

14. Zhang, Y., Sakai, T., Osuki, H., Yamamoto, T., Kokubu, A. (2011). Very High Cycle Fatigue Characteristics of Zr-Base Bulk Amorphous Alloy in Rotating Bending. *Journal of Solid Mechanics and Materials Engineering*, 5 (10), 519–533. doi: <https://doi.org/10.1299/jmmp.5.519>
15. *Standard Methods for Mechanical Testing of Welds* (2007). Miami: American Welding Society, 2007.
16. Kuhn, H., Medlin, D. (Eds.) (2000). *ASM Handbook Volume 8: Mechanical Testing and Evaluation*. Ohio: ASM International, 998.
17. Rafi, H. K., Ram, G. D. J., Phanikumar, G., Rao, K. P. (2010). Microstructure and tensile properties of friction welded aluminum alloy AA7075-T6. *Materials & Design* (1980-2015), 31 (5), 2375–2380. doi: <https://doi.org/10.1016/j.matdes.2009.11.065>

Застосування сучасних прикладних комп'ютерних програм розширює можливість проведення багатокомпонентного статистичного аналізу в матеріалознавстві. В роботі розглянуто процедуру застосування методу множинного кореляційно-регресійного аналізу для дослідження і моделювання багатофакторних зв'язків фізичних характеристик у кристалічних структурах. Розгляд здійснено на прикладі монокристалів нелегованого арсеніду галію. У виконаному статистичному аналізі був задіяний комплекс із семи фізичних характеристик, отриманих неруйнівними методами для кожної з 32 точок вздовж діаметра кристалічної пластини. Масив даних досліджувався методами множинного кореляційного аналізу. Була побудована розрахункова модель регресійного аналізу. На її основі з використанням програм Excel, STADIA і SPSS Statistics 17.0 проведено статистичну обробку даних і аналітичне вивчення взаємозв'язків всіх характеристик. Отримано і проаналізовано регресійні співвідношення при визначенні концентрації фонові домішки вуглецю, залишкових механічних напружень і концентрації фонові домішки кремнію. Була встановлена можливість коректного проведення множинного статистичного аналізу для моделювання властивостей кристала GaAs.

Виявлено нові взаємозв'язки між параметрами кристала GaAs. Встановлено, що концентрація фонові домішки кремнію пов'язана з вакансійним складом кристала і значенням концентрації центів EL2. Також встановлено відсутність зв'язку концентрації кремнію з величиною залишкових механічних напружень. Ці факти і термічні умови формування точкових дефектів при вирощуванні монокристалів свідчать про відсутність перерозподілу фонових домішок в процесі охолодження кристала нелегованого GaAs.

Використання методу множинного регресійного аналізу в матеріалознавстві дозволяє не тільки моделювати багатофакторні зв'язки в бінарних кристалах, а й здійснювати стохастичне моделювання факторних систем змінного складу

Ключові слова: кореляційно-регресійний аналіз, множинна регресія, арсенід галію, кристалічна структура

UDC 631:331.4+ 621.315

DOI: 10.15587/1729-4061.2019.188512

APPLICATION OF MULTIPLE CORRELATION ANALYSIS METHOD TO MODELING THE PHYSICAL PROPERTIES OF CRYSTALS (ON THE EXAMPLE OF GALLIUM ARSENIDE)

M. Litvinova

Doctor of Pedagogical Sciences,
PhD (Physics and Mathematics), Associate Professor*
E-mail: lmb965@gmail.com

N. Andrieieva

PhD, Associate Professor
Department of Heat Engineering**

V. Zavodyannyi

PhD, Associate Professor
Department of Physics and General Engineering Disciplines
Kherson State Agrarian University
Stritenska str., 23, Kherson, Ukraine, 73006

S. Loi

Associate Professor
Department of Welding**

O. Shtanko

PhD, Associate Professor*

*Department of Software Engineering,
Physics and Mathematics**

**Kherson Branch of Admiral Makarov National
University of Shipbuilding
Ushakova ave., 44, Kherson, Ukraine, 73022

Received date 25.11.2019

Accepted date 19.12.2019

Published date 26.12.2019

Copyright © 2019, M. Litvinova, N. Andrieieva, V. Zavodyannyi, S. Loi, O. Shtanko

This is an open access article under the CC BY license

<http://creativecommons.org/licenses/by/4.0>

1. Introduction

There is no doubt that in modern materials science modeling of the structural properties of crystals occupies

a fundamentally important place in the process of improving the technologies for their preparation and application. A significant problem in the correct analysis of the structure and corresponding physical parameters of crystals is the

multifactorial nature of their mutual influence [1, 2]. A standard approach to studying the relationship between the structure of crystals and their physical and chemical properties is based on establishing a correlation in two parameters with a fixed value of the others: their absolute values are taken into account, and not combinations and changes of these values. More perfect is the multiple correlation method, which is highly effective in creating accurate macromodels of physical systems. Although this method is used in materials science, the field of its application can be significantly expanded taking into account the capabilities of modern applied computer programs [3, 4].

