385 (1998) [Semiconductors **32**, 343 (1998)].
6. F. Hatami, N. N. Ledentsov, M. Grundmann, *et al.*, Appl. Phys. Lett. **67**, 656 (1995).
7. F. Hatami, M. Grundmann, N. N. Ledentsov, *et al.*, Phys. Rev. B **57**, 4635 (1998).
8. J. M. Rorison, Phys. Rev. B **48**, 4643 (1993).
9. U. E. H. Laheld, F. B. Pedersen, and P. C. Hemmer, Phys. Rev. B **52**, 2697 (1995).
10. S. Fukatsu, H. Sunamura, Y. Shiraki, and S. Komiyama, Appl. Phys. Lett. **71**, 258 (1997).
11. A. V. Kalameĭtsev, A. O. Govorov, and V. M. Kovalev, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 634 (1998) [JETP Lett. **68**, 669 (1998)].
12. A. V. Dvurechenskii, A. I. Yakimov, V. A. Markov, *et al.*, Izv. Akad. Nauk, Ser. Fiz. **63**, 307 (1999).
13. A. V. Dvurechenskii and A. I. Yakimov, Izv. Akad. Nauk, Ser. Fiz. **64**, 306 (2000).
14. O. P. Pchelyakov, Yu. B. Bolkhovityanov, A. V. Dvurechenskii, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 1281 (2000) [Semiconductors **34**, 1229 (2000)].
15. A. V. Dvurechenskii and A. I. Yakimov, Izv. Vyssh. Uchebn. Zaved., Mater. Élektron. Tekh., No. 4, 4 (1999).
16. A. I. Yakimov, V. A. Markov, A. V. Dvurechenskii, and O. P. Pchelyakov, J. Phys.: Condens. Matter **6**, 2573 (1994).
17. A. V. Rzhanov, *Electron Processes on Semiconductor Surface* (Nauka, Moscow, 1971).
18. D. Bimberg, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33**, 1044 (1999) [Semiconductors **33**, 951 (1999)].
19. M. M. Rieger and P. Vogl, Phys. Rev. B **48**, 14276 (1993).
20. T. Meyer, M. Klemenc, and H. von Känel, Phys. Rev. B **60**, R8493 (1999).
21. O. G. Schidt, K. Eberl, and Y. Rau, Phys. Rev. B **62**, 16715 (2000).
22. A. V. Nenashev and A. V. Dvurechenskii, Zh. Éksp. Teor. Fiz. **118**, 570 (2000) [JETP **91**, 497 (2000)].
23. P. N. Keating, Phys. Rev. **145**, 637 (1966).
24. Y. Kikuchi, H. Sugii, and K. Shintani, J. Appl. Phys. **89**, 1191 (2001).
25. C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963; Nauka, Moscow, 1967).
26. A. I. Yakimov, A. V. Dvurechenskii, N. P. Stepina, *et al.*, Zh. Éksp. Teor. Fiz. **119** (3), 574 (2001) [JETP **92**, 500 (2001)].
27. C. G. van de Walle, Phys. Rev. B **39**, 1871 (1989).
28. A. I. Yakimov, N. P. Stepina, A. V. Dvurechenskii, *et al.*, Semicond. Sci. Technol. **15**, 1125 (2000).
29. R. J. Warbuton, C. S. Dürr, K. Karrai, *et al.*, Phys. Rev. Lett. **79**, 5282 (1997).
30. K. H. Schmidt, G. Medeiros-Ribeiro, and P. M. Petroff, Phys. Rev. B **58**, 3597 (1998).
31. M. Grundmann, O. Stier, and D. Bimberg, Phys. Rev. B **52** (16), 11969 (1995).
32. F. V. Gasparyan, Z. N. Adamyan, and V. M. Arutyunyan, *Silicon Photodetectors* (Erevan. Univ., Yerevan, 1989).
33. A. I. Yakimov, A. V. Dvurechenskii, A. I. Nikiforov, and O. P. Pchelyakov, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 125 (1998) [JETP Lett. **68**, 135 (1998)].
34. A. I. Yakimov, A. V. Dvurechenskii, A. I. Nikiforov, and O. P. Pchelyakov, Pis'ma Zh. Éksp. Teor. Fiz. **72**, 267 (2000) [JETP Lett. **72**, 186 (2000)].
35. A. I. Yakimov, A. V. Dvurechenskii, A. I. Nikiforov, *et al.*, Phys. Rev. B **62**, R16283 (2000).
36. A. F. Ioffe, *Physics of Semiconductors* (Akad. Nauk SSSR, Moscow, 1957; Infosearch, London, 1960).
37. A. M. Danishevskiĭ, A. A. Kastal'skiĭ, B. S. Ryvkin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **10**, 470 (1969) [JETP Lett. **10**, 302 (1969)].
38. A. F. Gibson and P. N. D. Maggs, J. Phys. D **7**, 292 (1974).
39. S. S. Ishchenko, S. M. Okulov, A. A. Klimov, and Z. D. Kovalyuk, Fiz. Tekh. Poluprovodn. (Leningrad) **17**, 1230 (1983) [Sov. Phys. Semicond. **17**, 776 (1983)].
40. I. Lo, W. C. Mitchel, R. Kaspi, *et al.*, Appl. Phys. Lett. **65**, 1024 (1994).
41. J.-P. Cheng, I. Lo, and W. C. Mitchel, J. Appl. Phys. **76**, 667 (1994).
42. I. Lo, S. J. Chen, Y. C. Lee, *et al.*, Phys. Rev. B **57**, R6819 (1998).
43. S. A. Kaz'min, V. I. Kaĭdanov, and S. S. Shevchenko, Fiz. Tekh. Poluprovodn. (Leningrad) **19**, 530 (1985) [Sov. Phys. Semicond. **19**, 328 (1985)].
44. I. I. Zasavitskiĭ, B. N. Matsonashvili, and V. T. Trofimov, Fiz. Tekh. Poluprovodn. (Leningrad) **23**, 2019 (1989) [Sov. Phys. Semicond. **23**, 1249 (1989)].
45. A. G. De Oliveira, G. M. Ribeiro, D. A. W. Soares, and H. Chacham, Appl. Phys. Lett. **64**, 2258 (1994).
46. V. G. Kustov, Fiz. Tekh. Poluprovodn. (Leningrad) **10**, 2215 (1976) [Sov. Phys. Semicond. **10**, 1318 (1976)].
47. A. I. Yakimov, A. I. Dvurechenskii, N. P. Stepina, and A. I. Nikiforov, Thin Solid Films **380**, 82 (2000).
48. A. V. Dvurechenskii and A. I. Yakimov, Izv. Akad. Nauk, Ser. Fiz. **65**, 187 (2001).
49. A. I. Yakimov, A. V. Dvurechenskii, Yu. Yu. Proskuryakov, *et al.*, Appl. Phys. Lett. **75**, 1413 (1999).
50. P. Boucaud, V. Le. Thanh, S. Sauvage, *et al.*, Appl. Phys. Lett. **74**, 401 (1999).
51. S. Sauvage, P. Boucaud, F. H. Julien, *et al.*, Appl. Phys. Lett. **71**, 2785 (1997).
52. S. Sauvage, P. Boucaud, F. H. Julien, *et al.*, J. Appl. Phys. **84**, 4356 (1998).
53. S. Sauvage, P. Boucaud, J.-M. Gerard, and V. Thierry-Mieg, Phys. Rev. B **58**, 10562 (1998).
54. A. Weber, O. Gauthier-Lafaye, F. H. Julien, *et al.*, Appl. Phys. Lett. **74**, 413 (1999).

55. T. Cho, J.-W. Kim, J.-E. Oh, and S. Hong, in *Technical Digest of the International Electron Devices Meeting, 1998*, p. 441.

56. S. Kim, H. Mohseni, M. Erdtmann, *et al.*, Appl. Phys. Lett. **73**, 963 (1998).

57. A. I. Yakimov, A. I. Dvurechenskii, A. I. Nikiforov, and Yu. Yu. Proskuryakov, J. Appl. Phys. **89** (10), 5676 (2001).

58. R. P. G. Karunasiri, J. S. Park, and K. L. Wang, Appl. Phys. Lett. **59**, 2588 (1991).

59. H. Mohseni, E. Michel, Jan Sandoen, *et al.*, Appl. Phys. Lett. **71**, 1403 (1997).

60. A. I. Yakimov, A. V. Dvurechenskii, A. I. Nikiforov, and O. P. Pchelyakov, Phys. Low-Dimens. Struct. **3/4**, 99 (1999).

61. A. I. Yakimov, C. J. Adkins, R. Boucher, *et al.*, Phys. Rev. B **59**, 12598 (1999).

62. A. I. Yakimov, A. V. Dvurechenskii, A. I. Nikiforov, and C. J. Adkins, Phys. Status Solidi B **218**, 105 (2000).

63. A. I. Yakimov, A. V. Dvurechenskii, V. V. Kirienko, *et al.*, Phys. Rev. B **61**, 10868 (2000).

64. A. V. Dvurechenskii, V. A. Zinov'ev, V. A. Kudryavtsev, and Zh. V. Smagina, Pis'ma Zh. Éksp. Teor. Fiz. **72** (3), 190 (2000) [JETP Lett. **72**, 131 (2000)].

Translated by D. Mashovets

Semiconductors, Vol. 35, No. 9, 2001, pp. 1106–1109. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1154–1158.

PHYSICS OF SEMICONDUCTOR DEVICES

Switching Characteristics of Electron-Irradiated MOS-Controlled Thyristors

E. V. Chernyavskiĭ*, V. P. Popov, Yu. S. Pakhmutov, Yu. I. Krasnikov, and L. N. Safronov

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: evgen@isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—MOS-controlled thyristors (MCTs) were designed and fabricated. The effect of electron irradiation on static and dynamic characteristics has been studied. Electron irradiation was found to substantially reduce the MCT turn-off time. An increase in the controlled current density was observed. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A new generation of bipolar-field-effect devices for power electronics with a blocking voltage of up to 4.5 kV is developing vigorously. These devices include the MOS-controlled thyristor (MCT), the emitter-switched thyristor (EST), the insulated-gate bipolar transistor (IGBT), and the injection-enhanced gate transistor (IEGT). A combination of the low forward drop with the advantages offered by field control make these devices the most promising in power electronics. In contrast to the gate turn-off (GTO) thyristor, the MCT and the IGBT need no cumbersome control circuits. The turn-off process is very complicated for power devices, the current densities being as high as 100 A/cm^2. Local inhomogeneities leading to current pinching and the resulting device breakdown may arise at these current densities. The turn-off time is another critical parameter, determining the power dissipated by the device as it is turned off. The main ways to reduce the turn-off time are anode shunting and controlling the carrier lifetime by introducing recombination centers. The latter can be done, e.g., by electron irradiation. This technology is being successfully used to reduce the switching losses in GTO thyristors. In bipolar-field devices, such irradiation deteriorates the parameters of control MOSFETs and may reduce the controlled current, and, consequently, make the safe operation area (SOA) of the device narrower. The goal of the present study was to investigate the possibility of controlling the turn-off time of an MCT with anode shunting and to analyze the static and dynamic characteristics of devices with different carrier lifetimes in the *N*-base.

2. EXPERIMENT

2.1. Anode Structure

To achieve a blocking voltage of 2.5 kV at a low forward drop across a MCT, we used an *n*-type anode buffer layer in combination with anode shunts, as in power GTO thyristors [1–3] (Fig. 1). Floating-zone silicon wafers 420 μm thick with a resistivity of 2000 Ω cm were used to fabricate the devices. The buffer *N* layer of the anode was produced by phosphorus diffusion to a depth of 15 μm. The anode P^+ emitter was formed by the implantation of boron through a photoresist mask with its further spreading to a 1-μm depth. Open windows were left in the anode P^+-emitter layer for depositing the anode metal contact onto the anode buffer layer. In this way anode shunts were formed. The ratio of the anode emitter contact area S_{p^+} to the area of *N*-buffer contact S_n was 100 : 1. The procedure of anode fabrication has been described elsewhere [4–6]. The anode metallization was multilayered. The first layer consisted of 1000-Å-thick Ti. Further, Ni and Ag layers of total thickness 1 μm were deposited. Such metallization is necessary for mounting the crystal into its package with tin–lead solder. The anode structure is shown in Fig. 1.

2.2. Cathode Structure

MCT is a field-controlled bipolar device. Therefore, the *N*-channel MOSFETs used to turn the thyristor on (on-FETs) and *P*-channel MOSFETs used to turn it off (off-FETs) are located on the planar surface of the MCT.

The cathode has the form of a matrix of 80 × 80-μm unit cells consisting of 24 *N* emitters placed in a *P* pocket. A well into the N^- base is fabricated at the center of a unit cell, serving as the drain of an *N*-channel on-FET. Cathode emitters are 10 × 10-μm squares of N^+ layer. Circular P^+ regions on the emitter periphery serve as sources of the *P*-channel off-FETs. A contact to the cathode metallization is located at the center of an emitter. Circular P^+ sources also have contacts to the

1063-7826/01/3509-1106$21.00

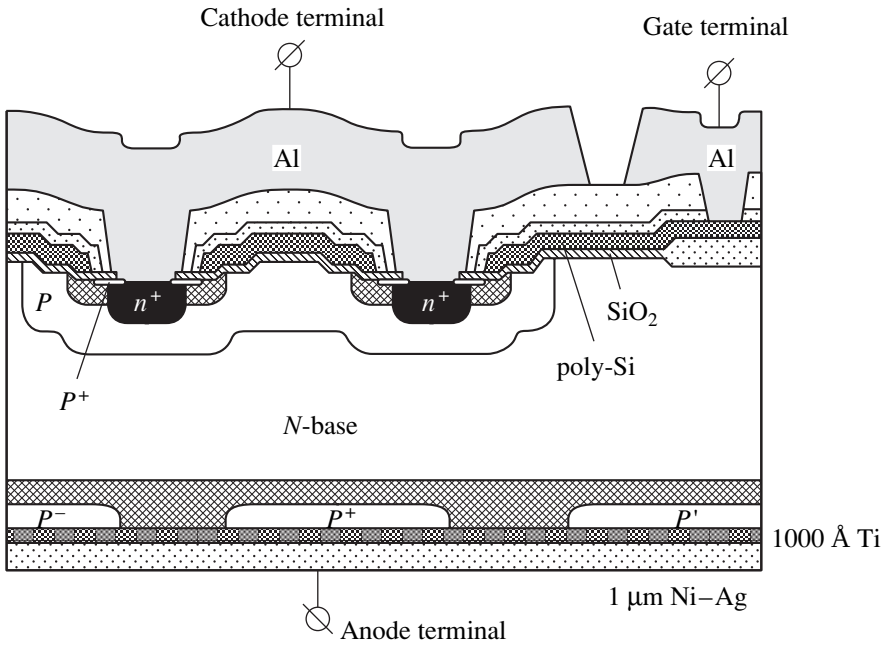


Fig. 1. MCT structure.

cathode metallization. The channels of *P*-channel MOSFETs are 0.8 μm long. The *P* pocket serves as a drain. The distance between the *N* emitters is 6 μm. The *N* emitters are fabricated by implanting phosphorus through a polysilicon mask, using the technology of self-alignment. The P^+ source is also implanted through a polysilicon mask. The polysilicon simultaneously serves as a mutual gate for N^- and *P*-channel MOSFETs. Polysilicon is doped with phosphorus to resistivity $R_s = 12\ \Omega/\square$. The *P* pocket is common to all the *N* emitters and has the form of a 7820 × 4820 μm rectangle. The total number of *N* emitters is 144042. The length of the emitter perimeter is 40 μm, and, therefore, the *W*/*L* ratio for the *P*-channel MOSFETs is 50. The aggregate width of the channels of all the *P*-channel MOSFETs is 5.76 m. The total number of the *N*-channel MOSFETs is 5901. The channel is 9 μm long, its aggregate width is $W = 0.354$ m. The contact to the polysilicon gate of the transistors is on a separate pad. Phosphosilicate glass is used as an interlayer insulation between the cathode aluminum and gate polysilicon. The MCT cathode structure is shown in Figs. 1 and 2.

2.3. Equivalent Circuit

To simplify the description of the processes occurring in MCT, we consider the equivalent circuit in Fig. 3. In this circuit, bipolar transistors *T*1 and *T*2 represent a vertical *NPNP* structure. Physically, this structure consists of the MCT layers: N^+ emitter–*P* pocket–*N* substrate–*P* anode. It is simulated by a tandem *NPN*–*PNP* transistor, with the base of the *T*1 transistor connected to the collector of *T*2. The *N*-channel MOSFET *M*2 serves as an on-FET. With a positive voltage applied to the gate, the transistor *M*2 injects carriers into the *N* substrate, which is the drain of the transistor *M*2, a positive current feedback arises between the *T*1 and *T*2 transistors, and the thyristor is turned on. To turn it off, this positive feedback should be broken. In a MCT, this is achieved by making the *P* channel conductive in the off-FET *M*1, whose drain is the *P* pocket. The hole current flows through this channel into the P^+ source connected to the N^+ emitter by metallization.

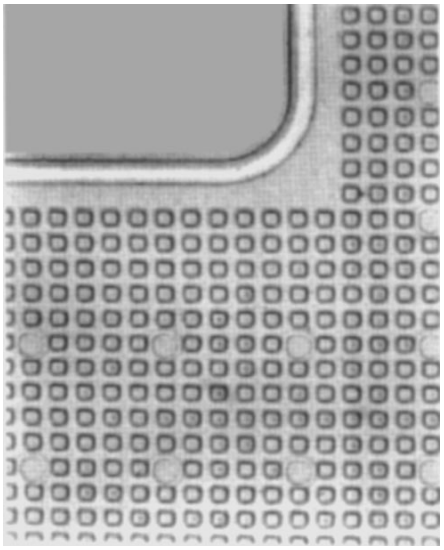


Fig. 2. MCT cathode structure.

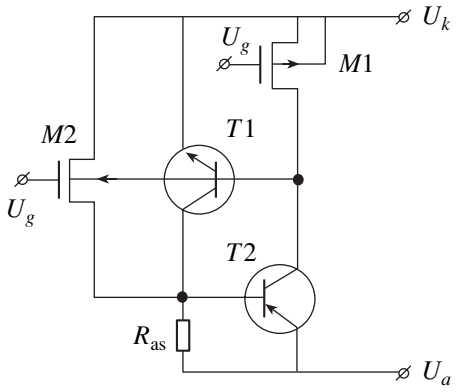


Fig. 3. Equivalent circuit of MCT.

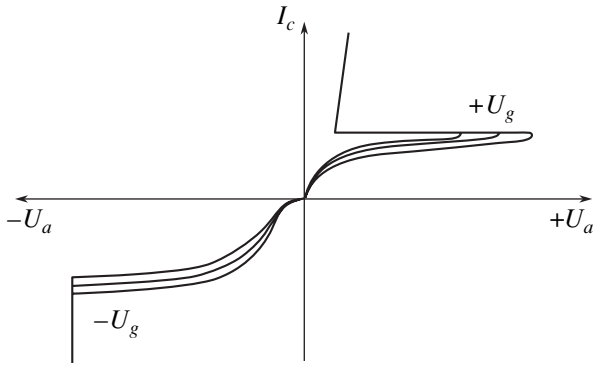


Fig. 4. Static *I–V* characteristics of an MCT with a shunted anode; gate bias U_g = 1–15 V is positive for the forward portion of the curve and negative U_g = –5–(–20) V for the reverse portion.

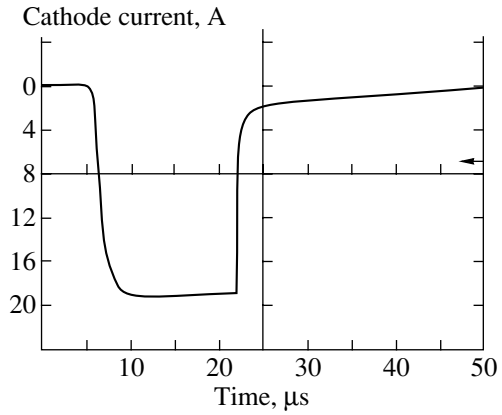


Fig. 5. Dynamic characteristic of unirradiated MCT.

This creates a bypass channel for excess carriers to pass into the cathode, and the thyristor is turned off. The turn-off capability of a thyristor depends crucially on the parameters of the *P*-channel off-FET. The resistance R_{as} between the base and the emitter of the transistor *T*2 is the same as the resistance of the anode shunts.

2.4. Static I–V Characteristics of a MCT

The MCT with an anode shunt is a unipolar device. Figure 4 presents the static current–voltage (*I–V*) characteristics of an MCT. With the positive anode voltage U_a and the negative or zero gate bias U_g, the MCT blocks the cathode–anode current. With a positive (relative to the cathode) bias applied to the gate, we obtain a family of characteristics of the *N*-channel on-FET. When the cathode current achieves a value equal to the thyristor holding current, the thyristor is latched. This portion of the cathode *I–V* characteristic is determined by the dynamic resistance R_d of the thyristor. Let us discuss the reverse portion of the characteristic at negative U_a and negative (relative to the cathode) gate bias $-U_g$. Since the thyristor anode is shunted, the ⟨anode P^+ emitter⟩–*N*-buffer junction is not a blocking one. In this case, the cathode current flows through the channel of the *P*-channel off-FET, forward-biased *P*-pocket–N^--base junction, and the resistance of the anode shunts flows into the anode. The voltage drop across the forward-biased *P*-pocket–N^--base junction is seen in the *I–V* characteristics at $U_a < 0$. This portion of the *I–V* characteristic is not the operating one, but it allows control over important parameters of the *P*-channel off-FET. A blocking voltage as high as 2500 V at leakage currents I_{leak} = 100–200 μA was obtained in the working mode (at positive U_a) with gate voltage $U_g \le 0$. The voltage drop across the device in the on state was 2.8 V at a current of 32 A.

