Selection of methods for increasing gas selectivity on the example of a sensor system based on In$_2$O$_3$–Ga$_2$O$_3$ semiconductor films

I E Demin
Omsk State Technical University, 11, Mira Ave., Omsk, 644050, Russia
e-mail: IvanDemin3467@gmail.com

Abstract. The article considers the possibility of quantitative determination of the selectivity of multisensor gas analyzers by applying methods of mathematical statistics on an example of a sensor system based on In$_2$O$_3$-Ga$_2$O$_3$ thin films operating in a mode with several working temperatures and in a dynamic mode with abrupt gas supply. The following methods were investigated: the quadrat sampling, the Thiessen-Voronoi polygons and the k-nearest neighbors. It is shown that the latter is the most attractive for quantitative determination of gas sensor system selectivity. The obtained results are applicable for determining the optimum ratios between the selectivity of the gas sensor system and the costs of achieving that selectivity by making improvements in the construction of sensing unit, a measuring circuit or an output signal processing unit.

1. Introduction

Gas sensors have found application in almost all areas where it is necessary to analyze, control or optimize the gas phase. Simple in design, cheap and reliable gas sensors are used to optimize the operation of the car engines and the mode of operation of household gas appliances, to monitor air pollution in industrial areas and to control indoor fires, to diagnose diseases and to analyze food freshness [1].

Thanks to significant scientific efforts, numerous types of gas sensors have been developed, including semiconductor, polymer, optical, electrochemical, surface acoustic wave sensors and others. Among the various sensors, the most interesting are semiconductor gas sensors, which have simple design, a wide range of detectable gases, compatibility of manufacturing technology with planar microelectronic technology, long-term stability and reliability [2].

The new applications arise in which precise determination of the composition of the gas phase is required. Multisensor gas analyzers are indispensable here [3]. Unlike simple gas sensors containing in their design

– a gas sensor,
– a unit for maintaining the operating mode
– and a measuring circuit,

multisensor systems are supplemented with a data processing unit that implements some mathematical algorithms for analyzing the array of output signals from sensors. The use of these units increases the cost of the sensor system, but it allows significant extension of its ability to determine the quantitative and qualitative composition of the air. To select the optimal relationship between the complexity of the algorithm for data processing unit of the sensor system and the cost of the system, it is necessary to find numerical characteristics that allow comparing different algorithms.
2. Problem statement
The most accepted method of increasing the selectivity of gas sensors is an increase in the number of information parameters of the sensor system [3–5]. Information parameters are parameters of the output signals of the system, which contain information on the concentration or type of analyzed gases. These parameters can be:

– the values of the gas induced resistance response in the steady state, in the abrupt gas supply mode or in the thermal cycling mode, as well as
– the parameters of the temperature or time dependences of the sensor response.

Previously obtained results showed the possibility of determining five information parameters using a sensor system based on an array of In$_2$O$_3$–Ga$_2$O$_3$ thin films operating in a multi-temperature mode and in a mode with abrupt gas supply [4]. Four of the proposed parameters are independent of the concentration of the analyzed gas and the operating temperature of the sensors. They can be used to identify the type of gas. The fifth parameter is the amplitude of the gas induced resistance response. It carries information on the gas concentration, so its value can be used to determine the quantitative composition of the gas phase. The values of the first four parameters for several common combustible gases are proposed in Table 1. In this table $E_{S\ LT}$ and $E_{S\ HT}$ are the activation energies of the temperature dependences of the gas induced resistance response in the ranges of low (350–500 °C) and high (500–750 °C) temperatures, $A_{\tau}$, $E_{\tau}$ are the pre-exponential factors and activation energies of the temperature dependences of the time constants of gas response in the mode with abrupt gas supply.

| № | Parameter | Test gas |
|---|-----------|----------|
|   |           | Ethanol  | Acetone | Ammonia | LPG     |
| 1 | $E_{S\ LT}$, eV | -0.805    | -0.898  | -0.743  | -0.589  |
| 2 | $E_{S\ HT}$, eV | 1.067     | 0.921   | 0.129   | 0.812   |
| 3 | $A_{\tau}$, ms | 495       | 512     | 0.548   | 308     |
| 4 | $E_{\tau}$, eV | 0.712     | 0.755   | 1.297   | 0.935   |

Obtaining additional information parameters of the sensor system requires the complication of its design and the increase in complexity of the algorithm for processing the output signals. The increase in selectivity of the sensor system provided by these improvements can be obtained at the expense of disproportionately high costs of

– increase of the cost of the system,
– enlargement of its dimensions,
– reduction of its reliability and usability.