The greatest problems with the multifactorial nature of physical property bonds arise in the study of non-stoichiometric crystals [5]. The above is typical for structural modeling of $A^{III}B^V$ compounds and, in particular, gallium arsenide. Therefore, multicomponent modeling of the properties of such crystals undoubtedly requires the implementation of multiple correlation analysis and the effectiveness of its application can be demonstrated on this material.

Gallium arsenide (GaAs) is highly effective for single-junction solar cells, but high production costs (under the conditions for obtaining structures with specified parameters) limit its use. The use of multiple correlation analysis to simulate the physical properties of GaAs crystals is a prerequisite for better predictability of their technological parameters and lower production costs. The calculation model built for gallium arsenide crystals can be used to model the physical properties of other materials.

2. Literature review and problem statement

Currently, a large amount of factual material has been accumulated on the physical properties of gallium arsenide. However, the volatility of arsenic atoms during the growth of GaAs single crystals leads to a significant variability in the deviation of their structure from stoichiometry and the complexity of the stable reproduction of parameters. Studies [6, 7] show that, as in any coupled multifactor system, in GaAs single crystals, a change in the concentration of one of the structural components can ambiguously affect the others. In particular, it is shown in [6] that fluctuations in the concentration of background carbon and silicon impurities in undoped GaAs lead to a local change in the band gap (the formation of local density state tails) in the crystal bulk. In optical manifestations, this effect is similar to the effect of localization of states considered in [7] in the bulk and surface regions of submicron grains of a crystal. Despite the variability of the deviation of the crystal structure from stoichiometry, in both studies the calculation model is limited by the relationship of only two parameters: carrier concentration – spectral characteristic of the material.

It is established in [8, 9] that the non-uniform distribution of defects (impurity atoms, complex centers, defect-impurity associations, etc.) in the bulk of crystals leads to the appearance of various macro- and microscopic electric and elastic fields. In [8], it is studied how external electric fields cause the formation of local internal fields and the corresponding redistribution of point defects. The authors of [9] showed in a compensated material that a weak electric field destroys exciton states in those regions of the crystal where the concentration of the background impurity fluctuates. However, the analysis and statistical processing of the exper-

imental results in the considered papers are carried out in a simplified way – based on pair parameter relationships. This reduces the validity of the conclusions and physical models proposed by the authors. In [1], it is stated that, due to the multifactorial nature of the structural bonds of the material, it remains a challenge to obtain large diameter gallium arsenide single crystals with a reliably predictable distribution of physical parameters in their cross section.

At the moment, there is experience in using multiple correlation analysis in X-ray diffraction studies, in medicine, chemistry, geology, etc. For example, the use of Minitab Software and SPSS Software in [4] makes it possible to conduct multivariate modeling of the growth rate of single-component borax single crystals from aqueous solutions. The use of cross-correlation functions in processing the data of X-ray diffraction analysis of solid colloidal crystals makes it possible for the authors of [12] to obtain valuable structural information. Corresponding calculations for single-component crystals and ordered monodisperse colloidal systems are a prerequisite for optimizing the modeling of the properties of binary crystals and, in particular, gallium arsenide single crystals.

3. The aim and objectives of research

The aim of research is to consider the effectiveness of using multiple correlation analysis to study and model the structural, electrophysical, and mechanical properties of crystals with multifactor bonds using single crystals of semi-insulating undoped (SIU) GaAs.

To achieve the aim, the following objectives are set:

- verify the data necessary for the correlation analysis of physical characteristics in a single crystal of SIU GaAs using Excel, STADIA, and SPSS Statistics 17.0 computer programs;
- build a calculation model and calculate the main correlation analysis coupling indicators for the SIU GaAs crystal parameters;
- analyze the results of multiple correlation analysis of GaAs crystals and generalize them to simulate the physical properties of other materials.

4. Materials and methods for studying the physical characteristics of crystals selected for conducting multiple statistical analysis

4.1. The investigated materials and equipment used in the experiment

The material for the statistical study is the set of physical characteristics of the semi-insulating undoped gallium arsenide (SIU GaAs). The crystal is grown by the company Pure Metals (Svitlovodsk, Ukraine) according to the Czochralski method with liquid sealing in the (100) direction from a melt close to stoichiometric. All parameters are measured in the cross section of the crystal along the diameter of the plate, cut perpendicular to the direction of its growth in the middle of the length of the ingot. The diameter of the plate is 6.5 cm.