2.5. Dynamic Characteristics

Figure 5 shows a dynamic characteristic of an unirradiated MCT. The anode current rises to 19 A in 3.5 μs. The turn-off front consists of two phases: fast and slow. During the fast phase, the anode current drops to 10% of the maximum value in 1.5 μs. In the slow phase, the residual current drops in 30 μs. This is due to the long carrier lifetime in the high-resistivity N^- base (~50 μs). Several methods of suppressing the carrier lifetime to reduce the turn-off tail are used. One of these is to introduce radiation defects acting as recombination centers. Irradiation with gamma quanta [7], electrons [8, 9], and protons [10] has been used. Each of these methods has its specific physical and technological features. While the proton irradiation is performed prior to wafer partitioning into chips and mounting into a package, the irradiation with gamma quanta and electrons is possible for packaged devices. In our case, the MCT was placed in a TO-218 package, so that only the electron and gamma irradiation was possible. Electrons with an energy of 2 MeV were used at an irradiation dose of 5×10^{13} cm^{-2}. After irradiation, the MCTs were annealed at 250°C for 2 h. The forward drop of irradi-

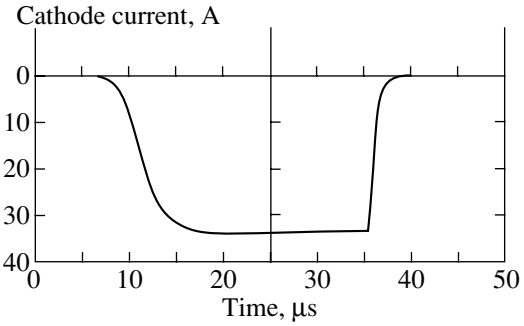


Fig. 6. Dynamic characteristic of irradiated MCT.

ated MCTs was higher—8 V at an anode current of 33 A, which corresponds to a current density of 100 A/cm^2. The turn-off tail was 2.5 μs long at a current of 33 A. The maximum turn-off current density exceeded 150 A/cm^2. A dynamic characteristic of an irradiated MCT is presented in Fig. 6.

3. DISCUSSION OF RESULTS

As shown above, the forward drop increases upon irradiation. To account for this increase, we use the model of a *p–i–n* diode, whose applicability to a thyristor was demonstrated in [11]. In a steady conducting state, a thyristor is similar to a *p–i–n* diode with holes flowing from the *P* emitter, and electrons, from the *N* emitter. The high density of electrons and holes leads to conductivity modulation, with a substantial decrease in the *N*-base resistance of the thyristor. The voltage drop across the *N* base is described by the expression [11]

$$V_N = 8kTbd^2/q(1+b)^2 D_a \tau_{\text{eff}}, \qquad (1)$$

where D_a is the ambipolar diffusion coefficient, $b = \mu_n/\mu_p$ is the ratio of electron and hole mobilities, $d = W_N/2$ is half the *N*-base width, q is the electron charge, and τ_{eff} is the effective carrier lifetime. According to [12], we assume that $D_a \approx 1$ cm^2/s, $d = 2.2 \times 10^{-2}$ cm, $b = 3$, $\tau_{\text{eff}} = 50$ μs, and $T = 100$°C. For an unirradiated thyristor, expression (1) yields $V_N = 1.17$ V, which is two times lower than the observed voltage drop of +2.8 V. It is worth noting, however, that we disregard here the voltage across the forward-biased N^+-emitter–*P*-pocket junction, and the voltage drop across the contacts. For an irradiated thyristor with a lifetime in the *N* base of $\tau_{\text{eff}} = 5$ μs, relation (1) yields $V_N = 5.4$ V, whereas the observed value is 8 V. This shows that the forward drop for an irradiated thyristor is to a large extent determined by the carrier lifetime in the *N* base, τ_{eff}.

In addition to the increased voltage drop, a rise in the maximum controlled current density was observed. No evidence of increase in the controlled current upon irradiation and annealing could be found in the literature. A possible explanation of this effect is as follows. The decrease in the lifetime shortens the hole diffusion length L_p. At $\tau_{\text{eff}} = 5$ μs, L_p is comparable with the distance between the *N* emitters (~10 μm), which reduces the probability of the switching of the hole current among the neighboring *N* emitters. This suppresses the effect of current pinching at large cathode areas and raises the controlled current density, which becomes as high as ~150 A/cm^2 for irradiated and annealed MCTs, with the overall controlled current exceeding 50 A.

4. CONCLUSION

The presented results demonstrate the possibility of controlling the turn-off tail of a high-voltage MCT within the limits from 50 to 1 μs and improving the dynamic characteristics of the MST by lowering the dynamic switching losses. An important result of this study consists in a two to threefold increase in the controlled current density and the widening of the device SOA. We believe that the latter effect is due to a more uniform distribution of the hole current. However, the higher voltage drop in the on-state and the resulting higher power loss in the steady-state mode should be taken into account in optimizing the turn-off tail length.

REFERENCES

1. T. Ogura, M. Kitagava, A. Nakagava, and H. Ohashi, IEEE Trans. Electron Devices **ED-38**, 1491 (1991).
2. M. Kegura, H. Akiyama, M. Tani, and S. Yamada, IEEE Trans. Power Electron. **5**, 430 (1990).
3. T. Ogura, A. Nakagava, M. Atsuta, *et al.*, IEEE Trans. Electron Devices **ED-40**, 628 (1993).
4. F. Bauer, E. Halder, K. Hoffman, *et al.*, IEEE Trans. Electron Devices **ED-38**, 1605 (1991).
5. F. Bauer, H. Haddon, T. Stockmeier, *et al.*, in *Proceedings of the International Conference MADEP 91, 1991*, p. 270.
6. H. Dettmer, W. Fichtner, F. Bauer, and T. Stockmeier, in *Proceedings of the International Symposium on Power Semiconductor Devices & ICs, ISPSD 94, 1994*, p. 13.
7. S. D. Brotherton and P. Bradley, J. Appl. Phys. **53**, 5720 (1982).
8. A. O. Evwaraye and B. J. Baliga, J. Electrochem. Soc. **124**, 913 (1977).
9. G. I. Andersson, M. O. Andersson, and O. Engstrom, J. Appl. Phys. **72**, 2680 (1992).
10. K. Nakayama and H. Matsuda, Jpn. J. Appl. Phys. **37**, 4751 (1998).
11. M. F. Chang, E. E. Wolley, and S. F. Bevacqua, in *Proc. Industry Appl. Soc., 1979*, p. 1068.
12. S. K. Chandi, *Semiconductor Power Devices* (Wiley-Interscience, New York, 1977).

Translated by D. Mashovets

Semiconductors, Vol. 35, No. 9, 2001, pp. 1110–1116. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1159–1166.

PHYSICS OF SEMICONDUCTOR DEVICES

Arrays of 128 × 128 Photodetectors Based on HgCdTe Layers and Multilayer Heterostructures with GaAs/AlGaAs Quantum Wells

V. N. Ovsyuk, Yu. G. Sidorov, V. V. Vasil'ev, and V. V. Shashkin*

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: shashkin@thermal.isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—A technology has been elaborated and photodetector modules based on $Hg_{1-x}Cd_xTe$/GaAs heterostructures and GaAs/AlGaAs multiquantum-well structures grown by molecular-beam epitaxy were fabricated for the 3–5 and 8–12 μm spectral ranges. The photosensitive HgCdTe layers were grown on the GaAs substrates with the intermediate buffer layer of CdZnTe. To decrease the surface effect on the recombination processes, the graded-gap $Hg_{1-x}Cd_xTe$ layers with x increasing towards the surface were grown. A silicon multiplexer was designed and fabricated by CMOS/CCD technology with a frame rate of 50 Hz. The hybrid microassembly of the photodetector array and the multiplexer was produced by group cold welding on indium columns while monitoring the connection process. The fabricated 128 × 128 modules based on HgCdTe layers with the cutoff wavelengths 6 and 8.7 μm had a temperature resolution of 0.02 K and 0.032 K, respectively, at a temperature of 78 K and a frame rate of 50 Hz. The photosensitive GaAs/AlGaAs multilayer quantum well structures were fabricated by MBE. It is shown that the technology developed allows 128 × 128 multielement photodetector arrays (λ_{peak} = 8 μm) to be produced with a temperature resolution of 0.021 K and 0.06 K at operating temperatures of 54 K and 65 K, respectively. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The following basic technologies are necessary for the development and fabrication of photodetecting modules:

(a) Molecular beam epitaxy (MBE) is needed for the growth of $Hg_{1-x}Cd_xTe$ layers and GaAs/AlGaAs multilayer structures;

(b) The capability to fabricate multielement photosensitive structures;

(c) The capability to fabricate a silicon multiplexer;

(d) A hybrid microassembly of photosensitive structures and a multiplexer;

(e) Development of the equipment and methods for testing and estimating the parameters of photodetecting modules.

In what follows, the main features of these technologies are considered and the parameters of the photodetecting modules fabricated are presented.

2. THE TECHNOLOGY OF MOLECULAR BEAM EPITAXY FOR $Hg_{1-x}Cd_xTe$ LAYER GROWTH

The most important initial parameters of material for producing high-sensitive infrared (IR) photodetectors are the electrical parameters (concentration of the majority charge carriers and their mobility, lifetime of the minority charge carriers) and the structural quality (the lowest dislocation density). To obtain the *p–n* junction, the optimal concentrations of carriers in the corresponding regions and a long lifetime are necessary. The latter, as a rule, is limited by the existence of traps in the film, which are caused by dislocations. This is especially important for the photodetectors operating in the 8–12 μm spectral range. The dislocation density in the material for such IR photodetectors should not exceed 10^5 cm^{-2}. For the detectors operating in the range 3–5 μm, the dislocations begin to affect their electrical characteristics for densities higher than 10^6 cm^{-2}.

MBE makes it possible to reduce significantly the cost of fabrication of CdHgTe material for large IR photodetectors by using alternative Si substrates with an intermediate CdZnTe layer. The advantages of MBE in the fabrication of epilayers in comparison with other methods make this method very easy to implement [1]. The application of analytical methods of monitoring based on the optoelectronic effects allows the epitaxial processes to occur with the parameters being monitored in real time. The best methods are the optical ones. Among them, ellipsometry has gained acceptance recently. This method makes it possible to measure the growth rate at the initial stage, the composition of the

growing layer for the complex compounds, variation in the growing surface relief, and the growth temperature.

The CdHgTe epilayers on the CdZnTe substrates are matched by the crystal lattice parameters and have a perfect crystallinity, which is mainly determined by the high crystal quality of the substrate. At present, these layers are used to fabricate multielement photodetector arrays with good photoelectric parameters. However, an increase in the photodetector sizes results in a sharp increase in their cost. The use of Si and GaAs substrates permits large-area epitaxial CdHgTe layers to grow. The structural quality of such layers turns out to be lower than that of similar layers on CdZnTe. Nevertheless, such layers are used to fabricate photodiode arrays with up to 256 × 256 elements for the wavelength range of 8–12 μm and arrays of larger sizes (up to 1024 × 1024) for the middle IR range.

An important parameter of the CdHgTe layers for the multielement photodetectors is the uniformity of the composition over area. In this case, for the photodetectors operating in the range of 8–12 μm, the variation of the composition should not exceed $\Delta x = 0.0002$ over 10 mm on the wafer with a diameter of up to 76 mm. For the growth of CdHgTe epitaxial films homogeneous over the thickness, it is necessary to develop molecular sources with almost unvarying molecular fluxes over a long period and to know the subsequent changes (resulting from a decrease in the material in the crucibles) to correct the temperature of the sources in order to maintain stable molecular fluxes [2].

We have developed and fabricated MBE equipment with an automated system for controlling the technological processes and facilities of the layer quality control in the growth process (*in situ*); this equipment is designed for the large-scale production of CdHgTe material on various substrates. The MBE setup represents a technological line designed on the basis of physicochemical studies of epitaxy from molecular beams on CdTe and GaAs substrates [3]. The technological module has a unique design and is intended for use in layer growth by MBE, including mercury-containing layers. The system developed for the automated control of the technological process precisely maintains the buffer layer growth conditions on the GaAs substrate and CdHgTe films. The use of an automated ellipsometer allows us to change the layer composition across the thickness according to the preset program and maintains this composition unchanged with a high accuracy by correcting the temperatures of the molecular sources. The heterostructure based on CdHgTe layers is shown schematically in Fig. 1; this structure is designed for the fabrication of multielement photodetectors.

CdHgTe with varying gap	$x = 0.22\ldots0.3$	$d = 0.5–1$ μm
Homogeneous layer of CdHgTe	$x = 0.22$	$d = 6–10$ μm
CdHgTe with varying gap	$x = 0.22\ldots0.3$	$d = 0.5–1$ μm
Buffer layer of CdZnTe		$d = 3–4$μm
GaAs substrate		$d = 300–400$ μm

Fig. 1. Schematic representation of the CdHgTe-based heterostructures grown by MBE.

3. PHOTODETECTOR 128 × 128 ARRAYS BASED ON CdHgTe HETEROEPITAXIAL LAYERS FOR MIDDLE AND FAR INFRARED RANGES

MBE which allows the CdHgTe layer composition in the growth process to be varied is a versatile method for fabricating photosensitive structures for infrared photodetectors. Variation in the operating-layer composition allows one to fabricate photodetectors possessing a high sensitivity in a wide spectral range. We have shown previously [4] that the introduction of a wide-gap layer at the boundaries allows one to eliminate the surface effect on the charge carrier recombination and to obtain high effective lifetimes.

When using multielement photodetecting devices based on CdHgTe layers obtained by MBE, the series spreading resistance of the substrate is an important parameter. A total current from individual *n–p* junctions, which flows in the CdHgTe *p* layer towards the base contact, can reach a magnitude when the voltage drop across the layer changes the operating point (the voltage bias) of the photodiodes. It has been shown recently [5] that MBE can be used to fabricate individual photodiodes with a low (down to 10 Ω) series resistance by means of layer growth with a specified profile. We have developed a technology for the fabrication of 128 × 128 photodetector arrays with a low series resistance; this technology ensures a uniform operating mode for each array element.

The photosensitive CdHgTe layers were grown on a GaAs substrate with a diameter of 76 mm and orientation (103) using an intermediate buffer CdZnTe layer. The growth temperature of the CdHgTe layer was 180–190°C, and that of the buffer layer, 240–300°C. The composition nonuniformity over the area of 1 cm^2 was no higher than $\Delta x = \pm 0.002$. The as-grown layers had *n*-type conductivity. After annealing at 230°C for

20 h, *n*-type conductivity converted to *p*-type. The hole concentration p and mobility μ_p in the film bulk at T = 77 K were 7×10^{15} cm^{-3} and 320 cm^2 V^{-1} s^{-1} (for a sample with λ_c = 6.0 μm) and 1.6×10^{16} cm^{-3} and 300 cm^2 V^{-1} s^{-1} (for a sample with λ_c = 8.7 μm).

The 128 × 128 photodetector arrays with spacing between the photosensitive elements of 50 μm were prepared on the CdHgTe layers using low-temperature planar technology [6]. Photodiodes were obtained by the ion implantation of boron with the ion energy of 120 keV. The area of *n*–*p* junctions was 17×17 μm^2.

Prior to the assembly with a multiplexer, the current–voltage characteristics and differential resistance dependences on the voltage bias were measured for arbitrary elements of photodetector arrays. The mean values of the product R_0A of elements of photodetector arrays were 6×10^3 and 22 Ω cm^2 for the middle and far IR regions, respectively.

4. TECHNOLOGY AND PHOTODETECTOR 128 × 128 ARRAYS BASED ON MULTILAYER STRUCTURES WITH GaAs/AlGaAs QUANTUM WELLS

Photodetecting elements based on multilayer structures with GaAs/AlGaAs quantum wells (MSQW) constitute a periodical structure of narrow-gap doped GaAs and wide-gap $Al_xGa_{1-x}As$ solid solution surrounded by conducting GaAs layers. For a GaAs layer thickness smaller than or comparable to the electron free-path length, a quantization of electron motion in the direction of the growth axis and splitting of energy states in the GaAs conduction band into minibands occurs; the miniband energies are given by $E_n \approx p_\parallel^2/2m^* + (h^2\pi^2/2m^*a^2)n^2$, where p is the electron momentum in the layer plane, m^* is the effective mass, a is the width of the GaAs quantum well (QW), h is the Planck constant, and n = 1, 2, … are integers. Photoconductivity in such a structure is effected by the optically active transitions (only allowed for the light polarization perpendicular to the quantum-confinement layers) between the lower occupied miniband and (i) bound states within the GaAs quantum well or (ii) the unbound states above the potential barrier (the conduction band edge of AlGaAs). The parameters of a photodetector (the quantum efficiency and the dark current) depend on the position of the quantized level with respect to each other and with respect to the potential barrier, on the value of the pulling electric field, on the electron concentration in the GaAs well, on the operating temperature, and on the size of an individual photodetector element.

Photosensitive structures were fabricated by MBE using a Riber 32P setup. As a substrate, the semi-insulating AGCHP-10 GaAs wafer with a diameter of 2 in. was used. The calculated thickness of the GaAs layers and the composition of the $Al_xGa_{1-x}As$ barriers were chosen so that the second level in the GaAs QW was below the $Al_xGa_{1-x}As$ conduction band edge by 5–15 meV. In this case, a maximum value of the ratio of the photocurrent to the dark current was attained for bias voltages of 2–4 V at the structure. The accuracy of obtaining the prescribed $Al_xGa_{1-x}As$ barrier composition and its uniformity over the area are of paramount importance in this case. A high uniformity of the layer thickness across the sample area in this setup (within $(x - x_0) < 0.01$) was achieved owing to the determination of an optimum angle of the manipulator position with respect to the sources and the substrate rotation axis during the growth process. It was established that the nonuniformity of the GaAs layer thickness over the wafer area did not exceed 1.5% for the optimum angle position of the substrate axis with respect to the molecular sources.

The specificity of Al evaporation implies a uniformity over the area no lower than for GaAs, which was confirmed by the measurements. In addition, we have used a method of growth interruption (which has been described in the literature) at the stage of completion of the GaAs well growth; this stage was characterized by smoothening of the surface. The smoothening time varies in the range of 10–30 s. The smoothening process was monitored by observing the intensity of the mirror reflection at the high-energy electron diffraction pattern. Studies of various MSQWs, which were grown under optimized conditions, by the methods of high-resolution transmission electron microscopy, low-temperature photoluminescence, and Raman scattering showed that the roughness of heterointerfaces does not exceed 1 monolayer. The thickness of the $Al_xGa_{1-x}As$ barrier layer was 400–500 Å and was chosen so that the tunneling component of the current between the ground quantum-confinement levels in the neighboring wells would be negligibly small compared to the thermal activation current in the temperature range of 60–80 K. The photosensitive layer consisted of 50 periods of GaAs QWs and $Al_xGa_{1-x}As$ barriers. The 1-μm-thick GaAs:Si conducting contacts were formed on both sides of a photosensitive layer. The Si concentration in the conducting contact layers was 2×10^{18} cm^{-3}. The Si concentration in quantum-confinement GaAs layers was selected in the range of $(4–5) \times 10^{11}$ cm^{-2}, which is close to the optimum concentration of MSQW for photoresistors. To eliminate the accumulation of Si impurities at the GaAs/AlGaAs interfaces, the doping of QWs was carried out in the central parts of GaAs at a distance of 2 monolayers from the heteroboundaries.

The reproducibility of the MBE-growth parameters of AlGaAs, homogeneity, the high structural quality of layers, and the accuracy of maintaining the $Al_xGa_{1-x}As$ composition allow one to fabricate structures with preassigned parameters. The calculated and experimental plots of dark current for MSQW structures with different spectral photosensitivity are shown in Fig. 2. It was assumed in calculations that the electron mobility μ_e =

2×10^3 cm^2 V^{-1} s^{-1} and the saturation velocity $v_s = 7 \times 10^7$ cm s^{-1} in AlGaAs with $x \approx 0.3$. The energy of a donor state in QW $E_D \approx 11.6$ meV.

The photodetector array was fabricated by photolitography and constituted periodically arranged mesa-structures obtained by wet etching, with a size of 40×40 μm and a period of 50 μm. The lower contact layer GaAs:Si is common to all the elements [7]. The nonrectifying contacts to the conducting GaAs:Si layers were obtained by the fusion of Ge/Au/Ni/Au. In order to ensure the IR radiation absorption in MSQW for a normal incidence of light at the photodetector array from the substrate side, a staggered two-dimensional diffraction grating with a period of 2.8–3 μm was formed at each photodetector-array element by electron and optical lithography. The etching of the grating was carried out in the CCl_2F_2 plasma to a depth of 0.7–0.8 μm. For the hybrid assembly of the photodetector array with a silicon multiplexer, the In-pillars 5–7 μm in height were prepared at the surface of a nonrectifying contact.

5. SILICON MULTIPLEXER

Two types of silicon multiplexers were developed. The first one featured direct access to the arbitrary elements of the photodetector array elements, had external storage, and was used for the detailed examination of the parameters of the array element massive: current–voltage characteristics, spectral sensitivity, noise parameters, etc. However, the multiplexer can also be used independently for obtaining an IR image for a low (no higher than 10 Hz) frame rate. Its advantage is the ability to operate with photodetectors for high dark currents (for a photodetector with a sensitivity cutoff larger than 12 μm) and elevated temperatures (up to 200 K).

The second type of multiplexer was designed and fabricated using CMOS/CCD technology with an *n* pocket. This multiplexer is used to read out photosignals from the array of CdHgTe photodiodes and from the array of GaAs/AlGaAs QW photoresistors. The readout scheme allows one to control the collection time for a fixed frame time, which provides the enhancement of the dynamic range of the IR photodetector device under a high level of input currents.

