Selectivity enhancement of SiC-FET gas sensors by combining temperature and gate bias cycled operation using multivariate statistics

Christian Bur, Manuel Bastuck, Anita Lloyd Spetz, Mike Andersson and Andreas Schuetze

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:

Christian Bur, Manuel Bastuck, Anita Lloyd Spetz, Mike Andersson and Andreas Schuetze, Selectivity enhancement of SiC-FET gas sensors by combining temperature and gate bias cycled operation using multivariate statistics, 2014, Sensors and actuators. B, Chemical, (193), 931-940.

<http://dx.doi.org/10.1016/j.snb.2013.12.030>

Copyright: Elsevier

<http://www.elsevier.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-104641>

Selectivity Enhancement of SiC-FET Gas Sensors by Combining Temperature and Gate Bias Cycled Operation Using Multivariate Statistics

Christian Bur^{1,2*}, Manuel Bastuck¹, Anita Lloyd Spetz², Mike Andersson² and Andreas Schütze¹

¹Lab for Measurement Technology

Department of Mechatronics

Saarland University

D-66123 Saarbrücken, GERMANY

²Div. of Applied Sensor Science

Department of Physics, Chemistry and Biology

Linköping University

SE-58183 Linköping, SWEDEN

* Corresponding author:

Mail: c.bur@LMT.uni-saarland.de

Phone: 0049 681 302 3904

Fax: 0049 681 302 4665

E-mail addresses of co-authors:

Manuel Bastuck: m.bastuck@lmt.uni-saarland.de

Anita Lloyd Spetz: spetz@ifm.liu.se

Mike Andersson: mikan@ifm.liu.se

Andreas Schütze: schuetze@lmt.uni-saarland.de

Abstract

In this paper temperature modulation and gate bias modulation of a gas sensitive field effect transistor based on silicon carbide (SiC-FET) are combined in order to increase the selectivity. Data evaluation based on extracted features describing the shape of the sensor response was performed using multivariate statistics, here by Linear Discriminant Analysis (LDA). It was found that both temperature cycling and gate bias cycling are suitable for quantification of different concentrations of carbon monoxide. However, combination of both approaches enhances the stability of the quantification, respectively the discrimination of the groups in the LDA scatterplot. Feature selection based on the stepwise LDA algorithm as well as selection based on the loadings plot has shown that features both from the temperature cycle and from the bias cycle are equally important for the identification of carbon monoxide, nitrogen dioxide and ammonia. In addition, the presented method allows discrimination of these gases independent of the gas concentration. Hence, the selectivity of the FET is enhanced considerably.

Keywords:

SiC-FET, temperature modulation, gate bias modulation, selectivity, feature extraction, pattern recognition

Highlights:

- Combination of temperature and gate bias modulation is presented
- Significant features describing the shape of the sensor response have been selected
- Linear Discriminant Analysis (LDA) is used for data evaluation
- Selectivity of the sensor can be increased by our suggested approach
- Discrimination of CO, NO₂ and NH₃ independent of the concentration is possible

1. Introduction

Field effect based gas sensors have been studied for many years [1]. Besides Schottky diodes and MOS capacitors, field effect transistors (FETs) are of great interest due to easy read out of the transistor and their improved stability. Besides silicon, which is commonly used as a substrate material, silicon carbide (SiC) or other wide band gap materials provide the possibility for high temperature application (up to 800 °C for SiC). Due to its chemical inertness, SiC is a suitable material for sensors operating in harsh environments, e.g. directly in the exhaust stream of combustion engines [2].

By using catalytically active gate materials like platinum, iridium or palladium excellent gas-sensitivity is achieved for SiC-FETs. The sensing properties of the FET depend mainly on the material and its structure as well as on the operating temperature of the device. For a dense, homogenous layer of palladium, hydrogen molecules adsorbing on the surface dissociate and rapidly diffuse through the dense metal layer. At the metal-insulator, here SiO₂, interface hydrogen atoms form a polarized layer of hydroxyl groups influencing the density of mobile carriers in the channel of the transistor [3].