The studied material at a temperature of 300 K had n -type conductivity, resistivity $r=1.2 \cdot 10^8$ Ohm-cm and mobility of the main carriers $\mu_n=4850$ cm²/V·s. The average dislocation density on the plate is $N_d=8.4 \cdot 10^4$ cm⁻². The dislocation distribution along the plate diameter is W-shaped,

i. e., their maximum density is in the central region and on the periphery of the crystal.

4. 2. Method for determining the physical characteristics of the crystal

For statistical processing, the physical characteristics of the crystal are used, the determination of which is possible on the basis of non-destructive methods at fixed positions along the plate diameter. Based on this, seven physical parameters are studied, which are experimentally determined as follows:

1) The density of dislocations N_d is found from the etching pits using a MIM-7 microscope [5].

2) The intensity of the edge band of photoluminescence (PL) with a maximum energy $h\nu_m=1.51$ eV (I_{FL}) is measured at 77 K by the standard method with normalization per unit (measured in rel. units) [7]. The PL is excited by a He-Ne laser with a radiation intensity of $(2.5-3.0) \cdot 10^{18}$ cm⁻²·s⁻¹ and a focusing diameter of less than 0.5 mm. The PL is recorded in the region up to 800 nm by a PMT-68, and above 800 nm by a cooled FD-9G germanium photodiode (threshold sensitivity 10^{-10} lm). The measurement of the PL spectra, considered further in paragraphs 3–5, is carried out by the same method.

3) The vacancy composition of the crystal (indicator Z is the ratio of the concentration of gallium and arsenic vacancies) is found by the ratio of the intensities of the edge PL band ($T=77$ K) and the band caused by the radiative transition from the conduction band to the acceptor level of carbon in the arsenic sublattice [13].

4) The concentration of uncontrolled background carbon impurities (N_C) is determined from the PL spectra at 77 K [9].

5) The concentration of uncontrolled background silicon impurity (N_{Si}) is found by the same procedure as N_{Si} .

6) The concentration of deep centers EL2 (N_{EL2}) is determined by the optical absorption of light quanta with an energy of $h\nu=1.04$ eV [7] using an MDR-2 monochromator. The excitation source is a quartz halogen compact lamp with a continuous spectrum of 60 watts.

7) The value of residual mechanical stresses σ is determined by the polarization-optical method using a Senarmon compensator [14]. The measurements are carried out at a wavelength $l=1.15$ μm. The source of polarized monochromatic light is the LG-126 quantum generator. The measurement error s in accordance with the parameters of the installation used is $2 \cdot 10^{-2}$ MPa.

All parameters are measured at 32 points from edge to edge of the crystal plate along its diameter. The displacement step of the measurement point is 2 mm (with an error of 0.1 mm), which corresponds to two diameters of the working field of view of the microscope when determining the dislocation density. Such a displacement step provided a condition under which the studied zones of the crystal did not overlap. The data obtained are given in Table 1.

Each column of the Table 1 corresponds to some indicator (in statistical analysis, the considered characteristics are called independent variables).

For statistical calculations and graphical constructions, Excel, STADIA, SPSS Statistics 17.0 are used. Terms and definitions correspond to the SPSS Statistic program [15]. All hypotheses and conclusions are presented using a significance level of 5 %. This means that the probability of error when adopting various statistical hypotheses and generating conclusions based on statistical analysis is 5 %. In some cases, a significance level of 10 % is adopted (in these cases, it is about less obvious manifestations of statistical dependence [15]).