Its scheme is presented in Fig. 3. The multiplexer consists of 128×128 array readout cells, horizontal and vertical registers, a collection time controlling system, and an output unit. Each readout cell of the multiplexer is directly connected with a photodiode through an indium pillar and contains an input transistor; accumulation capacitance; and two switches T1 and T2, whose gates are connected to the address bus. With the use of an input transistor, the voltage bias is set at the photodetector (PD) and the photocurrent is read out. The photocurrent is integrated at the CCD capacitance under the accumulation gate. The accumulation capacitance is about 1 pF.

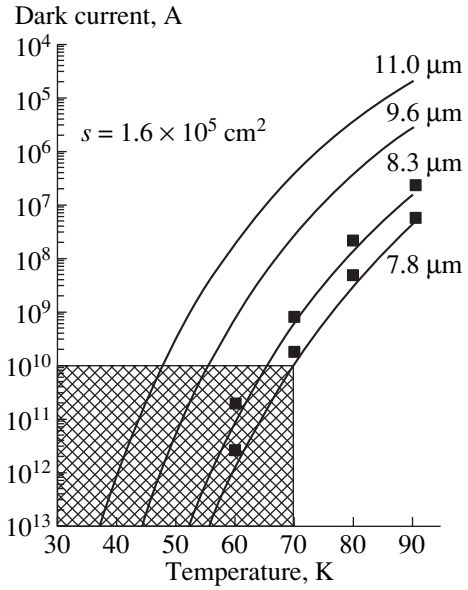


Fig. 2. Calculated (lines) and experimental (dots) dependences of the dark current of the MSQW GaAs/AlGaAs photoresistors with various positions of the photosensitivity peak. Hatched region represents the operating ranges of temperatures and currents for an array photodetector module with a charge capacitance amounting to 4×10^7 electrons.

The minimum collection time is equal to the readout time of two lines, and the maximum one, to the frame time. The mode of the electronic shutter provided in the multiplexer makes it possible to widen the dynamic range of input signals 64 times in comparison with the photocurrent integration during the entire frame time. The maximum input current of the multiplexer ranges from 0.3 to 20 nA for the maximum and minimum collection times, respectively.

6. HYBRID MICROASSEMBLY OF THE ARRAY PHOTODETECTOR MODULES

The assembly of the array photodetector modules based on HgCdTe was accomplished by cold welding of indium micropillars under pressure [6, 7]. The crystals of the multiplexer and photodetector array were oriented with respect to each other along the axes of the connected arrays so that the deviation from the ideal coincidence of the edge indium micropillars was no larger than 2 μm in the planar plane (*XY*), and the misalignment along the *Z*-axis did not exceed ±1 μm. In the course of crystal compression, the autoplanarization of their surfaces was provided. At the moment of completion of the cold welding, the planar surfaces become fixed at a preset distance with respect to each other. The lateral drift of crystals was no more than 2 μm in the course of welding. The limiting mechanical load is determined from the measurements of plastic-flow

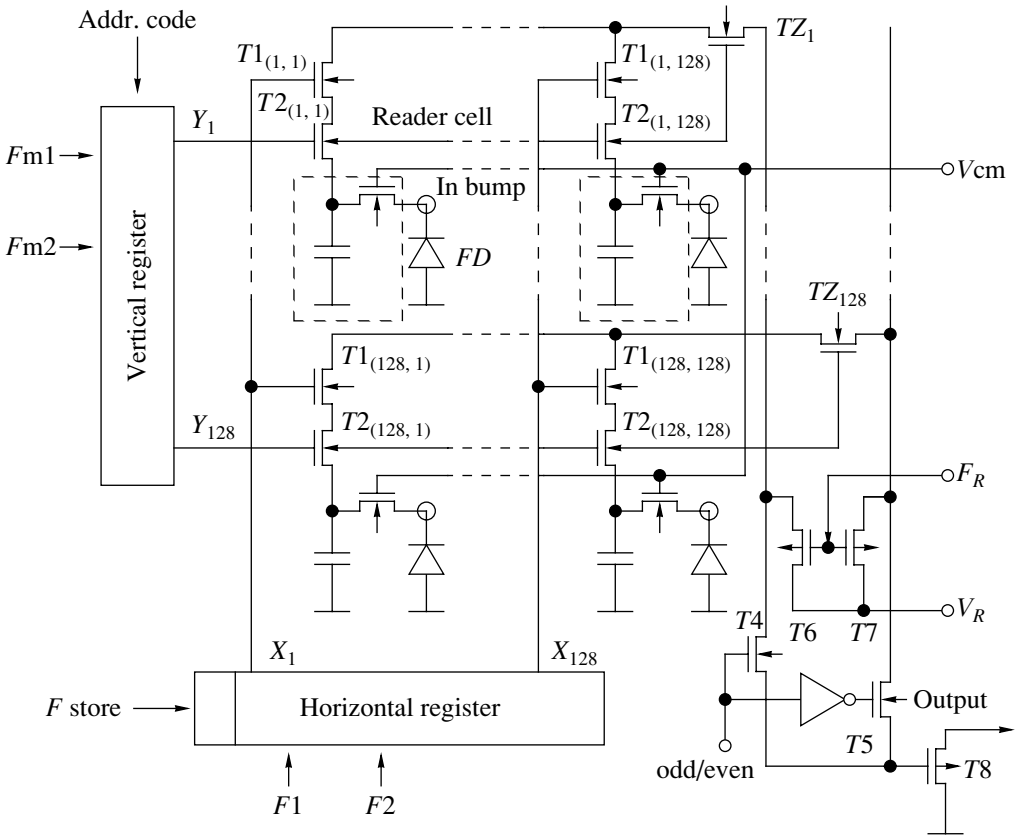


Fig. 3. Schematic diagram of the silicon multiplexer.

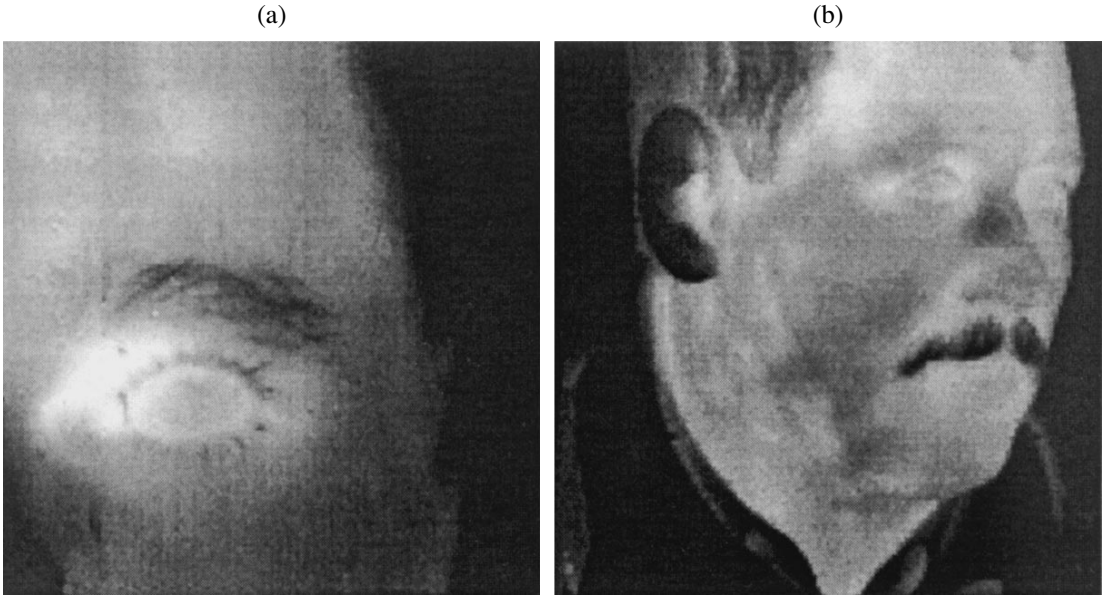


Fig. 4. Examples of the thermal images with CdHgTe-based 128 × 128 modules with a photosensitivity cutoff of (a) 6 and (b) 8.7 μm.

curves for the indium pillars [8], their height and area. The pressure necessary for the plastic flow of indium pillars ranges from 0.3 to 0.9 kg mm^{-2}. The studies of the effect of vertical compression on the *p*–*n*-junction parameters in the CdHgTe layers showed that their degradation starts at a pressure higher than 1.5 kg mm^{-2} [9]. Therefore, for reliable welding of indium pillars during the assembly of HgCdTe modules, we use a

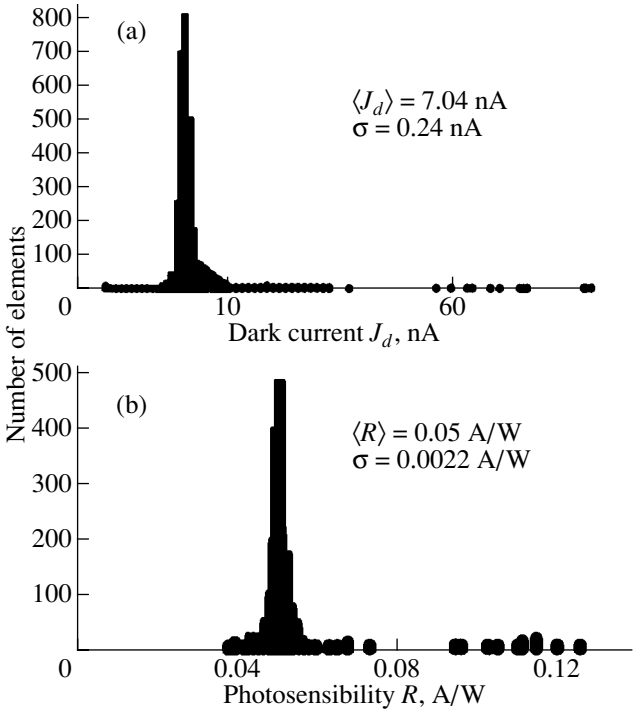


Fig. 5. Histograms of (a) the dark-current distribution in the photodetector module and (b) the photosensitivity for the array elements based on GaAs/AlGaAs MSQWs with the array size of 128 × 128.

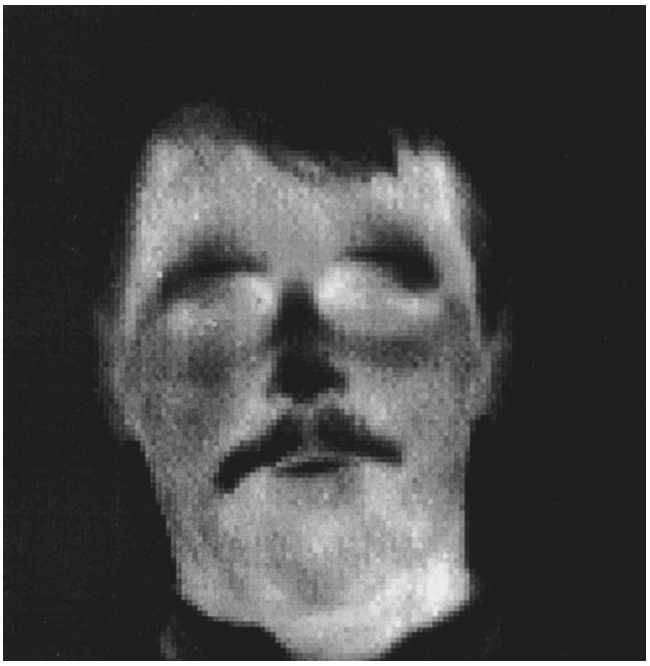


Fig. 6. The dark image obtained using a photodetector array module based on the GaAs/AlGaAs MSQWs with the size of 128 × 128. Spectral range is 7.5–8.5 μm, frame rate is 50 Hz. Operating temperature is 65 K.

pressure no higher than 1.0 kg mm^{-2}; in this case, the fraction of the cells connected reaches 99.9%. The assembly of GaAs/AlGaAs MSQW modules was performed by a similar technology, but, as opposed to HgCdTe, degradation was not observed for the electrical characteristics of the photoarrays when the pressure was increased to 5–6 kg mm^{-2}.

7. PARAMETERS OF THE PHOTODETECTOR MODULES BASED ON CdHgTe

Electrical parameters of the photodetecting modules were measured at the operating temperatures of 78–80 K and a background temperature of 293 K. The background flux was limited by the cooled diaphragm and, in the photosensitivity range, it amounted to 4.43×10^{15} and 2.97×10^{16} photon cm^{-2} s^{-1} for the first (λ_c = 6.0 μm) and for the second (λ_c = 8.7 μm) modules, respectively. The integration time of a signal for the first module was about 900 μs (collection during the lines for the frame rate of 50 Hz), and for the second module, 130 μs. In this case, for obtaining of an integration time shorter than the duration of a single line, the application of the bias at the photodiode was accomplished by a pulse, the duration of which actually determined the collection time. The measurement of the electrical signal was performed by a high-speed analog-to-digital converter (14 bit) with a fairly low intrinsic noise (less than 0.5 of the least significant digit), that allowed us to minimize the error in the noise determination.

In the table, the main parameters of the photodetector modules based on CdHgTe for the middle and far IR range are given.

The examples of the thermal images obtained with a lens F/1.6 at a frame rate of 50 Hz are shown in Fig. 4.

8. PARAMETERS OF THE PHOTODETECTOR MODULES BASED ON THE MULTILAYER STRUCTURES WITH GaAs/AlGaAs QUANTUM WELLS

As opposed to CdHgTe, the main source of the excess currents and noise is the thermal activation current in GaAs/AlGaAs MSQW, which makes it desirable to reduce the operating temperature of the module down to 45–65 K depending on the spectral range of the

Main parameters of the photodetecting modules

Long-wavelength sensitivity cutoff, μm	6	8.7
Array size	128 × 128	128 × 128
Interelement spacing, μm	50	50
Frame rate, Hz	50	50
Operating temperature, K	78	78
Temperature resolution, K	0.02	0.027

photoarray (Fig. 2). In our case, the modules with a photosensitivity peak near 8 μm were studied. The operating temperature of the module in the assembly of the silicon multiplexer described above was 60–65 K. In Fig. 5, the histograms of the photocurrent distribution of the photoarray and the dark current for a module with a photosensitivity peak near 8 μm are presented. The histogram of the dark current was measured in liquid nitrogen to minimize module overheating by the controlling currents of the switches and to provide precise monitoring of the temperature.

We relate the nonuniformity of the dark currents over the array elements to the growth defects, whose nature and intrinsic structure are not yet clear; however, we observed a correlation between the number of defects in the individual elements and the dark currents for several samples. Direct calculation shows that, for the photosensitivity values given above, the temperature resolution is 0.06 K at a temperature of 65 K. The example of a thermal image obtained with a lens $F/1.6$ for the frame rate of 50 Hz is shown in Fig. 6.

9. CONCLUSION

It is shown that the technologies developed for MBE growth of the CdHgTe/CdZnTe/GaAs heterostructures and multilayer structures with GaAs/AlGaAs quantum wells, the fabrication of the multielement photodetectors, and the fabrication of a silicon multiplexer and the hybrid microassembly make it possible to manufacture photodetector modules for the middle and far IR range with good photoelectrical parameters.

REFERENCES

1. J. M. Arias, M. Zandian, and S. H. Shin, J. Vac. Sci. Technol., B **9** (3), 1646 (1991).
2. V. S. Varavin, S. A. Dvoretsky, and V. I. Liberman, Thin Solid Films **267**, 121 (1995).
3. V. S. Varavin, S. A. Dvoretsky, and V. I. Liberman, J. Cryst. Growth **159**, 1161 (1996).
4. V. M. Osadchiĭ, A. O. Suslyakov, V. V. Vasil'ev, and S. A. Dvoretskiĭ, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33** (3), 293 (1999) [Semiconductors **33**, 297 (1999)].
5. V. S. Varavin, V. V. Vasil'ev, and T. I. Zakhar'yash, Opt. Zh. **66** (12), 69 (1999) [J. Opt. Technol. **66**, 1068 (1999)].
6. V. V. Vasilyev, D. G. Esaev, and A. G. Klimenko, Proc. SPIE **3061**, 956 (1997).
7. M. A. Dem'yanenko, O. R. Kopp, G. L. Kuryshev, *et al.*, Avtometriya, No. 4, 35 (1998).
8. A. G. Klimenko, V. G. Voĭnov, and A. R. Novoselov, Avtometriya, No. 4, 105 (1998).
9. L. N. Romashko, A. G. Klimenko, and A. P. Kravchenko, Proc. SPIE **3437**, 446 (1998).

Translated by T. Galkina

Semiconductors, Vol. 35, No. 9, 2001, pp. 981–984. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1026–1028.

THIS ISSUE IS DEDICATED TO THE MEMORY OF ANATOLIĬ VASIL'EVICH RZHANOV (1920–2000)

A Scientist, a Mentor, a Soldier

I. G. Neizvestnyĭ

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

e-mail: neizv@isp.nsc.ru

I met Anatoliĭ Vasil'evich Rzhanov for the first time in 1954 when I was assigned by the Moscow Power Institute to the Lebedev Physical Institute, first, to gain practical experience and then to do graduate studies. It is only now that I realize how young Rzhanov was at that time. Then, the 10-year difference in age gave the impression that the elder belonged to a quite different generation and embodied the image of a man engaged in activity in a mysterious field referred to as "science". Later, this image was complemented by new traits, sometimes unrelated to Rzhanov's main line of activity; however, these impressions always enhanced my positive opinion of him. Prior to my coming to the Lebedev Physical Institute, I had completed my undergraduate projects at one of the oldest semiconductor institutions (NII-35, later renamed Pul'sar), and I was greatly surprised by the situation at the Lebedev Physical Institute. Practically without any preliminaries, I was assigned to the task of translating several papers from English, to form an understanding of them, and to design and build the setup and develop a scheme for measurements. The task to study the surface properties of germanium in experiments with the field effect was given to me (still a student) without any doubts in the accomplishment of this study. I understood later that such an impromptu approach served in fact as a test of my ability to perform an experimental study at the Lebedev Physical Institute. In Rzhanov's team, there was a custom to convene weekly informal meetings. Each Thursday, all the research workers reported on the results obtained during the last week. Rzhanov used to form an understanding of the essence of the performed studies and drew all the other members of the team into the discussion. Such a collective method of discussing results and, of course, of planning the week ahead was very efficient. The task formulated and its deadline for completion were strictly observed, although they might be altered taking into account the opinions of members of the team and the external circumstances. Generally, it was very difficult to force Rzhanov to change his mind. Once a decision was made, he defended it to the end and brought it to fulfillment. Notwithstanding the fact that he was able to directly tell a person the truth (not always pleasant), Rzhanov was highly respected at the laboratory of semiconductor physics headed by B.M. Vul and, generally, at the entire Lebedev Physical Institute. It is only many years later that I became aware of Rzhanov's injuries and, to my complete surprise, of the fact that he was missing an eye.

The history of Rzhanov's arrival at the Lebedev Physical Institute and of the successive progress in his scientific career (as reconstructed from recollections of scientists that worked with him and from a few preserved official papers) can be confidently considered as heroic. In 1943, after a serious wound (the loss of an eye!) and a bad contusion, Rzhanov arrived in Moscow in order to draw up his papers for demobilization. There, he saw an announcement for students to enlist for postgraduate studies at the Lebedev Physical Institute and he decided to take a chance. Rzhanov had received his degree certificate from the Leningrad Polytechnical Institute in December, 1941. The entrance

1063-7826/01/3509-0981$21.00

exams were not easy, the examiners included well-known scientists such as Vul and G.S. Landsberg. At the beginning of a bleak winter, Rzhanov recovered from pneumonia with croup and then decided to go to his unit and get at least a greatcoat and boots, of which he was denied when he had been sent to hospital from the front. The marine's brigade in which he had served was deployed at the Oraniebaum bridgehead. An offensive was started at that time—a battle aimed at breaking the siege of Leningrad. The brigade suffered heavy losses, especially in officers, so that Rzhanov, although he was a "guest" and was officially discharged from the army on account of his injuries, had to take the command of his former reconnaissance company. Rzhanov was caught in heavy mortar shelling and was again heavily wounded and contused. For this battle, Rzhanov was awarded an order of Patriotic War. Having left hospital after a rather short time, he was once again in Moscow in January 1944, this time wearing a greatcoat and having spare boots, which was quite a treasure at that time.

In the spring of 1944, Rzhanov passed a postgraduate exam in electrodynamics. The examiners were again quite prominent scientists. In addition to those present at the exam, the list of examiners included S.L. Mandel'shtam, D.V. Skobel'tsin, and I.E. Tamm. Unfortunately, almost immediately after the exam, Rzhanov's head and hip wounds reopened. He suffered from the latter wound all his life; however, there were only a few who noticed and knew that, because Rzhanov never gave any indication. Rzhanov again spent many months in hospitals and sanitariums and only at the end of 1945 was he able to start his postgraduate studies. He defended his thesis on June 22, 1948. The results of Rzhanov's dissertation devoted to a new ceramic piezoelectric, barium titanate, were of much interest from the standpoint of both basic research and practical applications. Everything changed after his serious conversation with S.I. Vavilov who was at that time both the director of the Lebedev Physical Institute and the President of the Academy of Sciences of USSR. Vavilov persistently recommended Rzhanov to start research in the field of semiconductors. Thus started the period of Rzhanov's scientific activity related to semiconductors.

When I first appeared at the Lebedev Physical Institute, Rzhanov was in charge of a team of research workers experienced in growing germanium single crystals and involved in the development of a germanium transistor; by that time, he had started studies aimed at stabilizing the parameters of semiconductor devices, relating the causes of instability mainly to the surface properties of semiconductors.