In this connection, it is necessary to define numerical criteria characterizing the selectivity of the sensor system which allow to commensurate the increase in its selectivity with the deterioration of its other properties. Obtaining such criteria will provide answers to the following questions.

– How much increased selectivity with increasing number of sensors in the system?
– How much increased the probability of determining the type of gas after improving the measuring circuits?
– How much increased selectivity with increasing number of information parameters?
– How much increased selectivity after introducing new algorithms for processing the output signals of sensor systems?

Among the scientific results presented by the research groups, there is practically no information on the methods of processing the output signals of multisensor gas analyzers and on the criteria characterizing the selectivity of such systems [6–12]. All of the above indicates the high relevance of research on mathematical methods for increasing the selectivity of sensor systems.
Thus, the goal of this paper is to investigate the methods of processing the output signals of gas sensor systems using In$_2$O$_3$-Ga$_2$O$_3$ thin films for the purpose of increasing their selectivity.

Main goals:
– obtaining an array of response values of information parameters of the sensor system based on In$_2$O$_3$-Ga$_2$O$_3$ thin films operating in a mode with multiple working temperatures and in a dynamic mode with abrupt gas supply for cases of gas response to ethanol, acetone, ammonia and liquefied petroleum gas (LPG);
– analysis of mathematical methods for determining the selectivity of multisensor systems;
– application of the studied methods on the array of considered information parameters and
– comparison of the results obtained with the use of that methods.

3. Experimental
In$_2$O$_3$-Ga$_2$O$_3$ thin films were obtained by pulsed laser deposition from pressed powder targets. A detailed description of the deposition unit and deposition modes are given in [13]. Values of information parameters that are independent of the type of analysed gas were obtained from the temperature dependences of the gas induced resistance response and from the temperature dependences of the response time constants for abrupt gas supply. A detailed description of the procedure for obtaining these dependences, as well as the design of the experimental setup and the measurement regimes used are given in [13].

The results discussed in this paper were obtained for a sensor system consisting of four 50%In$_2$O$_3$–50%Ga$_2$O$_3$ films. This composition was chosen on the basis of an analysis of the dependence of the parameters of the gas induced resistance response on the film composition given in [14]. That article shows that films of 50%In$_2$O$_3$–50%Ga$_2$O$_3$ composition have the greatest response amplitude, the smallest response time constants and the lowest temperatures of the maximum response. The operating temperatures for the sensors of the system were 440 °C, 500 °C, 560 °C and 620 °C.

The following information parameters that do not depend on the type of analyzed gas and working temperature were used:
– the activation energies of the gas induced resistance response in the low-temperature (E$_{SLT}$) and high-temperature ranges (E$_{SHT}$);
– the pre-exponential factors (A$_{\tau}$) and the activation energies (E$_{\tau}$) of the temperature dependences of the time constants of the gas induced resistance response (\(\tau\)).

The values of E$_{SLT}$ and E$_{SHT}$ were determined by approximation of the sections of the temperature dependences of the gas resistance response by linear functions in the ln(S), 1/T coordinates in the temperature ranges 350–500 °C and 500–750 °C. The values of A$_{\tau}$ and E$_{\tau}$ were obtained by approximation of the temperature dependences of the time constants of the gas induced resistance response for abrupt gas supply by linear functions in the coordinates ln(\(\tau\)), 1/T in the temperature range 350–750 °C. Approximation was carried out by the method of least squares [4, 14].

4. Theory
The following mathematical model was used to analyze the selectivity of a multi-sensor system [15]. A multidimensional space is considered in which different directions correspond to different information parameters of the sensor system. Possible values of the response of the sensor system to each analyzed gas are placed in this space in the form of points. The regions of the considered space that span the points corresponding to a given gas are called classes. During the operation of the sensor system, the data processing unit tries to determine to which class the current values of the information parameters belong. As a result, a conclusion is drawn on the presence of an appropriate gas in the atmosphere at the moment [15].