Table 1

The values of the physical parameters along the diameter of the GaAs plate

No. of point	$N_d \cdot 10^{-4}$, cm ⁻²	I_{FL} , rel. unit	$Z \cdot 10^{-2}$, rel. unit.	$N_C \cdot 10^{-16}$, cm ⁻³	$N_{Si} \cdot 10^{-16}$, cm ⁻³	$N_{EL2} \cdot 10^{-16}$, cm ⁻³	σ , MPa
1	21.7	0.12	13.7	1.80	1.11	1.40	3.0
2	18.1	0.19	10.9	1.85	1.16	1.40	5.7
3	8.1	0.26	10.4	1.81	1.21	1.46	6.3
4	5.0	0.42	7.5	1.89	1.45	1.44	7.2
5	3.9	0.54	7.4	1.83	1.41	1.56	9.3
6	3.2	0.68	7.1	1.91	1.46	1.75	12.5
7	3.2	0.79	6.6	1.90	1.49	1.85	18.4
8	3.6	0.81	6.4	1.86	1.34	1.81	14.6
9	3.7	0.88	5.1	1.84	1.27	1.89	25.0
10	3.1	1.0	4.2	1.81	1.35	2.0	25.7
11	2.1	0.95	4.3	1.82	1.42	1.86	20.6
12	4.6	0.81	4.2	1.80	1.44	1.81	12.7
13	5.1	0.72	5.0	1.81	1.47	1.92	4.4
14	8.3	0.69	5.1	1.78	1.64	1.20	3.8
15	8.6	0.45	5.1	1.80	1.87	1.97	3.1
16	9.6	0.11	5.4	1.86	1.74	1.68	5.7
17	9.5	0.27	5.9	1.81	1.62	1.45	9.0
18	7.1	0.34	5.5	1.75	1.58	1.42	20.2
19	5.9	0.39	5.0	1.75	1.53	1.47	23.0
20	6.2	0.56	5.0	1.77	1.60	1.47	22.7
21	4.1	0.63	4.3	1.81	1.48	1.42	16.9
22	3.9	0.92	4.2	1.80	1.45	1.44	11.1
23	1.9	0.87	3.6	1.76	1.42	1.51	8.3
24	4.9	0.87	5.9	1.79	1.47	1.66	7.1
25	5.6	0.72	7.7	1.70	1.47	1.71	6.6
26	5.8	0.66	8.5	1.88	1.42	1.44	5.4
27	6.0	0.52	8.9	1.91	1.44	1.42	4.2
28	8.8	0.48	9.6	1.82	1.32	1.48	3.1
29	10.0	0.34	10.3	1.83	1.29	1.24	2.2
30	20.1	0.20	12.5	1.86	1.37	1.22	1.2
31	22.6	0.11	14.8	1.99	1.22	1.23	0.7
32	23.1	0.09	15.0	1.98	1.16	1.25	0.3

5. Research results of the effectiveness of multiple correlation analysis for modeling the properties of SIU GaAs crystals

5.1. Data verification for correlation analysis

For the correct statistical analysis, the obtained data are previously verified. The graphical method and the testing method according to the Kolmogorov-Smirnov criterion performed an analysis of the distributions of variables. It is found that all the studied parameters are distributed according to the law of normal distribution. Checking the data array for homogeneity by the Levene's test shows that all physical characteristics are homogeneous. The analysis of scattering diagrams makes it possible to establish that the data array belongs to a single population.

When studying multifactorial processes, it is also advisable to preliminarily investigate the degree of connection between individual factors in pairs. If all pairwise connections approach linear on average, then there is every reason to assume that the multiple connection will be linear (the linear regression model is considered in accordance with [16]).

When conducting a correlation analysis, the relationship between all the data in the Table 1 is investigated by the characteristics of SIU GaAs and for each pair of parameters the corresponding correlation coefficients are found. If the value of the correlation coefficient (modulo) is close to or equal to unity, then there is a linear relationship between the two variables. With a negative value, there is an inverse linear relationship (for example, between the indicators «dislocation density value» and «value of mechanical stresses» along the crystal diameter). If the correlation coefficient is zero, then there is no relationship between the two characteristics. Moreover, correlation coefficients close to unity are rarely observed. More often intermediate values take place. According to the classification adopted in [16], the values of the coefficients are interpreted as follows: with an absolute value of the correlation coefficient of 0.3–0.5 – as a weak relationship; in the range of values between 0.5 and 0.7 – as an average relationship; with values greater than 0.7 – as a high relationship. For a sample of 32 parameters (small sample size) there is a «threshold» coefficient value, which is significant. The threshold value of the coefficient is found on the basis of Student's t-criterion and amounted to 0.35. Therefore, if the correlation coefficient exceeds 0.35, then there is a significant correlation between the two parameters.

According to the results of statistical processing, the coefficients for each pair of parameters are given in the Table 2 (when calculating, the Pearson pair correlation coefficient is found [16]). Intersection cells show correlation coefficients for the corresponding pairs. If the coefficient value and its significance level are below the threshold value, then a dash is put in the cell. In cells with the «intersection of the same name», the coefficient is 1.

In the case where the parameters correlate with each other (the correlation coefficient is greater than 0.35), the scattering plot constructed in the corresponding coordinates is functionally close to a linear dependence, and the points on it are grouped near the regression line. An example of such dependence for the intensity of the edge band of photoluminescence on the dislocation density is shown in Fig. 1. The calculations show that for the analyzed dependence, the correlation relations are: for a linear form – 0.516; for hyperbolic – 0.064; for a second-order parabola – 0.085. The given values, as well as the value of the approximation error ($A=12\%$) make it possible to use the linear approximation form [16].