Rzhanov was invariably friendly in his relations both with colleagues and subordinates. It was interesting, difficult, but somehow enjoyable to work with him. After parting with him late in the evening having discussed in detail all the problems, we were encountered in the next morning with his invariable question "Have you anything new to tell me?". Rzhanov always wholeheartedly participated in the parties organized at the laboratory and at the institute, in pranks and practical jokes; he astounded me once when I watched him dancing classical Charleston.

Rzhanov's ability to discuss competently any paper reported at seminars was incomprehensible for a young researcher like myself. Only much later, when I have looked through Rzhanov's archive and come across his notes of the books and monographs he read, I realized the amount of effort and time it took for Rzhanov to change his field of research from piezocrystals to semiconductors. Apparently, this thoroughness formed the basis of Rzhanov's great erudition in diverse fields of semiconductor physics and physics of semiconductor devices.

A specific feature of Rzhanov's personality was his ability to face dangers and difficulties. The first occasion that helped me to recognize this trait of Rzhanov's character was related to his doctoral dissertation. In 1962, having prepared the thesis, he decided to defend it in Kiev. However, Ukrainian scientists led by V.E. Lashkarev and V.I. Lyashenko disagreed with certain concepts forwarded by Rzhanov in his dissertation. His decision was of cardinal importance: Rzhanov was ready to defend his thesis in the home territory of his opponents, to argue the correctness of his conclusions then and there, and to face possible defeat. However, Rzhanov was sure that he was right, he proved it and won.

In the same year, rumors started to circulate about a possible transfer to Novosibirsk.[1] When asked whether I was willing to go to Novosibirsk, I replied affirmatively and soon forgot about it. There was a lot of work to be done at the Physical Institute and elsewhere, and this completely erased from my memory the vague prospects of transfer. In July, while spending my vacation on the Black Sea coast in Simeiz at a sports camp of the Lebedev Physical Institute, I received a telegram from the Presidium of the USSR Academy of Sciences; I was informed that I had been appointed the deputy director of the new Institute of Solid-State Physics and Semiconductor Electronics in Novosibirsk and that I was requested to come and start my new job right away. Naturally, I considered it as a practical joke, which were quite popular among my friends and colleagues at the Lebedev Physical Institute and in which I often participated myself. Having completed a wonderful vacation and returned to the Lebedev Physical Institute, I was astonished to get a harsh reprimand from Academician Vul, who had been repeatedly bothered with inquiries from the departments in the Presidium of the USSR Academy of Sciences about my whereabouts. I found out that Rzhanov, having been acquainted with

[1] At that time, the Novosibirsk Scientific Center of the just founded Siberian Division of the USSR Academy of Sciences was being organized (*note added by the English version's editor*).

the government's resolution concerning the founding of the aforementioned institute and without waiting for the corresponding decision of the Academy of Sciences, had gone on a long-planned and quite extended scientific commission abroad. Before his departure, Rzhanov left a paper at the Presidium, in which he suggested my appointment and listed a large number of tasks to be performed by me. He did not doubt in the slightest that I would not change my mind, he was completely sure that I would run all his errands; therefore, I could do nothing but follow his instructions. Once again, I became aware of Rzhanov's astonishing ability not only to make decisions but also to plan ways to fulfill them completely.

Later, the Council of Ministers officially approved Rzhanov's suggestion to erect a thermostatted block, i.e., a building with clean rooms built according to a completely new design. The layout designed by Rzhanov made it possible to combine the required clean-room conditions with a pleasant view of the forest from the window. This layout is still admired by guest scientists at the Institute of Semiconductor Physics.

Having succeeded in getting an appointment with A.N. Kosygin (at that time, the Chairman of the Council of Ministers), Rzhanov managed to secure substantial hard-currency funding, which formed the basis for a large amount of the Institute's experimental equipment and technological facilities.

Rzhanov (in collaboration with S.I. Stenin) organized the production of installations for molecular-beam epitaxy (MBE), first at the Institute and then at the experimental plant of the Siberian Division of the USSR Academy of Sciences; these installations were virtually the first to be manufactured in the Soviet Union. The Institute of Semiconductor Physics has not only been awarded a State prize for this activity, but also remains an acknowledged leader in using MBE to produce diverse thin-film structures.

All this became possible only because Rzhanov had immediately realized the capabilities of MBE, made the decision to concentrate most of the efforts of the Institute on this problem, and again saw the matter through to the end.

No effort was spared in designing and developing diverse instrumentation and installations. As an example, I mention a series of ellipsometers, the design and production of which was honored with a prize from the Council of Ministers; this line of research is successfully developing at present. K.K. Svitashev, a follower of Rzhanov, has played an outstanding role in solving this problem.

This was the case in many situations. For example, studies aimed at stabilizing the characteristics of metal–insulator–semiconductor (MIS) transistors resulted in the conclusion that these characteristics can be stabilized by using a two-layer insulator composed of silicon oxide and silicon nitride layers. However, when the hysteresis properties of a metal–nitride–oxide–semiconductor (MNOS) system were simultaneously uncovered, Rzhanov immediately organized an able team of researchers and technologists from the members of the staff of our Institute, the Institute of Inorganic Chemistry of the Siberian Division of the USSR Academy of Sciences, and the NPO Vostok. The result consisted in the development of nonvolatile memory elements that form the basic components of the computers existing at present; in addition, the NPO Vostok became a leading industrial institution active in the above field.

In just the same way (by concentrating large scientific and technological resources), a solution to the problem of multielement photodetector arrays designed to operate in the infrared region of the spectrum was found and was then transformed into an impressive practical application; this achievement was also awarded with a State Prize.

I would like to emphasize that almost all the aforementioned examples (as well as other projects on which I cannot enlarge owing to limited space) were based on an enormous amount of preliminary efforts in the field of the surface physics of semiconductors and from studies of the interfaces of a semiconductor with a metal, insulator, vacuum, gas media, and other semiconductors. It is the Rzhanov school in this field of semiconductor physics, also including studies in the field of physics and technology and in the development of thin films and thin-film structures, that has formed the basis of past, present, and future achievements of the Institute of Semiconductor Physics, Siberian Division of the Russian Academy of Sciences.

It was also of interest for me to watch how Rzhanov used to squash from the outset attempts at spreading malicious gossip and at starting squabbles. If a person (*X*) came to him and started to tell tales about somebody (*Y*), *X* immediately heard "Just a minute, I shall ask *Y* to come here, and then you can tell him everything to his face." Most often, the conversation was completed at that point. However, if someone came with serious complaints and agreed to a three way discussion, then all the problems were solved quickly and in a businesslike manner.

Rzhanov spent much energy and time educating the younger generation. I do not only mean his work as a lecturer and a tutor. Rzhanov was always ready to discuss any problem troubling young people, be it scientific, housing, or personal problems. It should be added that Rzhanov could easily and professionally discuss topics of literature, art, and nature (nowadays, we would refer to the last topic as ecology). He knew about all the latest publications in respectable journals and in newspapers, about new movies, and so on. If we add to this his phenomenal memory, then it will be understandable why any conversation with Rzhanov involved always attracted interest. He took part with pleasure in "philosophical seminars", which were regularly convened at that time; it was a rare occasion when some-

body managed to shake his leading position in any problem. It is my opinion that Rzhanov was a typical representative of professors, who, as we know from literature, always had encyclopedic knowledge.

The same knowledge allowed Rzhanov to behave very confidently during his scientific trips abroad and not only at official conferences but also during private meetings, where discussions about political, ethical, and other issues often arose. His excellent command of the English language and his feeling for the weak spots of an opponent should be added to this. From all points of view, trips abroad with Rzhanov were not only extremely productive professionally but were also enjoyable for his companions.

When Rzhanov had spare time (which was rare), he spent it hunting and boating on the Ob and the Ob sea. In spite of the fact that he had poor eyesight, Rzhanov was usually a leader in hunting; as for boating, I never knew a more daring and skillful "captain". I know that well because I repeatedly had to follow his boat dashing through high waves in stormy weather.

Summing up Rzhanov's life, his military exploits, his scientific achievements, and his relations with young scientists, friends, and colleagues, we can confidently consider this man as great scientist, mentor, and soldier.

The collection of papers offered here to readers and written by his direct followers and coworkers, who belong to the Rzhanov scientific school of the physics of semiconductor surfaces and the physics of thin semiconductor films is intended to introduce specialists to the latest achievements in this field of semiconductor physics.

Translated by A. Spitsyn

Semiconductors, Vol. 35, No. 9, 2001, pp. 985–987. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1029–1031.

ATOMIC STRUCTURE AND NONELECTRONIC PROPERTIES OF SEMICONDUCTORS

Atomic Processes in Semiconductor Crystals

L. S. Smirnov

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

Submitted February 14, 2001; accepted for publication March 1, 2001

Abstract—The formation of a new field in the radiation physics of semiconductors and semiconductor technology under the general guidance and with the direct participation of the late A. V. Rzhanov is reviewed historically. This line of research gave rise to a multitude of practical applications; however, most importantly, it forced scientists to radically change the established concepts of reactions in semiconductor crystals by taking into account the mobile defect–impurity subsystem susceptible to external factors. The concepts developed form the basis for considering the processes at the atomic level, especially during the formation and modification of active clusters and nanoobjects. © *2001 MAIK "Nauka/Interperiodica".*

In this paper dedicated to the memory of Anatoliĭ Vasil'evich Rzhanov, I relate the history of the formation and development of a promising line of research, the radiation physics and technology of semiconductors; this is one of the basic research fields at the Institute of Semiconductor Physics, and is what our institute is known for and is proud of. As the leader of a large team of scientists, Rzhanov exerted a deciding effect on the problem as a whole, and also on sections of it such as the effect of accelerated particles on the surface properties of semiconductors and heterointerfaces, the doping of thin layers using ion implantation, the restoration of equilibrium in crystals by irradiation, and the formation of metastable states using pulsed treatments; for all the above subfields, Rzhanov acted as the initiator and as an active participator.

When observing radiation effects in semiconductors, researchers came across a number of unexpected, seemingly contradictory phenomena related mainly to the generation of crystal-structure defects with an accompanying modification of all the basic semiconductor characteristics, rather than to the excitation of the electron subsystem. The diversity of, and the spread in, the experimental data for various, especially multicomponent, semiconductors was so large that we would have drowned in these data had we restricted ourselves to reliable results in certain special cases. Looking back, I remember with much gratitude how Rzhanov tactfully and insistently encouraged us to change from germanium to silicon; he believed that adequate models and the basis for promising technological methods in microelectronics and nanoelectronics could be formed only when dealing with a pure and structurally perfect material.

When, in the 1950s, the problems of developing "atomic" semiconductor electric cells and doping the crystals using irradiation with particles were formulated, it seemed to many scientists that these problems could be solved easily because the foundations of solid-state physics had been well developed.

The reality was found to be much more complex and interesting; it took decades to form correct concepts of radiation-induced processes in semiconductors. Actual crystals with their nonequilibrium properties and high concentration of background impurities caused great difficulties to experimenters and necessitated the development of new experimental methods and the transition to a higher materials-science level.

The modern concepts were reached quite slowly. There are many reasons for this; however, in my opinion, the main reasons are the following.

First, it was wrongly (although, widely) accepted that elementary point defects, i.e., vacancies (the vacant crystal-lattice sites) and interstitial atoms (the atoms pushed from the sites), had low mobility.

Second, the concept of the dominant role of electronic excitations in the defect formation was automatically transferred from the physics of dielectrics to that of semiconductors.

The situation was found to be so tense and mysterious that international conferences on radiation effects in semiconductors in the 1960s–1980s were dominated by discussions of the "vacancy paradox" and "the mystery of interstitial atoms."

The first steps were the most difficult; these were related to substantiating the hypothesis that the dominant role in the processes of the formation of active irradiation-induced centers in crystals belongs to secondary reactions between the atoms forced by radiation out of the crystal-lattice sites (interstitial atoms) and vacancies and between these defects and the imperfections existing already in the crystal (impurities, dislocations, interphase boundaries, quenched-in defects, and tracks of heavy particles). The formation of complexes accounted for the appearance of many types of active radiation-induced centers with corresponding changes

1063-7826/01/3509-0985$21.00

in electrical, photoelectric, optical, luminescent, and mechanical properties of crystals; however, this implied a high mobility of interstitial atoms and vacancies even at room temperature and below and, thus, contradicted the totality of traditional concepts of solid-state physics, especially the data on the thermal treatment of materials. Many scientists attributed this discrepancy to a difference between the mobility of vacancies generated by thermal treatment and by radiation. However, the standpoint according to which the vacancies are no different from each other irrespective of the method of their generation was found to be correct. However, the previous concepts of vacancy mobility were proved to be erroneous. For example, in crystals of germanium, silicon, and many other semiconductors, vacancies are found to be in a free state only during their transition from one trap (a complex) to another; the vacancy migration energy amounts to several tenths of an electronvolt and depends heavily on the vacancy charge state. It is this circumstance that led to a wrong conclusion about the low mobility of vacancies, which resulted in serious fallacies and wrong concepts.

The development of notions about the role of the second component of a Frenkel pair (the interstitial atom) was no less dramatic. The high mobility of interstitial atoms (even higher than that of vacancies), their neutrality, and the relatively low activity of interstitial-containing complexes formed the origin of a long-standing belief in (and the problem of) the "unobservability" of products of reactions involving interstitials. We had studied in detail the interaction of interstitial atoms with dislocations and interfaces and explained the effect of pushing the impurities out of lattice sites by self-interstitials; electron microscopy made it possible to observe the interstitial clusters. As a result, interstitial atoms were shown to be equal participants of the secondary radiation effects alongside vacancies; the aura of mystery around interstitials disappeared because it originated from lack of knowledge.

These were, perhaps, the most difficult and paradoxical situations and solutions to them. There were also others. It is only the discoveries of the role of energy barriers for the complex formation, the annihilation of Frenkel pairs at certain centers, and the dependence of all energy parameters and the mobility of point defects on their charge state (which, in turn, is a function of irradiation conditions and temperature) that made it possible to devise the models (at first, qualitative, and then, in certain special cases, quantitative) of radiation-induced processes.

All the aforesaid is related to structural rearrangements in crystals; these rearrangements encompass merely several atoms. Most often, the sizes of active centers generated by radiation prevent them from being directly observed. It is only the use of a combination of methods, both already known and those just developed, that made it possible to obtain the fairly comprehensive information about the properties of the primary, secondary, mobile and stable defects, and defect–impurity complexes. It is not an easy matter to list all the methods whose employment was found to be necessary to solve the problems stated above. These methods involve studies of electrical conductivity and the Hall effect; infrared spectroscopy; studies of luminescence, electron spin resonance, and photoconductivity; deep-level transient spectroscopy; studies of Rutherford backscattering and the channeling of light particles and of microhardness; electron microscopy; and so on.

At present, we have a new line of investigation that emerged at the interface between solid-state physics, the physics and chemistry of semiconductors, and atomic physics. This line, with its powerful basis of the theory of radiation-induced modifications and the methods for forming the metastable systems, gave rise to numerous practical applications. The latter include ion implantation; pulsed treatment of materials (which won the State Prize in 1988); ion-implantation synthesis; the introduction of active centers, especially, using high-temperature irradiation; the formation of disordered systems; radiation-enhanced diffusion; controlling the properties of a metal–insulator–semiconductor system; the doping of semiconductors using neutron-induced nuclear transmutations; and the enhancement of the stability of materials and devices.

Original experimental results and generalizations of them were published in several books [1–5] written by teams of authors and translated into foreign languages. The relevance of all the problems listed continues to heighten as the knowledge about actual crystals becomes more profound and as the requirements imposed on the developed devices and technologies become more stringent.

The aforementioned basic facts radically changed the established notions about reactions in crystals with the elementary structural defects involved. The modification of a defect–impurity subsystem in semiconductors became an important field of research and technological efforts.

We believe that an urgent task at present consists in extending the results of radiation physics of semiconductors to a wide field of problems in the physics of actual crystalline materials. This requires deep theoretical insight and the continuous improvement of experimental methods. Below, I dwell on two of the aforementioned problems, which have a complex history and are most general.

The first problem is related to diffusion in crystals. If experimentally observed point defects are mobile even at low temperatures, we may assume that any impurity atom at an interstice and any elementary complex of this atom with vacancies or self-interstitials (an impurity atom at the lattice site may be also considered as trapped) are relatively mobile. In this respect, diffusion is the result of repeated events of trapping and of the trapping-complexes dissociation stimulated ther-

mally. As a result, the results of diffusion depend heavily on the presence of observed and, especially, unobserved potential participants in the complex formation; these results also depend on the diffusant's chemical activity and on the charge state of reactants. In practice, we encounter here serious difficulties related to the fact that the initial impurity–defect composition of a crystal is, as a rule, inadequately known and the energy parameters of the defect–impurity reactions depend heavily on the experimental conditions; in addition, the same impurity may be involved in forming dissimilar trapping centers.

The processes of diffusion in actual materials are very complicated, and it is difficult to gain deep insight into these processes; this is also typical of these processes in crystals. As a result, we have requirements imposed on the reliability of experimental data and the necessity of taking into account the relative value of models.

The second problem is related to the nonequilibrium properties of a defect–impurity system. The crystals, especially synthesized, are found to be in a nonequilibrium state at the technological and operation temperatures; the defect–impurity subsystem of these crystals is rearranged when the mobile point defects are generated by irradiation and tends to a more equilibrium state. This effect is most pronounced when the initial matrix subsystem is far from equilibrium. The kinetics of the transition depends on both the initial parameters and new conditions (temperature, the generation rate of defects, and the ionization level). The process is common to all solids, even for those as perfect and ultimately pure as silicon of state-of-the-art quality. I now illustrate this using the example of doped silicon. The equilibrium solubility of impurities decreases rapidly upon transition to operating conditions, and silicon crystals are found to be supersaturated even at a low level of doping. The supersaturation is removed by irradiation, and the impurity–defect subsystem tends to equilibrium under the given new conditions; the transition is fairly complicated, because the process involves not only the dopant but also the entire set of impurities; the crystal boundaries and the entire history of mechanical and chemical treatments should be also taken into account.

The above is related to the actual structure of a semiconductor crystal and to the complexity of atomic-level processes occurring in this crystal; understanding these processes amounts to solving the many-body problem, with the bodies endowed with individual properties. It is because of this that a correction of conventional notions about reactions in crystals is usually met with resistance. It is pertinent here to cite the following simple (radiation-related) example. It has been proven that the formation energy of a Frenkel pair (E_d) decreases severalfold as the temperature of irradiated materials increases, which is related to introducing oscillations into the crystal lattice and to the "collective" mechanism of the primary event. However, in the majority of cases, researchers are still reluctant even today to consider E_d as a variable, although they deal with a wide temperature range. This fallacy becomes basic when we attempt to devise quantitative models for ion implantation or transmutation-induced doping.

The technological capabilities of radiation-related methods when applied to semiconductor crystals and structures are wide and the implementation of these methods is apparently just beginning. The term "defect engineering" has appeared in recent publications; this term implies a nonequilibrium modification of the impurity–defect subsystem with the aim of obtaining new characteristics of the crystal, structure, or devices by forming the required active centers or nanoclusters. It may be expected that the most interesting results should be obtained when forming a system with nanoobjects by radiation-related methods; of course, this requires new experimental and theoretical efforts, an improvement in the general technological level, and the development of new methods and techniques. We envisage in this the realization of Rzhanov's ideas and principles in creating the component basis for solid-state electronics of a new generation.

REFERENCES

1. V. V. Bolotov, A. I. Vasil'ev, N. N. Gerasimenko, A. V. Dvurechenskiĭ, G. A. Kachurin, V. I. Popov, L. S. Smirnov, and V. F. Stas', *Physical Processes in Irradiated Semiconductors* (Nauka, Novosibirsk, 1977).
2. L. S. Smirnov, S. P. Solov'ev, V. F. Stas', and V. A. Kharchenko, *Semiconductor Doping by Nuclear Reactions* (Nauka, Novosibirsk, 1981).
3. V. V. Bolotov, A. I. Vasil'ev, A. V. Dvurechenskiĭ, G. A. Kachurin, N. B. Pridachin, L. S. Smirnov, and V. F. Stas', *Problems of Semiconductor Radiation Technology* (Nauka, Novosibirsk, 1980).
4. A. V. Dvurechenskiĭ, G. A. Kachurin, E. V. Nidaev, and L. S. Smirnov, *Pulsed Annealing of Semiconducting Materials* (Nauka, Moscow, 1982).
5. A. L. Aseev, L. I. Fedina, D. Hoehl, and H. Barsch, *Clusters of Interstitial Atoms in Silicon and Germanium* (Nauka, Novosibirsk, 1991; Akademie Verlag, Berlin, 1994).

Translated by A. Spitsyn

Semiconductors, Vol. 35, No. 9, 2001, pp. 988–991. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1032–1035.

ATOMIC STRUCTURE AND NONELECTRONIC PROPERTIES OF SEMICONDUCTORS

Investigation of Ge Film Growth on the Si(100) Surface by Recording Diffractometry

A. I. Nikiforov*, V. A. Cherepanov, and O. P. Pchelyakov

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: nikif@isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—A diagram of the structural and morphological state of a Ge film on a Si(100) surface is constructed using *in situ* recording of the reflection high-energy electron diffraction patterns. The diagram involves the following regions: the continuous film, hut- and dome-shaped clusters, and dome-shaped clusters with misfit dislocations at the interface. The variation in the lattice parameters of the Ge film during MBE growth on the Si(100) surface is measured for the first time, and the oscillations of the variation in the lattice parameter for the (100) plane during two-dimensional layer-by-layer growth are found. *© 2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Interest in Ge nanoclusters is related to advances in the development of the technology for obtaining an array of islands fairly uniform in size. Researchers succeeded in diminishing the nanoclusters to sizes permitting the manifestation of quantum-well effects up to room temperature [1]. The Ge-on-Si heterostructure is ideal for investigating the transition from the layer-by-layer film growth to the formation of three-dimensional (3D) islands (the Stranski–Krastanov mechanism). At relatively low synthesis temperatures, there are no misfit dislocations (MDs) in such islands even if their thickness significantly exceeds the critical one [2].