For the detection of non-hydrogen containing gases like carbon monoxide (CO), but also for ammonia a dense gate layer is not suitable. In order to detect these gases a porous layer of e.g. platinum is necessary allowing interaction with the oxide surface like spill over from the metal and detection of dipoles formed on the oxide surface. Three phase boundaries (metal, oxide and gas) have shown to be very important for the gas response [4, 5].

Schalwig et al. suggested that the sensing mechanism of hydrogen and non-hydrogen containing gases can be explained by spill-over effects of adsorbed oxygen. Negatively charged oxygen ions on the sensor surface will influence the electric field in the underlying oxide and hence the sensor characteristics, i.e. the IV curve [6]. Reducing gases like CO would react with adsorbed oxygen and thereby lower the density of oxygen on the surface similar to the sensing mechanism of resistive type metal oxide sensors.

A common challenge for chemical sensors in general is their poor selectivity and long-term stability. It was reported earlier that the selectivity of SiC-FET can be enhanced by dynamic operation, i.e. temperature cycled operation (TCO) [7]. Temperature modulation gives rise to several advantages [8]: Due to the continuously changing temperature the sensor is never in steady state; instead, the transient response results in a unique response pattern or signature for each gas thereby increasing the selectivity. Another advantage of a broad temperature range covered by the cycle is that measurements are made at the temperature corresponding to

maximum sensitivity for each gas. Hence, the sensitivity of the sensor system is increased. Finally, the high temperature part of the cycle facilitate desorption of adsorbates from the sensor surface resulting in improved stability.

Compared to two-terminal devices or resistive type metal oxide sensors, transistor structures offer another possibility for dynamic operation: changing the gate bias. The influence of the gate bias on the sensing properties has been studied by Nakagomi et al. [9]. The gate bias mainly influences the threshold voltage of the transistor but also its sensitivity. Similar to temperature modulation, the gate bias can be cycled leading to a similar increase in selectivity as TCO [10]. Gate bias cycling leads to pronounced hysteresis effects which is influenced dramatically by the gas exposure [11]. Thus, one can expect to gain more information from this mode of operation. Hysteresis effects were also found for other transistor type sensors and it was reported that the hysteresis is strongly affected by loading and unloading of trapping states mostly occurring in the underlying oxide layer of the FET [12, 13]. Most often, hysteresis is unwanted and great effort is spent to decrease this effect. In a sensor, however, hysteresis can be used to obtain additional information contained in the transient behavior [14].

Pohle et al. suggested using gate pulses of a suspended and floating gate FET for transient readout of the sensor. They could show that transient operation of field effect transistors can be used to enhance the selectivity and baseline stability of the sensor [15, 16].

In this work, the combination of temperature cycled operation (TCO) and gate bias cycled operation (GBCO) is investigated in order to increase the selectivity of gas sensitive field effect transistors further. Different methods of data pre-processing, feature extraction and feature selection were studied and compared showing that TCO and GBCO are complementary in the obtained information.

2. Experimental

2.1. Sensor Setup

For all measurements an n-channel depletion type SiC-FET was used (Fig. 1a, SenSiC AB, Sweden). The catalytic gate metallization is porous platinum with a thickness of 25 nm. The porous layer including cracks that appear during deposition and annealing has a pore size in the range of 0-30 nm width and 0-200 nm length with a wide distribution over this range. The gate length is 10 μm and the gate width is 300 μm . A detailed description of the sensor incl. SEM images of the gate layer and of the manufacturing process has been reported previously [17].

The SiC die is glued onto a ceramic heater (Heraeus GmbH, Germany) to allow precise heating of the sensor. A Pt-100 temperature sensor was attached next to the sensor as temperature reference. The heater with the FET and temperature sensor is mounted on a 16-pin TO8 header (Fig. 1b). Electrical contacts to the FET are realized via gold wire bonding. As shown in Fig. 1c, each sensor die holds four transistor structures: two with short-circuited gate connected to the drain contact (two-terminal devices) and two transistors with separate gate contact allowing independent control of the gate bias.