Table 2

Correlation matrix

Indicator	N_d	I_{PL}	Z	N_C	N_{Si}	N_{EL2}	σ
N_d	1	-0.91	0.87	–	0.18	–	-0.79
I_{PL}	-0.91	1	-0.84	–	–	0.45	0.51
Z	0.87	-0.84	1	0.66	-0.54	-0.86	-0.48
N_C	–	–	0.66	1	–	0.38	-0.16
N_{Si}	0.18	–	-0.54	–	1	0.62	0.44
N_{EL2}	–	0.45	-0.86	0.38	0.62	1	0.22
σ	-0.79	0.51	-0.48	-0.16	0.44	0.22	1

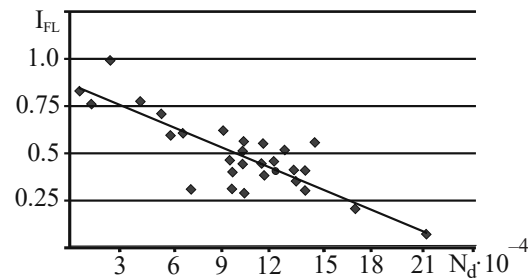


Fig. 1. Dependence of the intensity of the edge band of photoluminescence (I_{FL} , rel. units) on the dislocation density (N_d , cm^{-2})

It should be noted that in this case only the general fact of the correlation of two parameters is manifested – when one of them changes, the other changes. Such an indicator does not always indicate the mutual dependence of variables. This is because other variables that distort the true picture of dependence can influence the correlation of two factors.

For example, from Table 2 it follows that there is a close linear relationship between the concentration of EL2 centers and the intensity of edge radiation I_{FL} : with an increase in N_{EL2} , the I_{FL} value increases (the correlation coefficient is 0.45). Further, the analysis of the obtained data shows that such a correlation is classified as false (i. e., the interdependence does not exist). The reason is that the intensity of the edge radiation correlates with the concentration of the EL2 centers due to a change in the vacancy composition of the crystal (i. e., deviations from stoichiometry). By itself, the N_{EL2} value for the intensity of edge luminescence I_{FL} does not matter.

As a result, pair correlation coefficients in Table 2 do not always reflect the true degree of relationship. Therefore, in the course of the calculation, the correlation degree of parameters is clarified based on the analysis of particular correlation coefficients. The Table 3 shows the corresponding partial correlation coefficients, which show the degree of linear dependence in each pair of parameters. Such dependence is not distorted by the total influence of other parameters.

Table 3

Partial correlation matrix

Indicator	N_d	I_{FL}	Z	N_C	N_{Si}	N_{EL2}	σ
N_d	1	-0.84	0.71	–	–	–	-0.44
I_{FL}	-0.84	1	-0.68	–	–	–	–
Z	0.71	-0.68	1	–	–	–	-0.41
N_C	–	–	–	1	–	–	–
N_{Si}	–	–	–	–	1	–	0.22
N_{EL2}	–	–	–	–	–	1	–
σ	-0.44	–	-0.41	–	0.22	–	1

5. 2. Construction of a calculation model and calculation of the main indicators of communication

5. 2. 1. Calculation model

The calculation model of multiple linear regression is implemented as follows. For consideration, one of the independent variables is selected and parameters related to it are extracted from the entire data array. Based on the selected indicators, a regression equation is compiled [17]. For the remaining parameters, a study is conducted that establishes that the dependence is absent or statistically insignificant. From the correlation indicators, the statistical model includes those parameters that are most closely related to the characteristic selected for consideration.

Of greatest interest for analysis in a dataset is a parameter, the determination of which is statistically the most time-consuming. Such a parameter is chosen as the ultimate goal of the study and is called the response [17]. During the analysis, the influence of all other parameters on it is investigated. Three parameters are selected for consideration by stepwise regression: the concentration of the background silicon impurity; value of residual mechanical stresses; concentration of background carbon impurity.

Stepwise regression consists of two stages. The first stage is the backward stepwise [15]. All variables are included in the statistical analysis: dislocation density, edge radiation intensity, relative concentration of gallium and arsenic vacancies, concentration of EL2 centers, carbon, silicon atoms, and residual mechanical stresses. In the regression equation, on the one hand, all of the above parameters are present, and on the other hand, there is a response. Each term of such an equation passes the test for significance. If its significance is less than a given value, then the term is excluded from the equation and a test is performed for the next term. The process continues until all remaining terms are significant. As a result, only those variables remain on the regression equation, on which the response actually depends.

The second stage is forward stepwise [15]. Only the variables selected in the first stage are involved. Regression equations are compiled and analyzed step by step, in which variables with the maximum correlation coefficient with response are used in turn. Such variables are checked for statistical significance. Variables that fail verification are discarded.

For the correct regression analysis, parameters can't be used whose correlation coefficient is close to unity (0.8 or more) [15]. The quality criterion of the statistical model built as a result of stepwise linear regression is the coefficient of determination R^2 . This coefficient shows the extent to which the selected parameters explain the response [18].

Let's consider the calculation model for each of the selected parameters.