Several stages of island evolution are experimentally observed for the Ge–Si system in the course of increasing the effective film thickness. The appearance of islands is observed subsequent to the formation of the continuous Ge film. The onset of the formation of 3D clusters is accompanied by the appearance of elongated spots in the reflection high-energy electron diffraction (RHEED) patterns. These spots correspond to electron scattering at the {105} faces. Owing to their configuration, such islands were termed hut-shaped clusters [3]. As the average film thickness increases, the {113} and {102} faces manifest themselves in the RHEED patterns in addition to the {105} faces. This stage is characteristic of the formation of dome-shaped clusters. The last stage of morphological and structural evolution of the Ge_xSi_{1-x} islands on Si(001) is the formation of 3D plastically strained islands with MDs at the layer–substrate interface, which is accompanied by a rapid increase in the island size (see, for example, [4–6]).

The dependence of the island parameters on the conditions of a technological process calls for the continuous monitoring of the growth surface. A conventional technique suitable for this purpose is RHEED. The comparative analysis of RHEED patterns and surface images obtained by scanning tunneling microscopy, which was aimed at correcting the interpretation of diffraction patterns in the course of film growth, was carried out previously [7]. This approach may be used, for example, to determine the critical thickness of the pseudomorphic film and the orientation of island faces [6]. Through using RHEED, the phase diagrams of the structures existing on the surface during Ge-on-Si epitaxy were constructed. The data on strains in the growing layer are of special interest, since these strains constitute the major driving force for the morphological transformations observed [8]. These strains can be evaluated from the variation in the lattice parameter of the growing Ge film with transition from the layer-by-layer growth to the island growth as well as with changes in the island shape. The variation in the lattice parameter in the course of heteroepitaxy may be determined from reflection positions in a RHEED pattern [9, 10].

The purpose of this study was to refine the structural-phase diagram and to determine the structure of the crystal lattice of the growing Ge film on the Si(100) surface for various stages of morphological transformation.

RESULTS AND DISCUSSION

The synthesis was carried out in a Katun'-S MBE installation provided with two electron-beam evaporators for Ge and Si. The dopant (Sb and B_2O_3) was evaporated from effusion cells. The analytical part of the

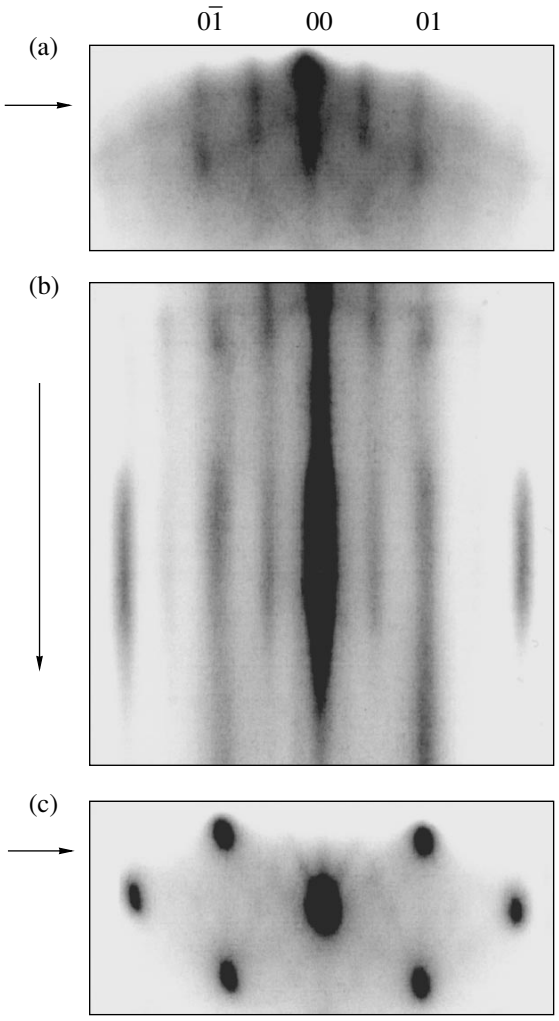


Fig. 1. The RHEED intensity variation during Ge film growth on Si(100): (a) and (c) are initial and final RHEED patterns, and (b) demonstrates the intensity variation during growth along the line shown by an arrow.

chamber consists of a quadrupole mass-spectrometer, a quartz thickness gage, and a high-energy (20 keV) electron diffractometer. During growth, the diffraction pattern is recorded with a CCD camera, and the image is fed to a personal computer. The software allows us to monitor both the image as a whole and selected parts of the diffraction pattern at a rate of 10 frame/s. The growth rate of the Ge layer was 10 monolayers (ML) per minute, and the temperature (T) was varied from 200 to 700°C. The Si (100) wafers with a misorientation less than 0.5° were used as the substrates. Prior to the growth of the Ge film, a high-temperature annealing of the wafer with the Si buffer layer growth was carried out to recover the initial surface.

The variations in the diffraction pattern qualitatively represent the variation of the growing film morphology. Quantitative information may be obtained from measuring the diffraction pattern intensity. The analysis of

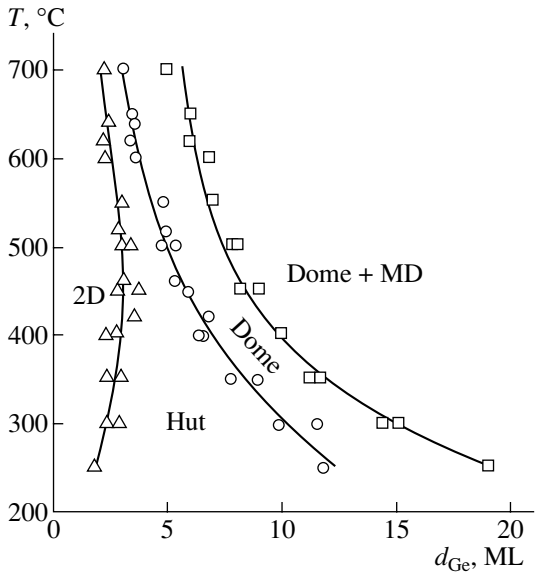


Fig. 2. Phase diagram of the Ge film structural state on the Si(100) surface.

the intensity variation in the azimuth [110] along the zero streak permitted us to separate the regions of two-dimensional (2D) growth and the growth of hut-shaped and dome-shaped clusters [8]. To determine the instant of the plastic relaxation of the dome-shaped clusters, the variation in the diffraction pattern intensity along the line intersecting the streaks and bulk reflections was also recorded (Fig. 1). The distance between them corresponds to the size of a 2D atomic cell $a_{\|}$ in the substrate plane. The initial and final diffraction patterns are shown in Figs. 1a and 1c. The arrow indicates the line, along which the intensity was recorded. The variation in intensity along this line with variation in the thickness of the Ge film deposited is shown in Fig. 1b. From the variation in the distance between the bulk reflections, which correspond to dome-shaped clusters, the instant of the plastic relaxation of the Ge film, which contains the MD network at the interface, was determined. The variation in the $a_{\|}$ quantity is 4% and coincides with the Ge–Si lattice mismatch. Based on these experimental data, the refined structural-phase diagram was constructed (Fig. 2). The diagram demonstrates the regions of the existence of stressed and plastically-strained continuous and island films depending on the Ge layer thickness d_{Ge} and the deposition temperature T.

As was demonstrated above, recording the variation in the diffraction pattern intensity can yield valuable quantitative information on the variation of the growing film morphology.

It can be noted from Fig. 1 that the distance between the spots of the $0\bar{1}$ and 01 reflections changes during growth. This may correspond to the variation in the

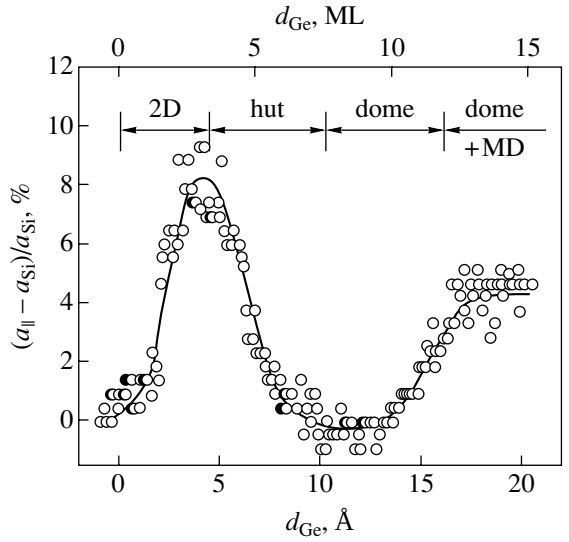


Fig. 3. Variation in the lattice parameter during Ge film growth on Si(100).

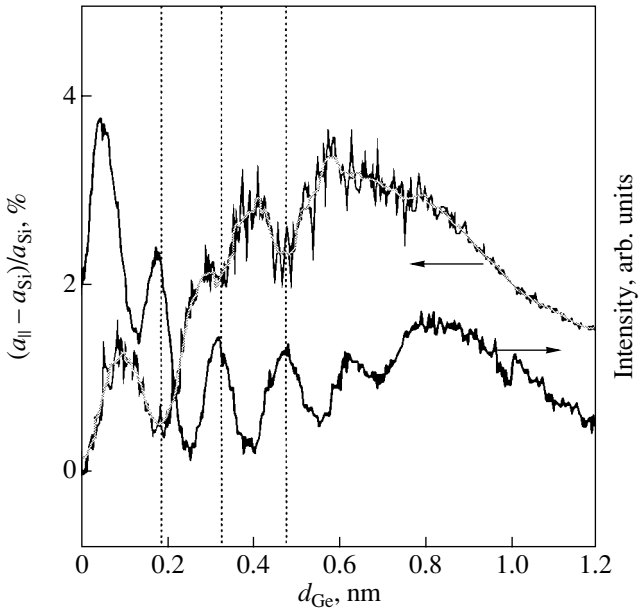


Fig. 4. Oscillations of the lattice parameter and specular reflection intensity in relation to the Ge film thickness. Dotted lines specify the position of peaks and valleys.

interatomic spacing $a_{\|}$ of the Ge film in the (100) plane. Reflections from the (2 × 1) superstructure also undergo a similar variation in their angular position. The variation in the lattice parameter $a_{\|}$ relative to the lattice parameter for Si a_{Si} with variation in the effective thickness of the deposited Ge layer d_{Ge} is shown in Fig. 3. As can be seen from the dependence demonstrated, the lattice parameter undergoes substantial variations. From the comparison of the dependence obtained with the phase diagram and a diffraction pattern, we can separate the regions corresponding to various stages of Ge film growth on the Si(100) surface. During the 2D-layer growth and the initial stage of the appearance of the hut-shaped clusters, the variation in the parameter $a_{\|}$ is as large as 8%. The next region corresponds to the formation and growth of elastically relaxed Ge islands in the form of hut-shaped clusters. In this case, the $a_{\|}$ magnitude for the island surface decreases to a value characteristic of the bulk Si crystal. Then, the hut-shaped clusters transform into dome-shaped clusters, which grow further. It is known that the MD network is formed at the Ge/Si interface during the growth of dome-shaped clusters. This network relieves the stresses caused by the Ge–Si lattice mismatch. Actually, it can be seen from Fig. 3 that the lattice parameter for the Ge film gradually approaches the value for the bulk material.

The variation in the lattice parameter for the plane parallel to the interface is measured continuously by *in situ* recording diffractometry. The reflection from the surface introduces the major contribution to the diffraction pattern. In this case, we may assume that an increase in $a_{\|}$ the during growth of the wetting layer indicates that the elastic relaxation begins to play a noticeable role even for 2D islands formed during the 2D-layer growth. Similar processes on the growth surface were observed previously [9]. The 2D Ge island rather than the continuous Ge layer on the Si surface is more likely to undergo elastic distortions in the growth plane. With an increase in the thickness of the continuous pseudomorphic Ge film, the elastic strain energy increases. This causes a progressive increase in $a_{\|}$ compared to the lattice parameter for Si. At the initial stage of the formation of a hut-shaped cluster, its size is comparable to the size of 2D islands. For this reason, no abrupt change in the thickness dependence of the lattice parameter is observed. As the cluster grows, the contribution of the surface to the intensity of the diffracted electron beam decreases, whereas that to the intensity of the beam passing through the bulk increases. Thus, a decrease in $a_{\|}$ down to the initial value apparently indicates that the lattice parameter for the bulk of the Ge island tends to a value equal to that for Si. Elastically strained regions may be found only in the (100) plane constituting a part of the faceting plane {105} of the hut-shaped cluster, which may be represented as the step $a_0/4$ in height and $5a_0/4$ in width, with the terrace lying in the (100) plane. The variation in the lattice parameter for the growth plane is confirmed by a periodic variation in the $a_{\|}$ value during the 2D growth similar to oscillations of specular reflection intensity. The variation in the lattice parameter $a_{\|}$ and the specular reflection intensity depending on the growing Ge film thickness are shown in Fig. 4. The periodic variation in the lattice parameter for the (100) plane is similar to the intensity oscillations, but is shifted by a half-period. The largest increase in the lattice parameter coincides

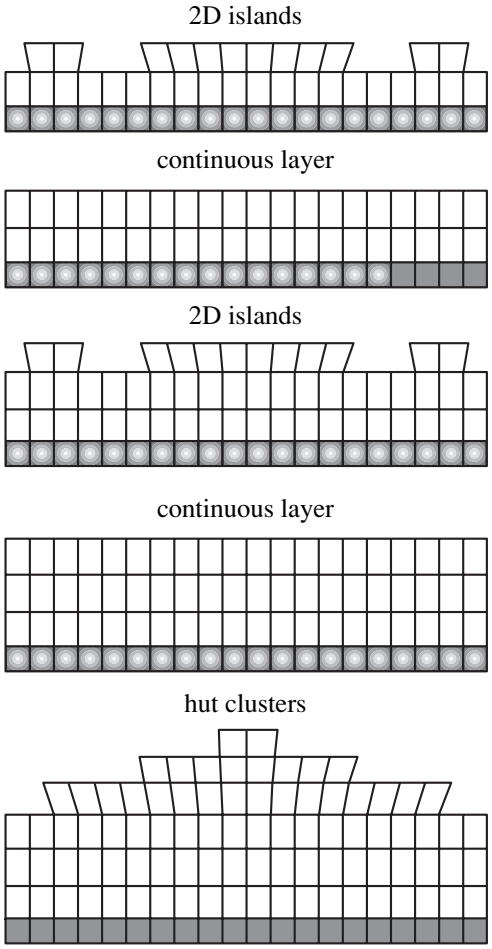


Fig. 5. Schematic representation of film growth.

with the minimum of the oscillation intensity. Such behavior is caused by the fact that the largest increase in $a_{\parallel}$ is observed for the roughest surface. A similar dependence for the growth of metals was observed previously [11]. The variation in the lattice parameter for the growing Ge layer is schematically shown in Fig. 5. The region of the 2D growth of the Ge layer (island and continuous film) and the initial stage of formation of the hut-shaped cluster are illustrated.

CONCLUSION

Thus, this study demonstrates the possibility of obtaining quantitative information about the atomic structure of the growing epilayer using the *in situ* recording of RHEED patterns. The refined diagram of the structural and morphological states of the Ge film on the Si(100) surface, which includes all transitions known, is constructed. The variation in the lattice parameter of the Ge film on the Si(100) surface during MBE growth is determined for the first time, and oscillations of the variation in the lattice parameter during 2D growth are observed. During 2D growth and the formation of hut-shaped clusters, the increase in the lattice parameter for the (100) plane is controlled by the elastic strain, whereas, for dome-shaped clusters, it is governed by a plastic relaxation due to MD network formation at the interface.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project nos. 00-02-18012, 00-02-17638, and 00-15-96806.

REFERENCES

1. A. I. Yakimov, A. V. Dvurechenskii, Yu. Yu. Proskuryakov, *et al.*, Thin Solid Films **336**, 332 (1998).
2. D. J. Eaglesham and M. Cerullo, Phys. Rev. Lett. **64**, 1943 (1990).
3. Y.-W. Mo, D. E. Savage, B. S. Swartzentruber, and M. G. Lagally, Phys. Rev. Lett. **65**, 1020 (1990).
4. F. K. LeGoues, M. C. Reuter, J. Tersoff, *et al.*, Phys. Rev. Lett. **73**, 300 (1994).
5. H. T. Johnson and L. B. Freund, J. Appl. Phys. **81**, 6081 (1997).
6. V. A. Markov, A. I. Nikiforov, and O. P. Pchelyakov, J. Cryst. Growth **175/176**, 736 (1997).
7. I. Goldfarb and G. A. D. Briggs, Surf. Sci. **433–435**, 449 (1999).
8. O. P. Pchelyakov, V. A. Markov, A. I. Nikiforov, and L. V. Sokolov, Thin Solid Films **306**, 299 (1997).
9. L. Kubler, D. Dentel, J. L. Bischoff, *et al.*, Appl. Phys. Lett. **73**, 1053 (1998).
10. A. Ohtake, M. Ozeki, and J. Nakamura, Phys. Rev. Lett. **84**, 4665 (2000).
11. P. Turban, L. Hennet, and S. Andrieu, Surf. Sci. **446**, 241 (2000).

Translated by N. Korovin

Semiconductors, Vol. 35, No. 9, 2001, pp. 992–996. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1036–1040.

ELECTRONIC AND OPTICAL PROPERTIES OF SEMICONDUCTORS

A Study of Galvanomagnetic Phenomena in MBE-Grown *n*-$Cd_xHg_{1-x}Te$ Films

V. S. Varavin*, A. F. Kravchenko, and Yu. G. Sidorov*****

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

** e-mail: varavin@isp.nsc.ru*

*** e-mail: krav@thermo.isp.nsc.ru*

**** e-mail: sidorov@isp.nsc.ru*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—The magnetic-field dependences of the Hall coefficient and the conductivity of *n*-type $Cd_xHg_{1-x}Te$ epitaxial structures were measured at 77 K. The structures were grown by molecular-beam epitaxy with a prescribed solid solution composition profile across the thickness. A specific feature of the obtained dependences is that the conductivity and the absolute value of the Hall coefficient decrease with an increasing magnetic field. The obtained experimental dependences can only be described in terms of a model including low-mobility electrons. It is shown that anodic oxide deposited onto the $Cd_xHg_{1-x}Te$ film surface makes the concentration of low-mobility electrons higher and that of anodic fluoride lower. The possible reasons for the appearance of low-mobility electrons are discussed. The most probable sources of such electrons are surface layers and electrical microheterogeneities in $Cd_xHg_{1-x}Te$ films. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The solid $Cd_xHg_{1-x}Te$ (MCT) solution is widely employed in developing high-efficiency IR detectors operating in the 3- to 5- and 8- to 14-μm spectral ranges. Successful use of this narrow-gap material as a sensitive IR cell requires that the electrical parameters of carriers should be determined with high precision. Commonly, the parameters are calculated from the dependences of the Hall coefficient and electrical conductivity on the magnetic field and temperature. However, these calculations encounter severe difficulties associated with the complex structure of the electronic spectrum and with the manner in which electrically active centers are formed in the bulk and at the boundaries of the layers: the existence of several kinds of carriers, the possible formation of two-dimensional layers at interfaces, microscopic heterogeneities, a large electron to hole mobility ratio, a pronounced nonparabolicity of the conduction band, etc. [1]. The most important question concerns the number of different kinds of free carriers. Without having an answer to this question, it is impossible to determine the electronic transport parameters and, consequently, to assess the quality of the material grown. In the *n*-type material, the problem of unambiguous and adequate characterization is made even more complicated by the presence of low-mobility ("heavy") electrons, leading to a strong dependence of the Hall coefficient and conductivity on the magnetic field. Heavy electrons are observed in MCT in bulk crystals [2] and films grown by liquid-phase epitaxy (LPE) [3] and molecular beam epitaxy (MBE) [4]. The reasons for the appearance of low-mobility electrons in a particular material are not always clear.

The most frequently mentioned reason for the field dependence of the Hall coefficient and conductivity in MCT is the compensation of the material by holes. However, this kind of compensation is absent in many cases and, in particular, measurements of the mobility spectrum suggest the presence of low-mobility electrons [5].

The aim of this study was to analyze the influence exerted by various treatments and techniques employed to prepare boundaries of *n*-type layers in MBE-grown MCT structures on the galvanomagnetic phenomena relied upon to determine the kinds of charge carriers and evaluate their electrical parameters.

LAYER GROWTH AND EXPERIMENTAL PROCEDURE

The MCT layers used in this study were grown by MBE [6] on GaAs substrates. A unit for MCT growth in the multiunit MBE machine is equipped with a set of annular sources ensuring a high composition uniformity on substrates up to 76 mm in diameter and a built-in fast automated ellipsometer for measuring the growth rate and continuously monitoring the MCT layer composition during the growth process. This monitoring allows the fabrication of structures with a prescribed composition profile across the thickness [7]. The MCT layers were 7–10 μm thick.

1063-7826/01/3509-0992$21.00 © 2001 MAIK "Nauka/Interperiodica"

The concentration and mobility of carriers in MCT layers were measured using the van der Pauw method. Samples 1×1 cm in size were cut for measurements from wafers 50.8 mm in diameter with heteroepitaxial MCT structures. Gold-plated pressure contacts were used. Control measurements on classical Hall structures with a strip width of 100 μm, fabricated by means of photolithography, yielded similar results. In measurements, the samples were submerged in liquid nitrogen in a special holder with a screen precluding background illumination. The current through the sample was in the range 0.1–3.0 mA; the results of measurements were independent of the current.