The selectivity of the considered multisensor system is judged by the distance between classes. If the classes are located close to each other, that is, the distances between them are small in comparison with the linear dimensions of the classes, then a conclusion is drawn about the low selectivity of the sensory system. A uniform distribution of classes over the space indicate a high selectivity.
To judge the uniformity of the distribution of classes over space, the following main methods of mathematical statistics are used [16]: k-nearest neighbors method [17], quadrat sampling method [18] and Thiessen-Voronoi polygons method [19].

To apply the k-nearest neighbors method one should begin with the calculation of the Euclidean-Mahalanobis distances between the centers of the classes [17]. Next, for each class, the «nearest neighbor» is defined, that is, the class, distance to which is minimal among distances from the given class to the remaining classes of the population. The characteristic of the uniformity of the distribution, and hence the selectivity of the sensor system is the mean distance to the nearest neighbor. The advantage of this method is low computational costs. The disadvantages of this method were not revealed.

The application of the method of quadrat sampling begins with the partition of the entire investigated space of information parameters into equal areas, called quadrats or squares. Then the density of class distribution by quadrats is calculated. In this method, the dispersion of the distribution density is a characteristic of the uniformity of the distribution of classes over space. The advantage of the method of squares is low computational costs. Its drawback lies in the fact that the reliability of the result depends on the number of classes. With a small number of classes in the studied population, the method is not applied.

A characteristic feature of the Thyssen-Voronoi polygon method is the decomposition of the entire investigated space of information parameters into regions called polygons. Each polygon contains only one class. Into the inner space of the polygon fall points, the distance from which to the inner class is less than the distance from them to other classes of the studied population. The name "polygon" was originally used for the two-dimensional case, but it was also retained for the multidimensional case. Characteristic of the selectivity of the sensor system is a dispersion of areas (volumes) of polygons. The advantage of this method is the high clearness of the result. Its disadvantage is the relatively high computational complexity.

The next section demonstrates that the results obtained by the methods studied are equivalent. The method of the k-nearest neighbors is of the greatest interest for determining the selectivity of sensor systems, because it has the optimal combination of advantages and disadvantages.

5. Results and discussion

For the application of the k-nearest neighbors method, an array of the response values of the information parameters of the sensor system based on thin films of In2O3–Ga2O3 operating in the multi-temperature mode and in the mode with abrupt gas supply, for cases of ethanol, acetone, ammonia and LPG was obtained. The description of the information parameters used, as well as their values for all the gases studied, were presented above in the section "Experimental".

A procedure for normalizing the values of information parameters was carried out. It consists in a linear transformation, as a result of which the zero value of the coordinate for each axis is associated with the minimum of the experimentally obtained values of the information parameters. The value equal to one on that axis is adjusted to the maximum of the obtained values of the given information parameter.

Thus obtained coordinate values of the centers of classes corresponding to different studied gases in the space of information parameters are shown in Table 2.

| №  | Coordinate | Test gas | Ethanol | Acetone | Ammonia | LPG |
|----|------------|----------|---------|---------|---------|-----|
| 1  | ES_LT      |          | 0.698   | 1.000   | 0.497   | 0.000|
| 2  | ES_HT      |          | 1.000   | 0.844   | 0.000   | 0.728|
| 3  | A_r        |          | 1.000   | 0.993   | 0.000   | 0.943|
For the application of the k-nearest neighbor method, Euclidean-Mahalanobis distance matrices between the centers of classes were obtained for the cases of different information parameters from the studied population, either individually or in different combinations of two, three or four parameters. This corresponds to one-, two-, three- or four-dimensional spatial representations of classes. For all the matrices obtained, the average distances to the nearest neighbor (\(<k\text{-NN}>\)) were determined, as well as the minimum distances to the nearest neighbor in the matrix (\(k\text{-NN}_{\text{min}}\)) and the number of «small» distances \((N_{\text{k-NN}})\). The distance to the nearest neighbor was considered as «small» if its value is less than the sum of the radii of these classes along the line connecting their centers. The form of the classes was considered to be ellipsoidal. The average diameter of classes for the population under study was 0.30.