5. 2. 2. Determination of the concentration of the background silicon impurity

For this parameter, the results of the second stage of stepwise regression coincided with the results of the first stage: a statistical dependence of the concentration of silicon atoms N_{Si} on the concentration of centers EL2 and the vacancy composition index Z is established. An analysis of the constructed statistical model indicates that both parameters are equally significant for determining N_{Si} . The following regression equation is obtained:

$$N_{Si} = 3.59 + 0.12N_{EL2} - 5.15Z \text{ at } R^2=0.54. \quad (1)$$

Based on the value of the determination coefficient of 0.54, it is possible to conclude that the vacancy composition of the crystal and the concentration of deep centers EL2 account for 54 % of the changes in the concentration of silicon atoms (i. e., the response is determined by the selected parameters by 54 %). The remaining 46 % of the response changes are determined by other factors not taken into account in this model.

Note that due to the limited sampling, the value of $R^2=0.54$ for this technique [18] is acceptable for stating a good quality model. The corresponding result of a three-dimensional graphical construction of the dependence of the concentration of the background silicon impurity on the concentration of deep centers EL2 and the index of the vacancy composition of the crystal is shown in Fig. 2.

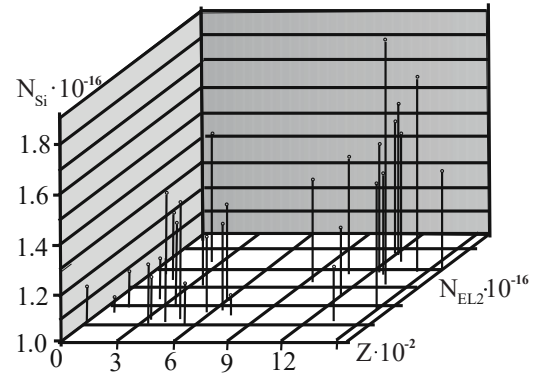


Fig. 2. Dependence of the concentration of the background silicon impurity (N_{Si} , cm^{-3}) on the concentration of deep centers EL2 (N_{EL2} , cm^{-3}) and the vacancy composition of the crystal (Z , rel. units)

From the Table 2 it can be seen that the vacancy composition index Z correlates with mechanical stresses σ (the correlation coefficient is minus 0.48). The concentration of silicon impurity also depends on σ – the correlation coefficient is 0.44 (Table 2). To clarify the true degree of dependence of the concentration of silicon impurity on residual mechanical stresses, let's find the value of the partial correlation coefficient for the pair « σ – N_{Si} » without taking into account the influence of the vacancy composition. Based on the corresponding statistical analysis, the interdependence of the selected parameters is established at a «fixed» value of Z . The calculated partial correlation coefficient is less than 0.25 (Table 3). Consequently, the concentration of silicon impurity, in fact, does not depend on the value of residual mechanical stresses and changes when they increase only due to a change in the vacancy composition of the crystal.

5. 2. 3. Determination of residual mechanical stress

According to the results of stepwise regression, it is found that the value of residual mechanical stresses σ in the presented sample varies depending on changes in the concentration of dislocations N_d and the vacancy composition of the crystal Z . A weak correlation is observed between σ and the concentration of silicon impurity (significance level 10 %). Other parameters for determining the value of σ are insignificant.

The resulting regression equation has the form:

$$\sigma = 28.91 - 0.27N_d - 0.18Z \text{ at } R^2=0.55. \quad (2)$$

The corresponding results of three-dimensional graphical construction of the dependence of residual mechanical stresses on the dislocation density and the vacancy composition index of the crystal are shown in Fig. 3.

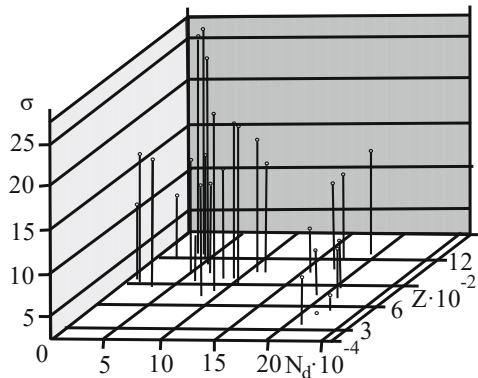


Fig. 3. Dependence of residual mechanical stresses on the dislocation density (N_d , cm^{-2}) and the vacancy composition index of the crystal (Z , rel. units)

5. 2. 4. Determination of the concentration of uncontrolled carbon impurities

The established character of the distribution of the concentration of uncontrolled carbon impurity N_C in the sample under consideration (Table 1) is different from the normal one. In addition, there is some heterogeneity in the dispersion of its parameters over the diameter of the crystal. Therefore, conducting regression and correlation analyzes based on the least squares method is incorrect, and the resulting regression equation may contain a significant error. Therefore, instead of stepwise regression, there are general statistical dependencies that replaced its result.