Sensor control and data acquisition was performed by a combined system developed by 3S GmbH, Germany. The system controls the sensor temperature with an analog control circuit with a resolution of 1 $^{\circ}\text{C}$ and can adjust the gate bias from -7 V to +7 V using an 8 bit DAC, i.e. with a resolution of approx. 50 mV. Data acquisition is performed with a 10 bit ADC for measuring the drain-source current with a measurement range of 1-1000 μA , resulting in a resolution of approx. 1 μA . The acquisition rate for all measurements was 10 Hz. Most important for the subsequent data acquisition, the system allows precise synchronization of the temperature and gate bias cycles with the data acquisition.

2.2. Sensor operating mode

A constant drain-source voltage of $V_D = 4\text{ V}$ was chosen as the typical operating mode keeping the transistor in the saturation region at all times. The recorded sensor response is the drain-source current I_D (saturation current). Operating in the saturation region provides a larger dynamic range than the linear regime, and the former is necessary when changing temperature and gate bias at the same time, i.e. when combining TCO and GBCO.

Figure 2a shows the combination of the temperature cycle (T-cycle) and gate bias cycle (GB-cycle) used in this work and the standardized sensor response in b (for signal pre-processing see section 3.1). The T-cycle consists of two temperature plateaus (260 °C and 200 °C); on each plateau, the gate bias is ramped continuously from -1 V to 3 V and back again, respectively (GB-cycle) with a slope of 0.4 V/s.

The chosen cycle consist of different parts where the sensor response is influenced by the temperature change at zero bias (cf. Fig. 2 part A and C) and where the response is influenced by the gate bias at constant temperature (cf. Fig. 2 part B and D). This allows studying both operating modes nearly independently. The total length of the T-cycle containing two GB-cycles is 90 s to account for the thermal time constant of the sensor setup (approximately 4 s) and for the slope of the gate bias ramp. In general, effects observed during continuous changes of the gate bias are slow (deep traps with time constants in the order of seconds) [13]. The chosen cycle is therefore a compromise between a short cycle length to allow fast measurements and maximum information gained. The slope of the bias ramp has an influence on the selectivity of the sensor as reported in [11].

2.3. Gas profile

For validation of the novel sensor operating mode in controlled laboratory tests typical test gases like carbon monoxide (CO, 200, 400, 600 and 800 ppm), nitrogen dioxide (NO₂, 20, 30, 40, 50 ppm) and ammonia (NH₃, 30, 45, 60, 75 ppm) were chosen. The total flow over the sensor was kept constant at 200 ml/min. The carrier gas was 5 % oxygen in nitrogen under dry conditions; thus, the gas profile is similar to studies of gas sensors for exhaust gas applications, i.e. for SCR systems [18, 19]. Each gas exposure had a duration of 30 minutes during which approx. 20 combined T/GB-cycles are recorded. The intervals between two different concentrations of the same gas and between two different gases were 1.5 and 4 hours, respectively, in order to avoid memory effects or carry-over from the previous exposure.

3. Signal Processing

The applied signal processing consists of three main parts as shown in Fig. 3: The signal pre-processing aims to reduce noise and drift by smoothing and normalization of the data. In dynamic operation smoothing and normalization is typically performed for each cycle. Smoothing can enhance the quality of the discrimination [20] but is not compulsory. In this work, smoothing is implicitly performed by the feature extraction mechanism, see below.

The main core of the signal processing is then the feature extraction and feature selection. Here, features describing the shape of the sensor response are extracted and later used for discrimination. Discrimination is based on Linear Discriminant Analysis (LDA) projecting the high dimensional feature space into a lower dimensional space for classification, see below. The discrimination is validated afterwards in order to avoid over-fitting. Feature selection can be done before and after the discrimination. However, our approach is to perform feature selection after the discrimination to make use of the fact that more features result in a better discrimination. Afterwards, redundant or less significant features can be removed (top-down approach). In general the applied signal processing is an iterative optimization process based on calibration (training) data.