The Thiessen-Voronoi polygon method was used for the same combinations of information parameters as for the k-nearest neighbor method, that is, for spatial representations from one-dimensional to four-dimensional. The dispersions of polygon areas \((\sigma_S)\) as well as the minimum polygon areas \((S_{\text{min}})\) and the numbers of «small» polygons \((N_S)\) were determined. The polygon was considered as «small» if its volume is smaller than the volume of an ellipsoid describing a given class. The average volume of the ellipsoid for the considered population was 0.25.

The results of calculations of the uniformity characteristics of the class distribution over the space of information parameters are presented in Table 3. The k-nearest neighbors method and the Thyssen-Voronoi polygon method were used. The method of squares was not used because of its inapplicability for a small number of classes (in a considered case there are only four classes).

### Table 3. Characteristics of the distribution uniformity of the classes in the space of information parameters

| Case                      | Information parameter | Characteristic |
|---------------------------|-----------------------|----------------|
|                           | \(E_{SLT}\)           | \(<k\text{-NN}>\) | k-NNmin | Nk-NN | \(\sigma_S\) | \(S_{\text{min}}\) | \(N_S\) |
| One-dimensional           |                       | \(E_{SLT}\) | 0.534   | 0.201 | 2     | 0.282   | 0.151   | 2     |
|                           |                       | \(E_{SHIT}\) | 0.519   | 0.116 | 3     | 0.281   | 0.085   | 3     |
|                           | \(A_t\)               | 0.508   | 0.007   | 3     | 0.282   | 0.006   | 3     |
|                           | \(E_t\)               | 0.547   | 0.073   | 2     | 0.296   | 0.052   | 3     |
| Two-dimensional           |                       | \(E_{SLT}, E_{SHIT}\) | 0.830   | 0.340 | 0     | 0.730   | 0.255   | 1     |
|                           | \(E_{SLT}, A_t\)      | 0.867   | 0.302   | 1     | 0.699   | 0.227   | 2     |
|                           | \(E_{SLT}, E_t\)      | 0.761   | 0.172   | 2     | 0.532   | 0.131   | 2     |
|                           | \(E_{SHIT}, A_t\)     | 0.745   | 0.126   | 3     | 0.533   | 0.102   | 3     |
|                           | \(E_{SHIT}, E_t\)     | 0.761   | 0.172   | 2     | 0.550   | 0.143   | 3     |
|                           | \(E_t, A_t\)          | 0.772   | 0.073   | 2     | 0.632   | 0.061   | 2     |
| Three-dimensional         |                       | \(E_{SLT}, E_{SHIT}, A_t\) | 1.036   | 0.340 | 0     | 1.062   | 0.366   | 0     |
|                           | \(E_{SLT}, E_{SHIT}, E_t\) | 1.016   | 0.347   | 0     | 1.073   | 0.372   | 0     |
|                           | \(E_{SLT}, A_t, E_t\) | 1.043   | 0.311   | 1     | 1.099   | 0.246   | 1     |
|                           | \(E_{SHIT}, A_t, E_t\) | 0.936   | 0.172   | 2     | 0.894   | 0.133   | 2     |
| Four-dimensional          |                       | \(E_{SLT}, E_{SHIT}, A_t, E_t\) | 1.181   | 0.348 | 0     | 1.521   | 0.381   | 0     |

The results presented in Table 3 demonstrate the equivalence of the application of the k-nearest neighbor and Thyssen-Voronoi polygon methods. Thus, from a practical point of view, the first method is more attractive, because it requires less computational costs.

Also, according to the proposed data, it is clear that the selectivity of the sensor system increases with an increase in the number of information parameters, in other words, with an increase in the dimensionality of the considered space. This is indicated by an increase in the values of \(<k\text{-NN}>\) and \(\sigma_S\), an increase in the values of k-NNmin and \(S_{\text{min}}\), as well as a decrease in the values of \(N_{k\text{-NN}}\) and \(N_S\).
The most attractive of the one-dimensional cases is the case of the $E_{SLT}$ parameter, from the two-dimensional ones – the case of the parameters $E_{SLT}$, $E_{SHT}$, from the three-dimensional cases – the case of $E_{SLT}$, $E_{SHT}$, $A_t$. Among all considered the best for all the studied characteristics is the four-dimensional case, in which all information parameters ($E_{SLT}$, $E_{SHT}$, $A_t$, $E_t$) are used.