EXPERIMENTAL RESULTS AND DISCUSSION

Immediately after growth, MCT layers with cadmium telluride content $x = 0.2$–0.4 are *n*-type [8]. Annealing the structures at 200–300°C makes these layers *p*-type. In annealed *p*-type MCT layers, the Hall coefficient R_H is independent of magnetic field B (Fig. 1). This indicates that only carriers of a single kind are present, and the *p*-type MCT with a single kind of carrier [9, 10] is more likely to be prepared. An *n*-type MCT with a single kind of carrier can also be prepared [11], but in most cases *n*-type materials exhibit a field dependence of the Hall coefficient R_H and conductivity σ (Fig. 2). Points in Fig. 2 represent the experimental Hall effect data, and lines, the results of calculations for two kinds of carriers: mobile electrons with concentration $n_1 = 4 \times 10^{14}$ cm^{-3} and mobility $\mu_1 = 86000$ cm^2/(V s) and electrons with concentration $n_2 = 5 \times 10^{15}$ cm^{-3} and low mobility $\mu_2 = 600$ cm^2/(V s).

When a field dependence exists, the Hall effect data were processed in terms of the model including two kinds of carriers. The following expressions were used for the Hall coefficient R_H and conductivity σ [12]:

$$R_H = \frac{R_{H1}\sigma_1^2(1 + \sigma_2^2 B^2 R_{H2}^2) + R_{H2}\sigma_2^2(1 + \sigma_1^2 B^2 R_{H1}^2)}{(\sigma_1 + \sigma_2)^2 + \sigma_1^2\sigma_2^2 B^2 (R_{H1} + R_{H2})^2}, \quad (1)$$

$$\sigma = \frac{(\sigma_1 + \sigma_2)^2 + \sigma_1^2\sigma_2^2 B^2 (R_{H1} + R_{H2})^2}{(\sigma_1 + \sigma_2) + B^2\sigma_1\sigma_2(\sigma_1 R_{H1}^2 + \sigma_2 R_{H2}^2)}. \quad (2)$$

In expressions (1) and (2), R_{H1}, R_{H2}, σ_1, σ_2 are the Hall coefficients and conductivities associated with the first and second kinds of carriers, respectively, and B is magnetic induction.

It should be noted that the accuracy of determining the parameters of low-mobility carriers is very poor. For example, if electrons are taken instead of holes, the results of fitting the calculated field dependences to experimental points remain virtually unchanged. To improve the accuracy, it is necessary to measure the field dependence in a wide range of magnetic induction values.

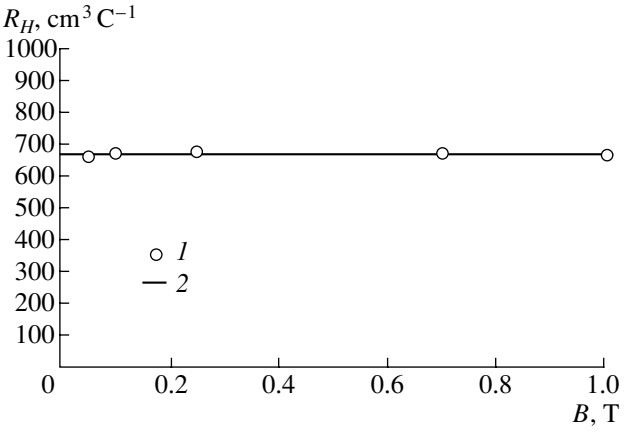


Fig. 1. Field dependences of the Hall coefficient for an MBE-grown *p*-type $Cd_xHg_{1-x}Te$ film with $x = 0.22$. (*1*) Experimental data and (*2*) results of calculation for hole concentration $p = 1 \times 10^{16}$ cm^{-3} and mobility $\mu_p = 585$ cm^2/(V s).

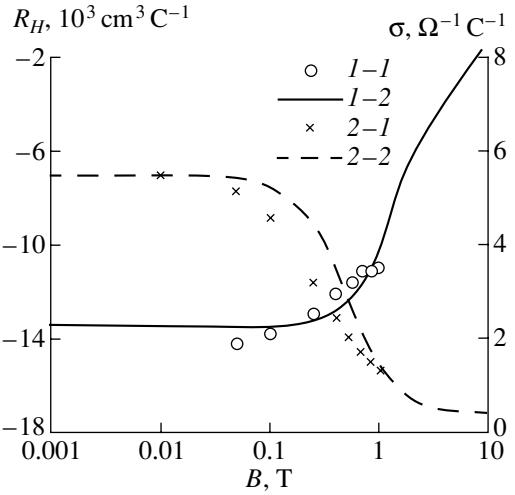


Fig. 2. Field dependences of the Hall coefficient and conductivity for an MBE-grown *n*-type $Cd_xHg_{1-x}Te$ film with $x = 0.22$. (*1–1*) Measured Hall coefficient, (*2–1*) measured conductivity, (*1–2*, *2–2*) calculated Hall coefficient and conductivity, respectively, for $n_1 = 4 \times 10^{14}$ cm^{-3}, $n_2 = 5 \times 10^{15}$ cm^{-3}, $\mu_1 = 86000$ cm^2/(V s), and $\mu_2 = 600$ cm^2/(V s).

The hole conductivity in MCT is due to vacancies in the metal sublattice [13]. The concentration of donors in MBE-grown MCT layers must be at a minimum because of the low growth temperature; indeed, this method does produce MCT layers with a low concentration (on the order of 10^{14} cm^{-3}) and a high mobility of conduction electrons. It follows from calculations of the equilibrium concentrations of donor centers introduced both by intrinsic point defects and by the most

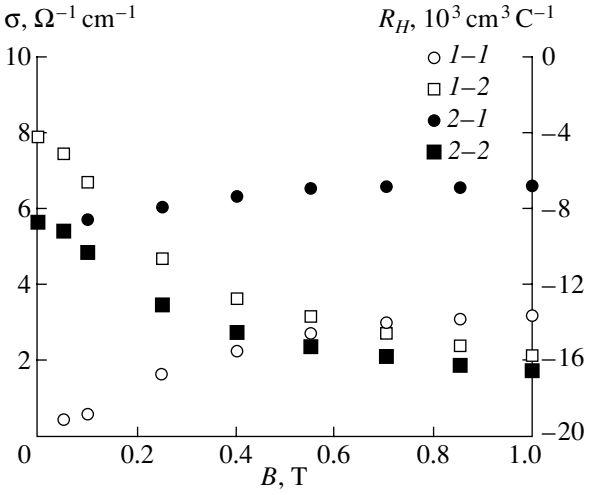


Fig. 3. Field dependences of the Hall coefficient and conductivity for an MBE-grown *n*-type $Cd_xHg_{1-x}Te$ film with $x = 0.22$ before (*1–2, 1–1*) and after (*2–1, 2–2*) deposition of anodic oxide.

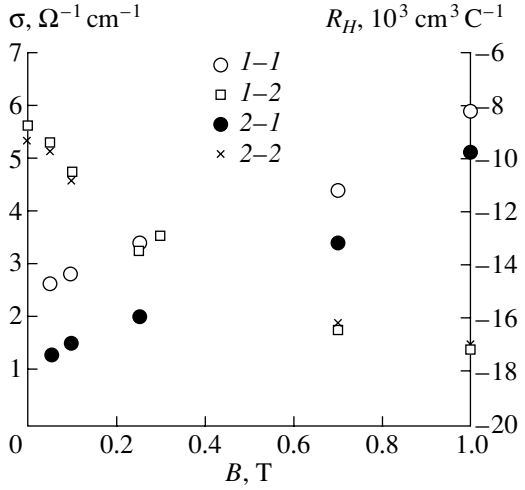


Fig. 4. Field dependences of the Hall coefficient and conductivity for an MBE-grown *n*-type $Cd_xHg_{1-x}Te$ film with $x = 0.22$ before (*1–2, 1–1*) and after (*2–1, 2–2*) deposition of anodic fluoride.

probable impurities that the equilibrium concentrations of donor centers are no higher than 10^7–10^{10} cm^{-3} under MBE conditions. As shown in [14], the most probable donor center in MBE-grown MCT layers is antisite tellurium, whose concentration is determined by the MCT crystallization conditions deviating from equilibrium. Calculations of the concentration of antisite tellurium in MBE-grown MCT with account of the deviation from equilibrium predict an increase in this concentration to 10^{14}–10^{15} cm^{-3}. The high concentrations of low-mobility electrons cannot be related to the bulk doping of a MCT film, being in all probability associated with the surface. It should be noted that low-mobility electrons observed in bulk MCT crystals [2] had a lower concentration, which is quite natural if these electrons are associated with the surface, since the films with low carrier concentration, studied in the present investigation, are rather thin (7–10 μm) and, therefore, the relative contribution of carriers from the surface layer is large.

If low-mobility electrons exist in the surface layer, then compositional changes in the surface layer must change the mobility of these electrons. For example, the mobility of low-mobility electrons must fall with increasing CdTe content in the surface layer.

Structures with graded bandgap layers on the surface were grown with the CdTe content at the surface varying from $x = 0.4$ to $x = 0.95$. However, no influence of the surface layer composition on the mobility of low-mobility electrons could be revealed. On etching off the graded bandgap layer (whose growth ended with the deposition of virtually pure CdTe), no noticeable changes in the field dependences of the Hall coefficient and conductivity were observed.

At the same time, various surface treatments affect the parameters of low-mobility electrons. For example, the deposition of anodic oxide leads to characteristic changes in the field dependences of the conductivity and Hall coefficient, shown in Fig. 3. The processing of the measurement results shows that the concentration of low-mobility electrons increases from 5×10^{15} to 3×10^{16} cm^{-3} at invariable characteristics of high-mobility electrons.

The deposition of anodic fluoride (by a procedure described in [15]) also affects the field dependences (Fig. 4). An anodic fluoride film lowers the rate of surface recombination and affects the band bending at the surface [16]. A sample onto which anodic fluoride was deposited was cut from the same wafer with MCT heteroepitaxial structure, which was used to prepare a sample for the deposition of anodic oxide. Processing of the obtained field dependences of the conductivity and Hall coefficient in terms of the model with two kinds of electrons shows that the anodic fluoride exerts the opposite influence to that of the anodic oxide, with the concentration of low-mobility electrons decreasing from 5×10^{15} to 1×10^{15} cm^{-3} and their mobility growing from <1000 cm^2/(V s) to >4000 cm^2/(V s).

The MBE technique employed to fabricate the structures allows the complete insulation of the *n*-layer from the surface. $Cd_xHg_{1-x}Te$ with $x > 0.19$ is easily made *p*-type by annealing at 200–220°C. At the same time, at a lower content of cadmium telluride, the layers remain *n*-type upon annealing of this kind. Therefore, since there are no surface electrons in *p*-type layers, it is possible to prepare a MCT structure having a layer with $x < 0.19$ confined between layers with a high CdTe content and to anneal this structure, with *n*-type conduction preserved in the layer with low CdTe content. This was actually done. Figure 5 shows the composition profile across the thickness *d*, measured ellipso-

metrically during structure growth. Prior to annealing, the structure had *n*-type conduction with field dependences of the conductivity and Hall coefficient similar to those presented in Fig. 3. The field dependences obtained upon annealing are shown in Fig. 6. A transition to *p*-type conduction is clearly seen at a high magnetic induction. Processing of the measurement results in terms of the model with two kinds of carriers (electrons and holes) satisfactorily describes the experimental data and gives the following concentrations $n = 8.4 \times 10^{13}$ cm^{-3}, $\mu_n = 125000$ cm^2/(V s) for electrons and $p = 1 \times 10^{17}$ cm^{-3}, $\mu_p = 200$ cm^2/(V s) for holes. No contribution from low-mobility electrons was revealed. This experiment not only excludes the surface as a factor leading to the appearance of heavy electrons but also eliminates, because of the low electron concentration, other reasons for the possible formation of heavy electrons (density-of-states tails [17]) and, even more importantly, an impurity band with hopping conduction. However, at high concentrations of donor impurity these factors cannot be eliminated.

The field dependence of conductivity was measured on long narrow samples cut from wafers with MCT heteroepitaxial structures, with the current flow direction coinciding with that of the magnetic field. Since the Lorentz force is zero in this case, there must be no field dependence of conductivity in a homogeneous isotropic sample. It follows from the experiment that, with the magnetic induction increasing from 0.05 to 1 T, the conductivity is changed to a varied extent (by up to 30%) in different samples. This indicates the presence of electrical heterogeneities in the samples studied.

The possible reason for the influence of the surface is the existence of a surface charge whose density depends on surface treatment. According to the theory of local electroneutrality, a foreign inclusion in MCT of, e.g., tellurium or cadmium, or a surface contaminant, shifts the Fermi level in MCT at the interface with the inclusion toward the electroneutrality level. In MCT, especially at low CdTe content, the Fermi level lies high in the conduction band [18], which means that the electron concentration near the inclusion is higher. The same result is obtained if the band structures of the inclusion material and MCT are considered—the Fermi levels of tellurium, cadmium, and mercury lie much higher than the local electroneutrality level, when their matching with the band structure of MCT shifts up the Fermi level in MCT at the boundary with the inclusion.

Low-mobility electrons may be present for at least the following four reasons:

(i) additional valleys in the conduction band;

(ii) enriched layers at the surface, which seems to be quite reasonable if account is taken of the high electron work function of MCT and the position of the local electroneutrality level above the conduction band bottom [18];

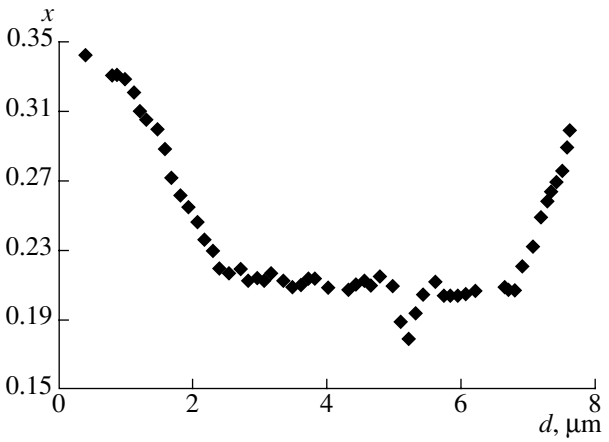


Fig. 5. Composition profile across the thickness of a MCT structure with narrow-gap interlayer, measured ellipsometrically during growth.

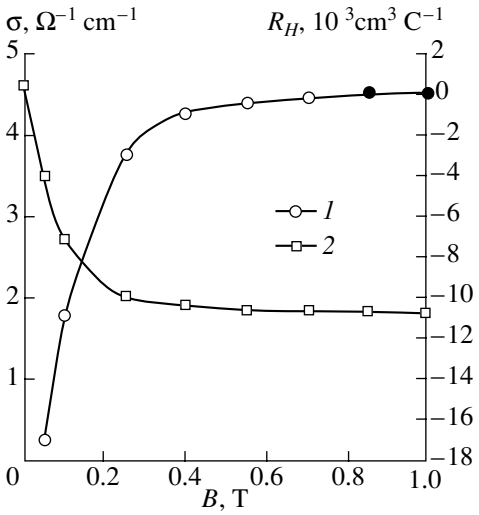


Fig. 6. Field dependences of (*1*) Hall coefficient and (*2*) conductivity of a *p*-type MCT structure with an *n*-type narrow-gap interlayer.

(iii) modulation of the conduction band bottom, due to random distribution of electrically charged centers giving rise to density-of-states tails;

(iv) formation of an impurity band with the appearance of hopping conduction.

The first factor is an inherent property of MCT itself and must be manifested irrespective of the MCT preparation technique or surface treatment. However, the specific features of the electronic structure must not give different concentrations of low-mobility electrons in samples with the same position of the Fermi level. In other words, if the presence of low-mobility (heavy) electrons results from the specific features of the band structure of MCT, then their concentration must unambiguously correlate with the concentration of light electrons (in the absence of holes, i.e., without compensa-

tion) and be independent of the method or growth conditions of MCT preparation. The results reported in [2] indicate unambiguously that, in the absence of pronounced compensation, the concentration of heavy electrons in bulk MCT [MCT crystals] falls with an increasing concentration of light electrons. This means that the experimental results obtained on bulk MCT crystals do not confirm the decisive influence of the band structure on the appearance of heavy electrons. Moreover, experimental investigations of the optical absorption spectra of MCT at photon energies exceeding the energy gap revealed no features related to the complex dispersion law [17].

If the free MCT surface is not (at least for the most part) the origin of low-mobility electrons, then it is necessary to consider the formation of regions in the bulk, in which the electron mobility is lower for some reason. The inclusion boundaries may well be regions of this kind. The density and size of inclusions must depend on the method of MCT preparation (bulk crystallization, LPE, MBE), since these techniques differ in relative component activities, growth temperatures, and growth rates reduced to temperature. However, MCT grown by any technique is characterized by the presence of low-mobility electrons and, consequently, the material should be characterized by the ratio of concentrations of high- and low-mobility electrons as a function of the concentration of high-mobility electrons. The higher this ratio at one concentration of high-mobility electrons, the smaller the number of inclusions with regions of low-mobility electrons around them. If, despite different conditions of MCT preparation, this ratio varies solely with the concentration of high-mobility carriers and is the same for different preparation techniques at equal concentrations of mobile carriers, then this is a fundamental property of the material.

CONCLUSION

The investigations performed indicate that the type of surface treatment of MCT films exerts strong influence on the results of Hall-effect measurements. Therefore, to clearly characterize *n*-type MCT films with a low concentration and a high mobility of conduction electrons, it is necessary to form a state of the surface with a known band bending and surface charge density and high uniformity of the charge density over the surface. The quality of the material can be characterized by the concentration ratio of high- and low-mobility electrons—films become more perfect when this ratio increases at a constant concentration of light electrons.

ACKNOWLEDGMENT

We thank the Program “Physics of Solid-State Nanostructures” for partial support of this study (grant no. 2000-3T).

REFERENCES

1. V. A. Pogrebnyak, I. M. Rarenko, D. D. Khalameĭda, and V. M. Yakovenko, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 319 (1998) [Semiconductors **32**, 288 (1998)]; Y. S. Gui, G. Z. Zheng, I. H. Chu, *et al.*, J. Appl. Phys. **82**, 5000 (1997); T. Thio and S. A. Solin, Appl. Phys. Lett. **72**, 3497 (1998); J. Antoszewski and L. Faraone, Proc. SPIE **2552**, 146 (1995).
2. E. Finkman and Y. Nemirovsky, J. Appl. Phys. **53**, 1052 (1982).
3. D. L. Leslie-Pelesky, D. G. Seiler, M. R. Loloee, and G. L. Littler, Appl. Phys. Lett. **51**, 1916 (1987).
4. J. Antoszewski and L. Faraone, J. Appl. Phys. **80**, 3881 (1996).
5. I. Vurgaftman, J. R. Meyer, C. A. Hoffman, *et al.*, J. Appl. Phys. **84**, 4966 (1998).
6. K. K. Svitashev, S. A. Dvoretsky, Yu. G. Sidorov, *et al.*, Cryst. Res. Technol. **29**, 745 (1994).
7. Yu. G. Sidorov, S. A. Dvoretsky, N. N. Mikhailov, *et al.*, Prikl. Fiz., No. 5, 121 (2000).
8. V. S. Varavin, S. A. Dvoretsky, V. I. Liberman, *et al.*, J. Cryst. Growth **159**, 1161 (1996).
9. V. S. Varavin, S. A. Dvoretsky, A. É. Klimov, and V. N. Shumskiĭ, Avtometriya, No. 4, 59 (1998).
10. L. He, J. R. Yang, S. L. Wang, *et al.*, J. Cryst. Growth **175/176**, 766 (1997).
11. O. A. Shegai, V. S. Varavin, S. A. Dvoretsky, *et al.*, in *Proceedings of the 8th International Conference on Narrow Gap Semiconductors, Shanghai, China, 1997*, p. 52.
12. R. A. Smith, *Semiconductors* (Cambridge Univ. Press, Cambridge, 1978; Mir, Moscow, 1982).
13. H. R. Vydyanath, J. Electrochem. Soc. **128**, 2609 (1981).
14. Yu. G. Sidorov, S. A. Dvoretsky, V. S. Varavin, *et al.*, in *Proceedings of the 2nd Russia–Ukraine Workshop “Nanophysics and Nanoelectronics”, Kiev, 2000*, p. 109.
15. E. Weiss and C. R. Helms, J. Electrochem. Soc. **138**, 993 (1991).
16. E. Weiss and N. Mainzer, J. Vac. Sci. Technol. A **6**, 2765 (1988).
17. V. Nathan, J. Appl. Phys. **83**, 2812 (1998).
18. V. N. Brudnyu, S. N. Grinyaev, and V. E. Stepanov, Physica B (Amsterdam) **212**, 429 (1995).

Translated by M. Tagirdzhanov

Semiconductors, Vol. 35, No. 9, 2001, pp. 997–1005. Translated from Fizika i Tekhnika Poluprovodnikov, Vol. 35, No. 9, 2001, pp. 1041–1049.

ELECTRONIC AND OPTICAL PROPERTIES OF SEMICONDUCTORS

Numerical Simulation of Intrinsic Defects in SiO_2 and Si_3N_4

V. A. Gritsenko*, Yu. N. Novikov*, A. V. Shaposhnikov*, and Yu. N. Morokov**

** Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia*

*** Institute of Computational Technologies, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

Submitted February 14, 2001; accepted for publication February 15, 2001

Abstract—The electronic structure of major intrinsic defects in SiO_2 and Si_3N_4 was calculated by the MINDO/3 and the density-functional methods. The defects that are of interest from the standpoint of their ability to capture electrons or holes were considered; these centers include the three- and two-coordinated silicon atoms, the one-coordinated oxygen atom, and the two-coordinated nitrogen atom. The gain in energy as a result of capturing an electron or a hole with allowance made for electronic and atomic relaxation was determined for these defects. The experimental X-ray spectra for both materials are compared with calculated spectra. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Amorphous silicon oxide (a-SiO_2) and silicon nitride (a-Si_3N_4) are the most important insulating materials in modern microelectronics. The oxide has a low surface-state density at its boundary with silicon and high chemical stability and features low leakage currents and a high breakdown voltage [1]. Silicon nitride is used in masks during the diffusion of impurities into silicon and its oxidation and has a high concentration of electron and hole traps ($\sim 5 \times 10^{18}$ cm^{-3}) [2].