The implementation of such a multiple analysis technique allows to establish that the concentration of carbon impurities within the significance level of 10 % depends only on the vacancy composition of the crystal. The influence of other parameters, within the limits of their existing changes, on the N_C value is insignificant. The graph in Fig. 4 shows a scattering diagram in corresponding coordinates.

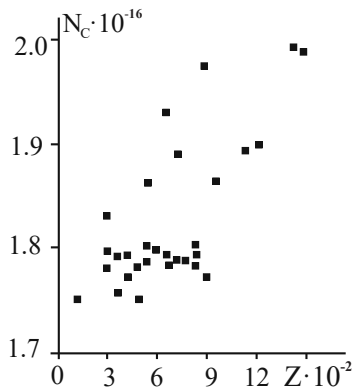


Fig. 4. The dependence of the concentration of the background carbon impurity (N_C , cm^{-3}) on the indicator of the vacancy composition of the crystal (Z , rel. units)

The scatter of values according to the parameter N_C in Fig. 4 is very large, which is obviously connected with the difference in the statistical dependences for the center and

periphery of the crystal. This effect is due to the release of arsenic from the side surface of the GaAs crystal during its cooling during growth.

6. Discussion of the results of multiple correlation analysis

Based on a generalization of the results of the regression analysis and accounting for the analysis of partial correlations, it is possible to conclude the following. In addition to previously known results – the relationship of mechanical stresses, dislocation density and the vacancy composition index of the crystal, as well as the relationship between the silicon concentration and the vacancy composition, there are quite unexpected ones. It is found that the concentration of silicon N_{Si} does not depend on the value of mechanical stresses, but is related to the concentration of cents EL2 (1).

The established dependence is most likely of the following nature. Defects EL2 are anti-structural – arsenic atoms at the position of the gallium atoms AsGa. Their concentration increases when the concentration of gallium vacancies, which occupy mobile arsenic, increases at the crystallization front. Despite the fact that in most studies gallium arsenide, silicon is considered an amphoteric impurity, it seems that in a specially non-alloyed material (at a low concentration), it predominantly precipitates onto gallium vacancies. Therefore, the conditions for its introduction at the crystallization front are similar to the conditions for the formation of EL2 cents. In this case, residual mechanical stresses are formed later during the cooling of the crystal and do not affect the redistribution of these defects.

For the same reason, mechanical stresses do not affect the concentration of carbon impurity, which, as established, in a significant range for N_{Si} values is determined only by the vacancy composition of the crystal (2). Therefore, it is possible to talk about the absence of redistribution of background impurities during the cooling of an undoped GaAs crystal. In specially doped crystals, where the concentration of impurities exceeds $1 \cdot 10^{17} \text{ cm}^{-3}$, the nature of the defect – impurity interaction requires additional statistical studies.

Thus, the calculated results of the multiple correlation analysis method are consistent with the known characteristics of GaAs crystals and the discovery of new previously unknown properties of gallium arsenide crystals. It is advisable to verify new calculation and theoretical data under the conditions of a technological experiment. The obtained result is a prerequisite for the transfer of the method under consideration to the process of modeling the structural properties of more complex systems of variable composition.

7. Conclusions

1. The use of multiple correlation analysis using Excel, STADIA, and SPSS Statistics 17.0 computer programs makes it possible to create a multivariate model of an undoped GaAs crystal with a varying vacancy composition. The considered procedure is effective for analyzing the influence of each of the factors included in the resulting model.

2. The calculation model of multiple linear regression according to three parameters has been implemented: the concentration of the background silicon impurity; the value of residual mechanical stresses; concentration of background carbon impurities. It is found that the vacancy composition

of the crystal and the concentration of deep EL2 centers account for 54 % of the changes in the concentration of silicon atoms. There is no relationship between the silicon concentration and the value of residual mechanical stresses. A change in the concentration of dislocations and an indicator of the vacancy composition of the crystal by 55 % determine the change in residual mechanical stresses. Within the significance level, the concentration of carbon impurity depends only on the vacancy composition of the crystal.

3. The calculated data obtained for the semi-insulating undoped gallium arsenide made it possible to identify and analyze both previously known and new correlation dependences and relationships between crystal parameters. It is concluded that there is no redistribution of background impurities during cooling of a GaAs crystal during its growth. The used regression model is promising for modeling multi-component bonds not only in binary crystals, but also for factor systems of variable composition.