Also, the results obtained can be used to determine the relationship between the increase in selectivity and the costs of increasing it. The costs of designing the sensor system possible to conduct measurements in a mode with multiple working temperatures or in the mode with abrupt gas supply may be arranged in ascending order of prices as follows. The design with multiple operating temperatures have lowest cost. The system with abrupt gas supply requires significantly higher expenses. Undoubtedly, the most expensive is the design, in which both approaches are realized. Thus, taking into account the ratio of the cost and selectivity of the sensor system, it is optimal to use two information parameters $E_{SLT}$ and $E_{SHT}$ (mode with multiple operating temperatures), in which a sufficiently high selectivity is realized (see Table 3) at a reasonable price of the sensor system.

In the same way the optimal relationships between the complexity of the sensor system design and its selectivity, between system reliability and selectivity, the optimal amount of sensors in the system and the required accuracy of obtaining the output characteristics by the measurement unit can be determined.

6. Conclusion
The results presented show the possibility of quantitative determination of the selectivity of multisensor gas analyzers by applying mathematical statistics methods by an example of a sensor system based on $\text{In}_2\text{O}_3$–$\text{Ga}_2\text{O}_3$ thin films operating in a mode with multiple operating temperatures and in a mode with abrupt gas supply. Among the methods studied: the Thiessen-Voronoi polygon method, the quadrat sampling method and the k-nearest neighbors method – the latter is the most attractive for quantitative determination of selectivity. The results obtained are applicable for determining the optimum relationships between the selectivity of the sensory system and the costs of achieving such selectivity by making improvements in the construction of sensing unit, a measuring circuit or an output signal processing unit.

Acknowledgments
This work was carried out within the project «Young scientist»; research effort № 18076B.

References
[1] Fleischer M and Lehmann M (eds.) 2012 Solid State Gas Sensors – Industrial Application (Berlin Heidelberg: Springer-Verlag, 2012)
[2] Korotcenkov G 2013 Handbook of Gas Sensor Materials. Properties, Advantages and Shortcomings for Applications Volume 1: Conventional Approaches (New York: Springer, 2013)
[3] Reimann P and Schutze A 2014 Sensor Arrays, Virtual Multisensors, Data Fusion, and Gas Sensor Data Evaluation (Berlin Heidelberg: Springer-Verlag, 2014)
[4] Demin I E and Kozlov A G 2018 J. Phys.: Conf. Ser. 944 012027
[5] Korotcenkov G and Cho B K 2013 Sens. Actuators B 188 709–28
[6] Maziarz W, Potempa P, Sutor A and Pisarkiewicz T 2003 Thin Solid Films 436 127–31
[7] Solorzano A, Rodriguez-Perez R, Padilla M, Graunke T, Fernandez L, Marco S and Fonollosa J 2018 Sens. Actuators B 265 1142–54
[8] Mascini M, Pizzoni D, Perez G, Chiarappa E, Di Natale C, Pittia P and Compagnone D 2017 Biosens. Bioelectron. 93 161–9
[9] Fernandez L, Marco S and Gutierrez-Galvez A 2014 Procedia Engineering 87 851–4
[10] Fonollosa J, Rodriguez-Lujan I and Huerta R 2015 Data in Brief 3 85–9
[11] Ziyatdinov A, Fonollosa J, Fernandez L, Gutierrez-Galvez A, Marco S and Perera A 2015 Sens. Actuators B 3 538–47
[12] Su P-G and Chuang T-Y 2017 Sens. Actuators A 263 1–7
[13] Demin I E and Kozlov A G 2015 Proc. 2015 Int. Siberian Conf. on Control and
Communications (SIBCON) (Hoboken: IEEE) 7147037
[14] Demin I E and Kozlov A G 2017 J. Phys.: Conf. Ser. 858 012009
[15] Kim D, Kim S-J and Kim S 2018 Sens. Actuators B 265 600–8
[16] Bhunia G S, Shit P K and Maiti R 2016 J. Saudi Soc. Agricultural Sci. 17 114–26
[17] Lu J, Zhu Q and Wu Q 2018 Eng. Appl. Artif. Intel. 72 213–27
[18] Fanini L and Lowry J K 2016 Ecol. Indic. 60 358–66
[19] Mu L 2009 International Encyclopedia of Human Geography (Amsterdam: Elsevier, 2009) 231–6