A reduction in the channel length of a metal–oxide–semiconductor (MOS) transistor is accompanied with a decrease in the gate-insulator thickness. An increase in the field in the channel leads to the heating of electrons and holes in the channel and to their injection into the insulator. The capture of electrons and holes by traps in the insulator results in the accumulation of charge in the insulator, a shift of the threshold voltage, the insulator breakdown, and the degradation of a MOS transistor.

Oxide–nitride–oxide (ONO) structures are widely used in silicon-based devices; these structures have a higher effective dielectric constant compared to silicon oxide. Silicon nitride is used as an active medium in the Si–SiO_2–Si_3N_4–SiO_2–Si memory structures. At present, the arrays of electrically programmable read-only memories (EPROMs) are being developed on the basis of ONO structures with a capacity of 10^{12} bit/chip. The lifetime of electrons and holes at the traps in Si_3N_4 is as long as ~10 years at 300 K. The silicon-based EPROMs replace the magnetic and optical memory media.

Many experimental and theoretical studies have been concerned with gaining insight into the identity of the traps (their atomic and electronic structure) in SiO_2 and Si_3N_4. However, only the origin of one type of trap has been identified so far. It has been shown that the oxygen vacancy in SiO_2 acts as a hole trap [3].

In this paper, we report the results of theoretical studies on the electronic structure of intrinsic defects in SiO_2 and Si_3N_4; the defects of interest from the standpoint of their ability to capture electrons or holes were considered. The main task consisted in determining the energy gain ΔE resulting from the capture of an electron or hole by the defect with allowance made for electronic and atomic relaxation. Calculations were performed by the MINDO/3 method and that based on the density-functional theory (DFT) in the cluster approximation. The MNDO method was used in a number of cases.

Calculations using the MINDO/3 method were performed in the approximation of the unrestricted Hartree–Fock method using the same parameters as were used previously [4, 5]. Calculations using the DFT method were performed using an ADF software package [6]. We used the double zeta basis of the Slater-type functions including the polarization functions for all atoms. We considered both the valence and shell electrons of atoms. Geometric parameters were optimized at the GGA level using the Becke formula for exchange interaction [7] and the Lee–Yang–Parr (LYP) formula for electronic correlation [8].

The α-quartz structure was used as the initial structure in forming the SiO_2 clusters; the β-Si_3N_4 structure was used in the case of silicon nitride. In all the clusters that were calculated, the dangling bonds of the boundary atoms were saturated with hydrogen atoms. For the clusters simulating the defects, the relaxation of atoms in the vicinity of the defect was taken into account for all charge states.

1063-7826/01/3509-0997$21.00

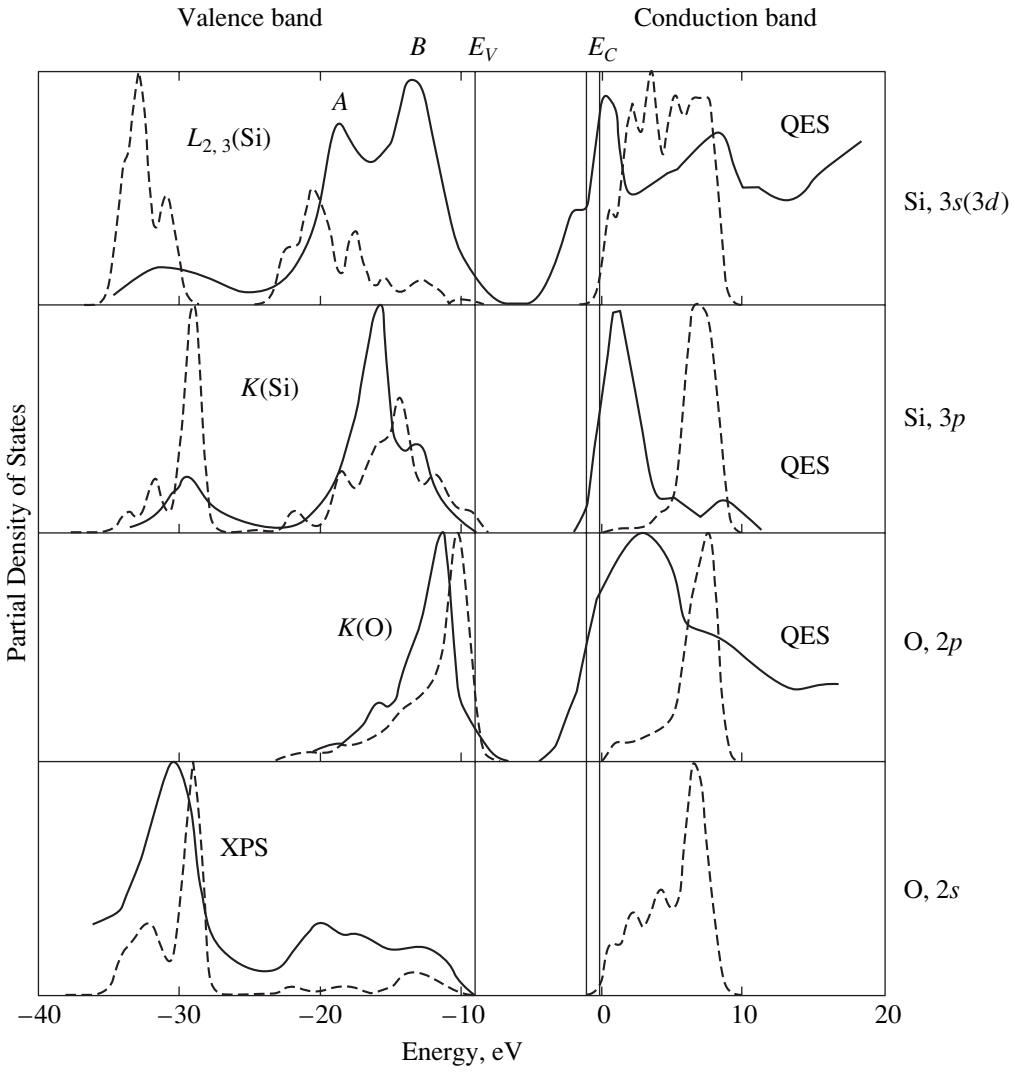


Fig. 1. Experimental X-ray emission spectra (solid lines) and the partial densities of states (dashed lines) calculated for atoms in the vicinity of the center of a $Si_{47}O_{52}H_{84}$ cluster that simulates the SiO_2 bulk.

The energy gain resulting from the capture of a hole or electron by the defect was calculated using the formula [9]

$$\Delta E^{h,e} = (E^0_{\text{vol}} + E^{\pm}_{\text{def}}) - (E^{\pm}_{\text{vol}} + E^0_{\text{def}}), \qquad (1)$$

where $E^{0,\pm}_{\text{vol}}$ and $E^{0,\pm}_{\text{def}}$ are the total energies of clusters that simulate the volume or defect in different charge states $(0, \pm 1)$.

Previously [10–12], the energy gain ("electrical level") for certain defects in SiO_2 was calculated by almost the same method, but the experimental values of the band edges (rather than their theoretical estimates) were used. However, in this case, it is necessary to take into account the polarization energy of the external medium in calculating the charged clusters, which was not carried out in [10–12]. If formula (1) is used, the polarization energy of the external medium for two charged clusters is reduced under the condition that clusters close in size are used for simulating the bulk and the defect. The difference is appreciable. The calculation of electron capture by the ≡SiO* defect in SiO_2 using the MINDO/3 method may serve as an example. The energy gain of $\Delta E = 1.44$ eV was reported in [11], whereas the calculation with formula (1) yields 2.3 eV.

2. ELECTRONIC STRUCTURE OF THE SiO_2 BULK

The electronic structure of silicon oxide has been considered by us previously [4]. In Fig. 1, the dashed lines represent the partial densities of one-electron states (PDOESs) calculated by the MINDO/3 method for a $Si_{47}O_{52}H_{84}$ cluster, which contains 183 atoms. All the PDOESs were calculated for the atoms located in the vicinity of the cluster center. The calculated PDOESs and the experimental spectra were normalized

to the maximum value (separately for the valence and conduction bands).

The solid lines in Fig. 1 represent experimental a-SiO_2 X-ray emission spectra (XES) $L_{2,3}$ (Si), K (Si), and K (O); the X-ray photoelectron spectrum (XPS, the lowest spectrum); and also the quantum-efficiency spectra (QES). These spectra have been reported previously [4]. The energy origin was taken equal to the electron energy in free space. The experimental positions of the edges of the E_C and E_V bands are indicated by vertical lines in Fig. 1.

It is typical that only the one-electron transitions are considered in interpreting the X-ray emission spectra. In this case, the intensities of the transitions allowed in the dipole approximation are proportional to the density of electronic states in the valence band and to the transition probability. If the transition matrix element depends only slightly on the energy, the XES approximately represent the PDOES in the valence band.

The plots represented in Fig. 1 show that the silicon dioxide valence band consists of two subbands separated by a gap. The lower narrow subband includes mainly the 2s oxygen states with a small addition of the 3s and 3p silicon states. The upper subband incorporates the 2p oxygen orbitals and the 3s and 3p silicon states. The valence-band top is mainly formed of $2p_\pi$ oxygen orbitals.

The $L_{2,3}$ (Si) XES represents the transitions from the Si 3s and 3d states to the 2p Si level. It is notable that the peak B in the vicinity of the valence-band top is not present in the calculated PDOES for the 3s Si state (see Fig. 1), although this peak is clearly pronounced in the experimental spectrum. This is characteristic of all calculations that use only the 3s- and 3p-silicon basis functions [4]. A similar situation occurs in silicon nitride [5].

It has been shown in a number of studies (see, for example, [13, 14]) that the 3d silicon state can make an appreciable contribution to the B peak. However, from a general standpoint, the $L_{2,3}$ (Si) spectrum is caused by transitions of electrons to the 2p Si state from the bulk-related extended states rather than from purely atomic states [5]. The transitions from the Si 3d and 4s states contribute to peak B, as do non-single-center transitions of electrons to the 2p Si states from the atomic 2p states of the nearest oxygen atoms.

We calculated the X-ray emission spectra using the DFT method and taking into account both the one- and two-center transitions. The emission spectra corresponding to one-center transitions have been calculated previously [13] using the self-consistent pseudopotential method.

In order to simulate the SiO_2 bulk by the DFT method, we used a 33-atom $Si_5O_{16}H_{12}$ cluster that was centered around the silicon atom and included three regular coordination spheres of the oxide.

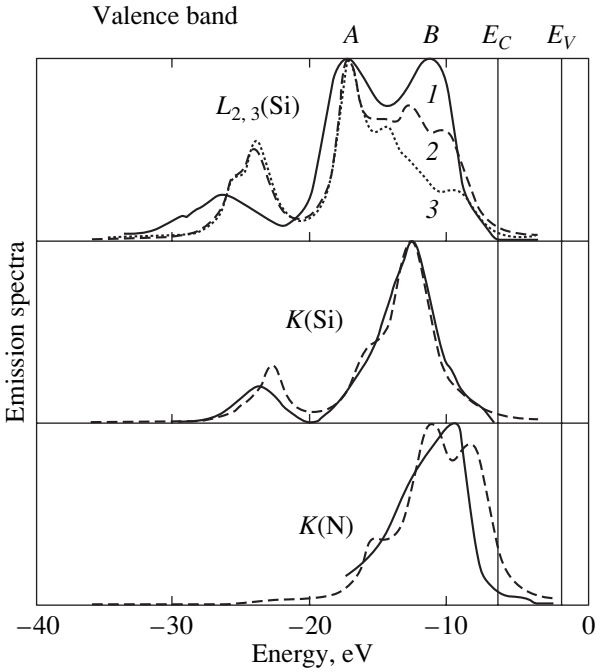


Fig. 2. The X-ray emission spectra for silicon oxide. Experimental spectra are represented by solid lines. The spectra calculated by the DFT method for a $Si_5O_{16}H_{12}$ cluster are shown by the dashed lines if the polarized 3d Si functions were used in the calculations and by the dotted line if these functions were not used.

The results of these calculations are compared with experimental spectra in Fig. 2. All the curves were separately normalized to a unit peak height; theoretical curves were shifted along the energy scale in order to ensure the coincidence of the calculated and experimental peaks (for the $L_{2,3}$ (Si) spectrum, the coincidence with peak A was ensured). The transition intensities were calculated in the frozen-orbital approximation: both the initial state (a hole at the inner orbital) and the final state (a hole in the valence band) were composed of molecular orbitals obtained as a result of the same calculation of the ground neutral state. The transition energies were assumed to be equal to the difference between the corresponding one-electron energies (thus, Koopmans' theorem was used). The matrix elements of transitions were calculated in the dipole approximation.

The theoretical curves represented in Fig. 2 were calculated with the introduction of the polarized Si 3d functions, except for curve *3* in the $L_{2,3}$ (Si) spectrum, which was calculated without introducing the above functions. Thus, the calculations indicate that the polarized 3d functions of Si contribute appreciably to peak B.

The experiment gives no way of separating the contributions of antibonding and bonding 2p oxygen states. However, it is important that the 3p Si states and, consequently, the related 2p O states are observed up to the

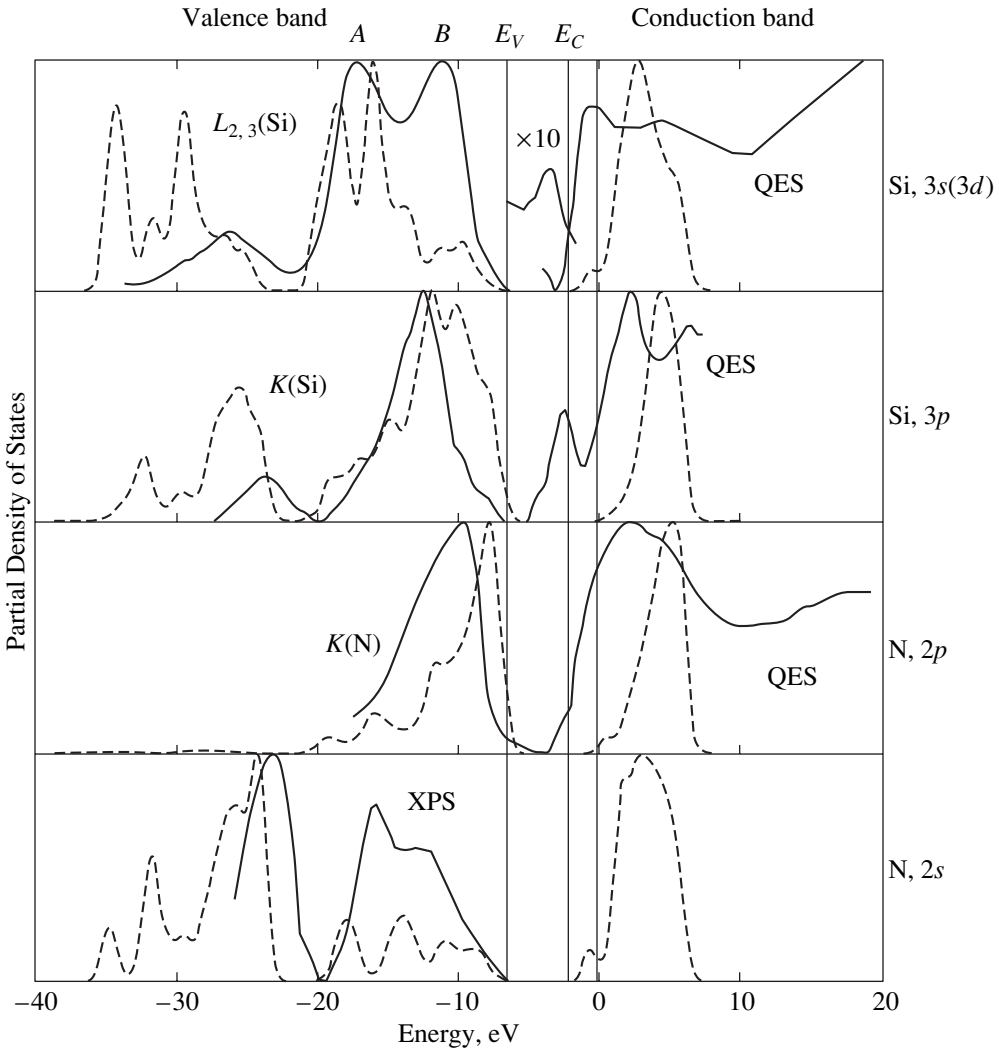


Fig. 3. Experimental X-ray emission spectra (the solid lines) and the partial densities of states (the dashed lines) calculated for the atoms near the center of a $Si_{61}N_{74}H_{76}$ cluster that simulates the Si_3N_4 bulk.

valence-band top. This means that not only the antibonding $2p_\pi$ O states but also the weakly bonding $2p$(O)–$3s,3p,3d$(Si) states exist in the upper part of the valence band. It has been argued [4] that this corresponds to the presence of the light and heavy holes in SiO_2. This assertion is based on the assumption that the rapid transport of holes in SiO_2 is accomplished via transport over the Si–O–Si bonds rather than via hops between the antibonding $2p_\pi$ orbitals of oxygen atoms.

3. ELECTRONIC STRUCTURE OF THE Si_3N_4 BULK

Electronic structures of the Si_3N_4 bulk have been considered previously [5]. The results of calculating the energy-band structure for the α and β phases of silicon nitride were reported, for example, by Xu *et al.* [14].

Figure 3 shows the PDOESs calculated for a $Si_{61}N_{74}H_{76}$ 211-atom cluster. Both the calculations and experimental data indicate that the Si_3N_4 valence band consists of two subbands. The lower narrow band is mainly formed by the $2s$ nitrogen orbitals, whereas the upper wide band is formed by the $2p$ nitrogen orbitals, which overlap the $3s$ and $3p$ silicon orbitals. The valence-band top is mainly formed by the nitrogen $2p_\pi$ orbitals.

Peak *B* in the upper part of the valence band in the $L_{2,3}$ (Si) spectrum (Fig. 3) has the same origin [5] as that of the peak *B* for SiO_2 considered above. This is supported by the DFT calculations of the Si_3N_4 emission spectra shown in Fig. 4.

In the vicinity of the Si_3N_4 valence-band top, a narrow band of antibonding $2p_\pi$ nitrogen orbitals is found. In addition, there is a nonzero density of states formed by the bonding $3s,3p,3d$(Si)–$2p,2s$(N) orbitals; holes with higher mobility may correspond to this density of states. The energy-band calculations yield a high

anisotropy of the effective mass of holes, both for SiO_2 and Si_3N_4 [14]. An appreciable component of the hole effective mass is attributed to unshared pairs localized at the $2p_\pi$ (N) or $2p_\pi$ (O) orbitals.

4. ELECTRONIC STRUCTURE OF TRAPS IN SiO_2

At present, the models of defects responsible for the localization of electrons and holes in SiO_2 continue to be discussed intensively [15–18].

4.1. The Three-Coordinated Silicon Atom O_3≡Si (the E' Center)*

The main paramagnetic defect in SiO_2 is an oxygen vacancy that has captured a hole [1, 10, 17–20]. The atoms closest to the vacancy experience an asymmetric relaxation, and an unpaired electron is mainly localized at one of two three-coordinated silicon atoms. It has been assumed that the O_3≡Si* group in this defect gives rise to transitions in the absorption spectra in the vicinity of 5.8 eV [17, 20, 21]. Several types of paramagnetic *E*' centers are recognized in relation to different atomic surroundings of the O_3≡Si* group [17, 20, 21].

The basic clusters used to simulate the defects in SiO_2 are represented in Fig. 5. In order to simulate an isolated ≡Si* defect, we used a $Si_4O_{12}H_9$ 25-atom cluster that was centered at a three-coordinated silicon atom and included three coordination shells of the oxide atoms. The defect is paramagnetic in its neutral state. Calculations by the MINDO/3 method have shown that 47% of the spin density of an unpaired electron is localized at the three-coordinated silicon atom and 25% is localized at the three nearest oxygen atoms.

The neutral defect introduces a one-electron level into the band gap; this level is located at 3.5 eV above the "cluster" valence-band top (the last but one occupied level). In publications, Koopmans' theorem has been often used to estimate the energies of capture of an electron or hole by a defect. However, Koopmans' theorem may give rise to large errors for highly localized states. Thus, our use of a more consistent (Δ*SCF*) method expressed by formula (1) yielded an energy gain of 2.9 eV for electron capture by the ≡Si* defect, which is lower than the value estimated above using Koopmans' theorem by 0.6 eV. The energy gain as a result of the capture of an electron by the isolated ≡Si* defect was equal to 1.4 eV.

4.2. Two-Coordinated Silicon Atom ≡Si: (a Sililene Center)

This defect in SiO_2 is often related to the absorption band peaked at 5.0 eV and to luminescence bands peaked at 2.7 and 4.4 eV [18, 22–24].

In order to simulate this defect, we used a $Si_3O_8H_6$ 17-atom cluster, which was obtained from a $Si_5O_{16}H_{12}$ cluster that simulated the oxide bulk by breaking two Si–O bonds of the central silicon atom and by removing the two corresponding groups of atoms.