References

1. Bombicz, P. (2019). A history and an industry perspective of crystallography. *Crystallography Reviews*, 25 (4), 263–263. doi: <https://doi.org/10.1080/0889311x.2019.1641098>
2. Luo, F., Cai, L.-C., Chen, X.-R., Jing, F.-Q., Alfè, D. (2012). Ab initio calculation of lattice dynamics and thermodynamic properties of beryllium. *Journal of Applied Physics*, 111 (5), 053503. doi: <https://doi.org/10.1063/1.3688344>
3. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J. et. al. (2009). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*, 66 (1), 12–21. doi: <https://doi.org/10.1107/s0907444909042073>
4. Suharso, S. (2010). Growth rate distribution of borax single crystals on the (001) face under various flow rates. *Indonesian Journal of Chemistry*, 6 (1), 16–19. doi: <https://doi.org/10.22146/ijc.21766>
5. Zlomanov, V., Zavrazhnov, A. (2008). *Nonstoichiometric compounds*. Intermetallics Research Progress. New York: Nova Science Publishers, 290.
6. Kovalenko, V. F., Shutov, S. V., Baganov, Y. A., Smykalo, M. M. (2009). Near band-edge luminescence of semi-insulating undoped gallium arsenide at high levels of excitation. *Journal of Luminescence*, 129 (9), 1029–1031. doi: <https://doi.org/10.1016/j.jlumin.2009.04.017>
7. Zhukov, N. D., Kabanov, V. F., Mihaylov, A. I., Mosiyash, D. S., Pereverzev, Y. E., Hazanov, A. A., Shishkin, M. I. (2018). Peculiarities of the Properties of III–V Semiconductors in a Multigrain Structure. *Semiconductors*, 52 (1), 78–83. doi: <https://doi.org/10.1134/s1063782618010256>
8. Gabibov, F. S., Zobov, E. M., Zobov, M. E., Kramynin, S. P., Pashuk, E. G., Khalilov, S. A. (2015). The effect of ultrasonic treatment on the energy spectrum of electron traps in n-GaAs single crystals. *Technical Physics Letters*, 41 (4), 362–365. doi: <https://doi.org/10.1134/s1063785015040239>
9. Shtan'ko, A. D., Litvinova, M. B., Kurak, V. V. (2010). Decrease of exciton radiation intensity in compensated gallium arsenide single crystals under influence of low electric field. *Functional Materials*, 17 (1), 46–51.
10. Zhukov, N. D., Kryl'skiy, D. V., Shishkin, M. I., Khazanov, A. A. (2019). On the Synthesis and Photoluminescence and Cathodoluminescence Properties of CdSe, CdTe, PbS, InSb, and GaAs Colloidal Quantum Dots. *Semiconductors*, 53 (8), 1082–1087. doi: <https://doi.org/10.1134/s1063782619080232>
11. Badea, A., Dragan, F., Fara, L., Sterian, P. (2016). Quantum mechanical effects analysis of nanostructured solar cell models. *Renewable Energy and Environmental Sustainability*, 1, 3. doi: <https://doi.org/10.1051/rees/2016003>
12. Lehmkuhler, F., Fischer, B., Müller, L., Ruta, B., Grübel, G. (2016). Structure beyond pair correlations: X-ray cross-correlation from colloidal crystals. *Journal of Applied Crystallography*, 49 (6), 2046–2052. doi: <https://doi.org/10.1107/s1600576716017313>
13. Lilvinova, M. B., Hertcova, N. Y., Seliverstova, S. R. (2003). The optical measurement technique of the definition of the GaAs structure deflection degree from stexiometry. *Proceedings of CAOL'2003. 1st International Conference on Advanced Optoelectronics and Lasers. Jointly with 1st Workshop on Precision Oscillations in Electronics and Optics (IEEE Cat. No.03EX715)*. doi: <https://doi.org/10.1109/caol.2003.1251297>
14. Litvinova, M. B., Shtan'ko, A. D. (2005). Influence of Structural Defects on the Mechanical Stress in the Impurity Diffusion Zone of GaAs Single Crystals. *Inorganic Materials*, 41 (8), 789–792. doi: <https://doi.org/10.1007/s10789-005-0211-0>
15. Elliott, A., Woodward, W. (2007). *Statistical Analysis Quick Reference Guidebook*. SAGE Publication. doi: <https://doi.org/10.4135/9781412985949>
16. Cankaya, S., Kayaalp, G. T., Sangun, L., Tahtali, Y., Akar, M. (2006). A Comparative Study of Estimation Methods for Parameters in Multiple Linear Regression Model. *Journal of Applied Animal Research*, 29 (1), 43–47. doi: <https://doi.org/10.1080/09712119.2006.9706568>
17. Kleinbaum, D., Kupper, L. L., Muller, K. E. (1988). *Applied Regression Analysis and Other Multivariable Methods*. Boston, USA: PWS-Kent, 664.
18. Saunders, L. J., Russell, R. A., Crabb, D. P. (2012). The Coefficient of Determination: What Determines a Useful R² Statistic? *Investigative Ophthalmology & Visual Science*, 53 (11), 6830. doi: <https://doi.org/10.1167/iovs.12-10598>