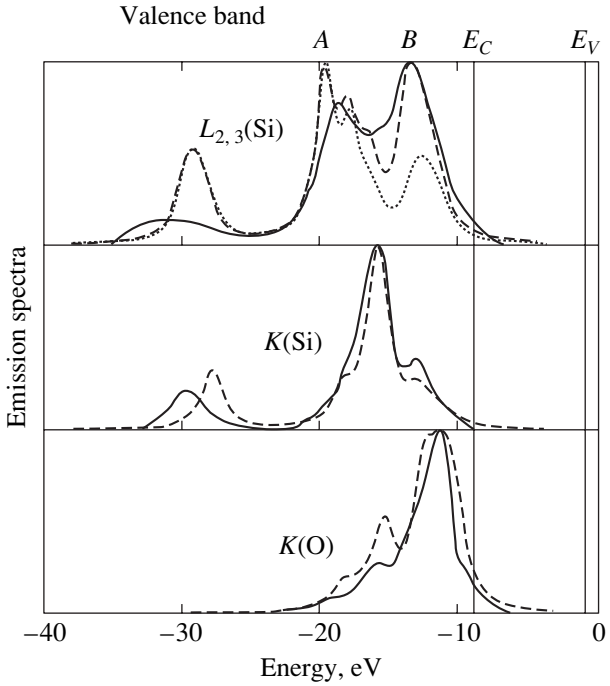


Fig. 4. X-ray emission spectra for silicon nitride. Experimental spectra are shown by solid lines. The spectra calculated by the DFT method for a $Si_{19}N_{26}H_{36}$ cluster are shown by dashed lines if the polarized 3*d* Si functions were used in the calculations and by the dotted line if these functions were not used.

Calculations show that the defect is diamagnetic in the neutral state. Two electrons of a two-coordinated silicon atom form an unshared pair that resides at the hybrid orbital lying in the O–Si–O plane. In calculations employing the MINDO/3 method, this hybrid orbital consists of 61% of the *s* atomic orbitals and 39% of *p* atomic orbitals of the silicon atom. Hole capture at this orbital transforms the defect into a paramagnetic state. As a result, 60% of the spin density becomes localized at the two-coordinated silicon atom. Calculations of the positively charged state performed by Dianov *et al.* [12] by the MNDO method indicated that 70% of the spin density is localized at the two-coordinated silicon atom.

The calculations employing the MINDO/3 method show that this defect is a trap for holes with a capture energy of about 1.5 eV. The calculation employing the DFT method shows that this defect is a hole trap with ΔE = 3.2 eV. The calculation employing the MNDO method yields a hole-capture energy of ΔE = 3.9 eV. No electrons are captured by this defect.

The ability of the =Si: defect to capture a hole indicates that this defect, in addition to the oxygen vacancy,

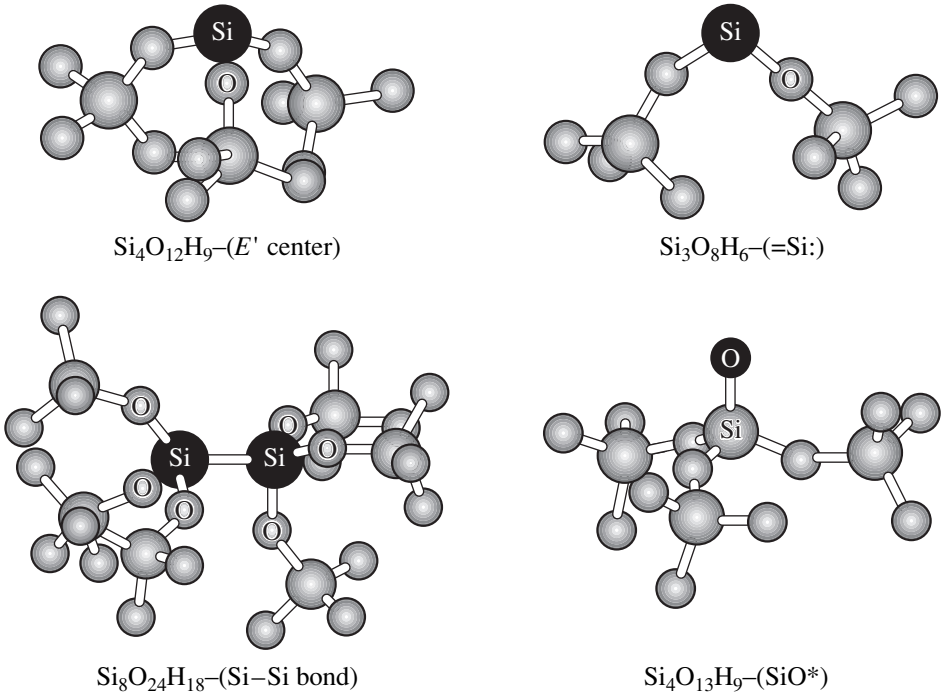


Fig. 5. Clusters used to simulate the defects in silicon oxide. Hydrogen atoms are not shown. The atoms forming defects are represented by shaded circles.

may be responsible for the accumulation of positive charge in the MOS devices subjected to radiation.

4.3. Silicon–Silicon Bond ≡Si–Si≡, Oxygen Vacancy

Experiments and theoretical calculations have shown that a neutral oxygen vacancy can capture a hole and, thus, give rise to the *E*' paramagnetic center [1, 10, 11, 18, 20]. The absorption band peak observed at 7.6 eV is typically related to the oxygen vacancy [17, 18]. It is assumed that the transition occurs between the bonding and antibonding states of the Si–Si bond formed.

As an initial cluster for forming the Si–Si bond, we chose a Si_8O24H_{18} 50-atom cluster that simulated an oxygen vacancy in the dioxide. The remote oxygen atom formed an angle of 144° with two neighboring silicon atoms. The $Si_8O_{24}H_{18}$ cluster consisted of two 25-atom halves, the distance between which was varied along the Si–Si bond. By choosing the distance L between these two halves, we then optimized the two silicon atoms in the Si–Si bond, with the positions of all other atoms remaining unchanged. The oxygen vacancy in dioxide corresponds to the initial distance of L = 3.1 Å. Calculations using the MINDO/3 method showed that, for L = 2.35 Å, this defect is an electron trap with ΔE = 1.0 eV. A hole can be also captured with ΔE = 3.0 eV. As the distance L increases, the gain in energy as a result of capturing an electron increases to 1.4 eV, which corresponds to an isolated ≡Si* defect. The gain in energy as a result of capturing a hole is virtually independent of the distance L.

The structure of neutral and positively charged oxygen vacancies has been theoretically studied in [10, 12, 18]. Snyder *et al.* [10] used several semiempirical methods: MINDO/3, MNDO, AM1, and PM3. A comparison of the results obtained by these methods shows that the MINDO/3 method is preferable [10]. The calculations yielded two local energy minima in the relaxation of the positively charged vacancy. One of the minima corresponds to an almost symmetric relaxation of atoms with a nearly symmetric electronic structure. The other minimum is deeper than the first one by 0.17 eV (MINDO/3). The transition to this state corresponds to an increase in the energy gain to 3.2 eV, which occurs when a hole is captured by an oxygen vacancy. This second minimum corresponds to the transfer of a silicon atom through the plane formed by three oxygen atoms bonded to the above silicon atom, with an accompanying formation of a weak bond with another (more remote) oxygen atom. The barrier height for such a transition is 0.4 eV (MINDO/3). An unpaired electron is localized at another three-coordinated silicon atom. Experiments with electron spin resonance (ESR) have shown that an unpaired electron is almost completely localized at a single silicon atom [20].

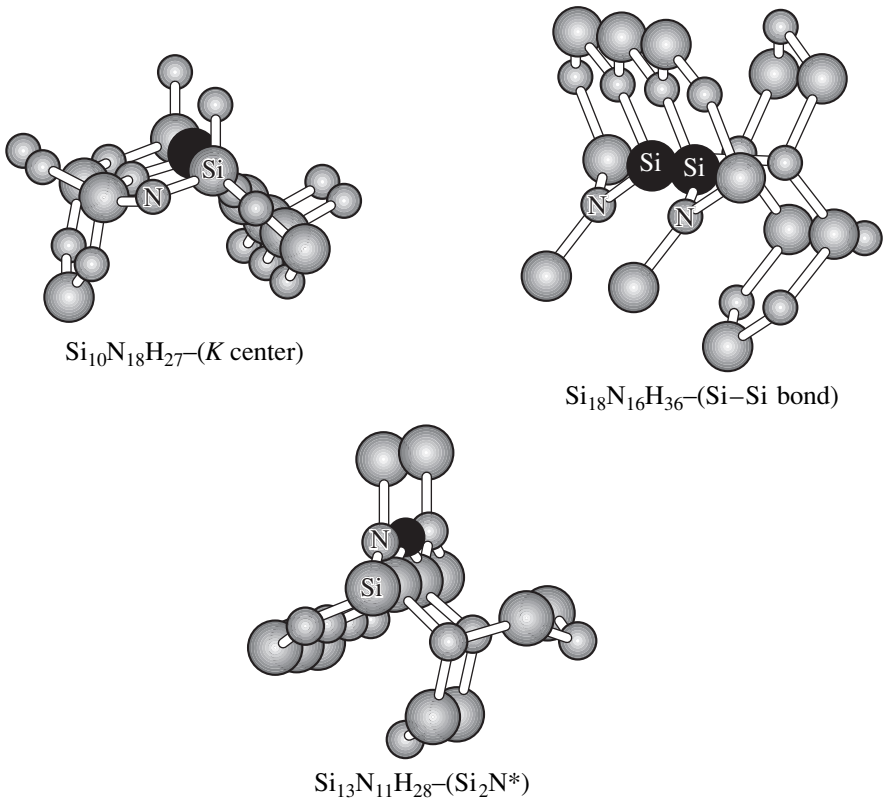


Fig. 6. The clusters used in simulating the defects in silicon nitride.

4.4. One-Coordinated Oxygen Atom ≡SiO (a "Nonbridging" Oxygen Atom)*

In order to reduce the amount of positive charge trapped in SiO_2, an empirical method has been developed; this method consists in the compensation of the above charge by a positive charge captured by electron traps, which are formed as a result of the wet oxidation of silicon [25]. However, the nature of such electron traps has not been clarified yet. It is believed that this electron trap may be attributed to a center that corresponds to the red luminescence line peaked at 1.9 eV and is related to the ≡SiO* defect [26]. This identification is supported by quantum-chemical calculations [17, 18]. It is assumed that this defect is responsible for the radiation resistance of oxide produced by wet oxidation.

In order to simulate a one-coordinated oxygen atom ≡SiO*, we used a $Si_4O_{13}H_9$ 26-atom cluster obtained from a "bulk" $Si_5O_{16}H_{12}$ cluster by breaking the outer O–Si bond of an oxygen atom in the first coordination sphere and by removing the corresponding atomic group.

The calculations using the MINDO/3 method [27] showed that this defect is only a trap for electrons with an energy gain of 2.3 as a result of electron capture. The defect is paramagnetic in its neutral state. An unpaired electron is localized at the $2p_\pi$ orbital of a one-coordinated oxygen atom. The calculation using the DFT method showed that the ≡SiO* defect is an electron trap with ΔE = 3.9 eV. The calculation by the MNDO method yields the value of the capture energy ΔE equal to 3.2 eV.

Thus, the three simulation methods that we considered verify the ability of the ≡SiO* defect to capture an electron; therefore, this defect may be used to interpret the electron traps with energies of 2.4–2.5 eV, which are observed in experiments with oxide depolarization [28].

5. ELECTRONIC STRUCTURE OF TRAPS IN Si_3N_4

Amorphous silicon nitride and oxynitride are considered to be alternatives to silicon oxide in silicon devices for the near future [29]; *a*-Si_3N_4 has a high concentration ($\sim 5 \times 10^{18}$ cm^{-3}) of electron and hole traps. It is believed that the capture of electrons and holes in *a*-Si_3N_4 is related to the defects induced by excess silicon atoms [30, 31]. The ESR signal was not detected either in the initial *a*-Si_3N_4 samples or in the samples subjected to the injection of electrons and holes [30, 31].

5.1. Three-Coordinated Silicon Atom N_3≡Si (the K Center)*

In order to simulate this defect, we used a $Si_{10}N_{18}H_{27}$ 55-atom cluster that is centered at the three-coordinated silicon atom and includes the atoms of three coordination shells of silicon nitride. The main clusters used in simulating the defects in Si_3N_4 are represented in Fig. 6.

Calculations by the MINDO/3 method show that an isolated ≡Si* defect captures an electron with an energy gain of 1 eV. The energy gain as a result of capturing a hole amounts to a mere 0.1 eV, which is within the model's error. The ≡Si* defect, which is paramagnetic in the neutral state, becomes diamagnetic after capturing an electron (a hole).

The absence of an ESR signal in the initial *a*-Si_3N_4 samples indicates that there are in fact no neutral isolated ≡Si* defects in these samples. In order to explain the absence of an ESR signal, it has been assumed that a pair of neutral defects ≡Si* gives rise to a pair of charged defects: ≡Si: and ≡Si* (the K^- and K^+ centers) as a result of the following reaction:

$$\equiv Si^* + \equiv Si^* \longrightarrow \equiv Si: + \equiv Si. \quad (2)$$

This model of "negative correlation energy" [30, 32, 33] implies that the energy gain (*–U*) is attained owing to the lattice relaxation.

We used the MINDO/3 method to calculate the characteristics of this defect in various charge states with allowance made for the relaxation of only the three-coordinated silicon atom. These calculations show that reaction (2) corresponds to the positive correlation energy equal to 4.0 eV. Nonempirical calculations at a level of 6-3G*/MP2 for a $Si(NH_2)_3$ 10-atom cluster yield a correlation energy of ~5.5 eV. Taking into account the polarization energy of the external medium according to the "classical model" [34], we obtain the following estimates for the correlation energy *U*: +2.7 eV (MINDO/3) and +3.5 eV (*ab initio*).

Recent DFT-based calculations for reaction (2) [34] yielded a positive correlation energy of $U = +0.9$ eV for isolated ≡Si* defects in silicon nitride. This comparatively small value is accounted for by a pronounced relaxation effect observed [34] for the positively charged ≡Si defect.

Thus, the calculations do not support the widely accepted model of negative correlation energy for the *K* centers in *a*-Si_3N_4.

5.2. Silicon–Silicon Bond ≡Si–Si≡, the Nitrogen Vacancy

The ≡Si–Si≡ defects can be most easily formed in the silicon nitride bulk by introducing a nitrogen vacancy. The latter was simulated using a $Si_{18}N_{16}H_{36}$ 70-atom cluster centered at the nitrogen atom to be removed.

As follows from calculations using the MINDO/3 method, a nitrogen vacancy serves as a trap for electrons ($\Delta E = 1.6$ eV) and holes ($\Delta E = 1.0$ eV). In its neutral state, the defect is paramagnetic. When a hydrogen atom is trapped by this defect, a dangling bond of one of the silicon atoms becomes saturated. The capture of an electron by such a center is accompanied by the dissociation of the Si–H bond and by the escape of the hydrogen atom to the nitride bulk. As a result, the total energy gain amounts to about 1 eV.

If a three-coordinated silicon atom (together with other atoms bonded to it) is removed from a $Si_{18}N_{16}H_{36}$ cluster, we obtain a $Si_{14}N_{12}H_{30}$ 56-atom cluster, which may be considered as a model of the Si–Si bond with an initial 2.9-Å distance between silicon atoms. Such a defect is an electron trap with $\Delta E = 1.76$ eV. The capture of a hole is unlikely because a small electron-capture energy of $\Delta E = 0.34$ eV may be related to errors in the model.

*5.3. Two-Coordinated Nitrogen Atom $(\equiv Si)_2N$**

Experimental data reported by Yount and Lenahan [35] indicate that the $(\equiv Si)_2N^*$ is an electron trap in silicon oxynitride. In order to eliminate the electron traps in the gate oxynitride of silicon devices, the oxynitride is repeatedly oxidized. Assuming in this situation that the electron traps are $(\equiv Si)_2N^*$ defects, we may easily explain the elimination of the traps in this technological stage by the replacement of a two-coordinated nitrogen atom by a regularly coordinated oxygen atom [27].

The $(\equiv Si)_2N^*$ defect was simulated using a $Si_{13}N_{11}H_{28}$ cluster that incorporated the coordination shells of nitride atoms. Calculations by the MINDO/3 method indicate that this defect is an electron trap with $\Delta E = 0.8$ eV and acts as a hole trap only in silicon oxynitride enriched with oxygen [27]. The ability of the $(\equiv Si)_2N^*$ defect to trap an electron is corroborated by the recent DFT-based calculations [34]. In its neutral state, this defect is paramagnetic. The unpaired electron is localized at the $2p_\pi$ state of the nitrogen atom, which is consistent with the ESR data [20, 35].

6. CONCLUSION

In order to identify the traps in SiO_2 and Si_3N_4, we used quantum-chemical methods to calculate the energy gain that results from the major intrinsic defects in these materials trapping an electron or a hole. The calculations were performed in the cluster approximation with allowance made for electronic and atomic relaxation. It is shown that the ≡Si* defects, oxygen vacancies, and the ≡SiO* centers are electron traps in SiO_2, whereas the ≡Si*, nitrogen vacancies, and the $(\equiv Si)_2N^*$ centers act as electron traps in silicon nitride. The hole traps are ≡Si*, oxygen vacancies, and the (=Si:) defect in SiO_2; in Si_3N_4, the nitrogen vacancy acts as a hole trap.

ACKNOWLEDGMENTS

This study was supported by INTAS, grant no. 97-0347.

REFERENCES

1. V. A. Gritsenko, *Structure and Electronic Properties of Amorphous Insulators in Silicon MIS-Devices* (Nauka, Novosibirsk, 1993).
2. *Silicon Nitride in Electronics*, Ed. by A. V. Rzhanov (Nauka, Novosibirsk, 1982; Elsevier, New York, 1988).
3. P. M. Lenahan and P. V. Dressendorfer, J. Appl. Phys. **55**, 3495 (1984).
4. V. A. Gritsenko, R. M. Ivanov, and Yu. N. Morokov, Zh. Éksp. Teor. Fiz. **108**, 2216 (1995) [JETP **81**, 1208 (1995)].
5. V. A. Gritsenko, Yu. N. Morokov, and Yu. N. Novikov, Fiz. Tverd. Tela (St. Petersburg) **39**, 1342 (1997) [Phys. Solid State **39**, 1191 (1997)].
6. C. Fonseca Guerra, J. G. Snijders, G. te Velde, and E. J. Baerends, Theor. Chem. Acc. **99**, 391 (1998).
7. A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
8. C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
9. A. X. Chu and W. B. Fowler, Phys. Rev. B **41**, 5061 (1990).
10. K. C. Snyder and W. B. Fowler, Phys. Rev. B **48**, 13238 (1993).
11. A. H. Edwards and W. B. Fowler, in *Structure and Bonding in Noncrystalline Solids*, Ed. by G. E. Walrafen and A. G. Revesz (Plenum, New York, 1986), p. 139.
12. E. M. Dianov, V. O. Sokolov, and V. B. Sulimov, J. Non-Cryst. Solids **149**, 5 (1992).
13. A. Simunek, J. Vackar, and G. Wiech, J. Phys.: Condens. Matter **5**, 867 (1993).
14. Y.-N. Xu and W. Y. Ching, Phys. Rev. B **51**, 17379 (1995).
15. G. Pacchioni and M. Vitiello, J. Non-Cryst. Solids **245**, 175 (1999).
16. V. A. Gritsenko, J. B. Xu, R. W. M. Kwok, *et al.*, Phys. Rev. Lett. **81**, 1054 (1998).
17. L. Skuja, J. Non-Cryst. Solids **239**, 16 (1998).
18. G. Pacchioni and G. Ierano, Phys. Rev. B **57**, 818 (1998).
19. A. R. Silin' and A. N. Trukhin, *Point Defects and Elementary Excitations in Crystalline and Vitreous SiO_2* (Zinatne, Riga, 1985).
20. W. L. Warren, E. H. Poindexter, M. Offenberg, and W. Muller-Warmuth, J. Electrochem. Soc. **139**, 872 (1992).
21. D. L. Griscom, J. Non-Cryst. Solids **73**, 51 (1985).
22. V. A. Radzig, J. Non-Cryst. Solids **239**, 49 (1998).
23. G. Pacchioni and R. Ferrario, Phys. Rev. B **58**, 6090 (1998).
24. V. A. Gritsenko, Yu. G. Shavalgin, P. A. Pundur, *et al.*, Microelectron. Reliab. **39**, 715 (1999).
25. Z. Shanfield and M. M. Moriwaki, IEEE Trans. Nucl. Sci. **NS-31**, 1242 (1984).
26. L. Skuja, J. Non-Cryst. Solids **179**, 51 (1994).
27. Yu. N. Morokov, Yu. N. Novikov, V. A. Gritsenko, and H. Wong, Microelectron. Eng. **48**, 175 (1999).
28. V. J. Kapoor, F. J. Feigl, and S. R. Butler, J. Appl. Phys. **48**, 739 (1977).
29. Y. Shi, X. Wang, and T.-P. Ma, IEEE Trans. Electron Devices **46**, 362 (1999).
30. W. L. Warren, J. Kanichi, J. Robertson, *et al.*, J. Appl. Phys. **74**, 4034 (1993).
31. V. A. Gritsenko and A. D. Milov, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 489 (1996) [JETP Lett. **64**, 531 (1996)].
32. P. W. Anderson, Phys. Rev. Lett. **34**, 953 (1975).
33. R. A. Street and N. F. Mott, Phys. Rev. Lett. **35**, 1293 (1975).
34. G. Pacchioni and D. Erbetta, Phys. Rev. B **61**, 15005 (2000).
35. J. T. Yount and P. M. Lenahan, J. Non-Cryst. Solids **164-166**, 1069 (1993).

Translated by A. Spitsyn