The final part of the developed signal processing scheme is the post-processing, i.e. classification of unknown (test) data. While the first two parts represent training of the system, the classification procedure is used in the evaluation step. Unknown data are projected using the same algorithms and classified by comparison with the training data.

For the signal processing Matlab programs developed in our group were used which are based on the Matlab statistics toolbox.

3.1. Pre-processing

As mentioned earlier the aim of the signal pre-processing is to reduce sensor drift. For static sensor operation this step is sometimes called baseline correction or baseline manipulation [21]. The baseline of the sensor operated in a cyclic mode is more complex. In fact, there is no single baseline as each point of a cycle plotted over the cycle number respectively the time (i.e. quasi-static sensor response) can be seen as a sensor signal at a certain temperature and/or bias. Therefore, drift correction is done separately for each cycle. There are different ways of normalizing a cycle. For example, additive drift, i.e. sensor offset, could be corrected by

subtracting a constant value from each point in the cycle, e.g. setting the cycle minimum to zero. An effective normalization to reduce multiplicative drift is setting the cycle mean to one:

$$y_{i,j}^{norm} = \frac{y_{i,j}}{mean(y_i)} = \frac{y_{i,j}}{\bar{y}_i} = \frac{y_{i,j}}{\frac{1}{p} \sum_{j=1}^p y_j} \quad (1)$$

Here, $y_{i,j}$ represents the data point j in cycle i and p is the number of data points per cycle. $y_{i,j}^{norm}$ is the normalized data point when dividing each data point $y_{i,j}$ by the mean value of the cycle $\bar{y}_i$. This normalization has been successfully applied to metal oxide gas sensors several times [14] and also for gas sensitive field effect transistors [7, 20]. In addition to reducing sensor drift, normalization can also imply a loss of information. When setting the mean of each cycle to one, information regarding discrimination of different gases which is contained in the raw data close to the mean of the cycle, is discarded. This can of course be taken into account by using the mean value as a separate feature in the subsequent discrimination.

Sensor drift cannot only occur in terms of a drifting baseline but also in scaling of the response over the cycle. Therefore, normalization in this work is done by mapping the cycle into a standardized cycle which has standard deviation of one and zero mean as shown in equation (2). In static operation this normalization is sometimes called *sensor auto-scaling* [21]. Again, the standard deviation of a cycle can also be retained as a feature to minimize the loss of information.

$$y_{i,j}^{stand} = \frac{y_{i,j} - mean(y_i)}{std(y_i)} = \frac{y_{i,j} - \bar{y}_i}{\sqrt{\frac{1}{p-1} \sum_{j=1}^p (y_j - \bar{y}_i)^2}} \quad (2)$$

Here, $y_{i,j}$ represents the data point j in cycle i and p is the number of data points per cycle. $y_{i,j}^{stand}$ is the standardized data point when subtracting the mean value of the cycle $\bar{y}_i$ from each data point $y_{i,j}$, i.e. to obtain zero mean, and dividing the result by the standard deviation of the cycle, i.e. to obtain a standard deviation of one.

In [22] the normalization (eq. 1) has been applied to the combined temperature and gate bias cycle (cf. Fig. 2a) and can hence be compared with the result of the standardized cycle presented in this work (cf. Fig. 2b).

Normalization or standardization of sensor data (here: cycle) is effective to reduce the influence of sensor drift. However, it causes also a loss of information in certain parts of the cycle, e.g. when setting cycle mean to one, but on the other hand other parts of the cycle are emphasized. This can be seen in the standardized dynamic sensor response in

Fig. 2b. Comparing interval 2 (260 °C and zero bias) with interval 11 (200 °C and zero bias) ammonia has the lowest signal of the three gases in 2 but the highest in 11. While temperature does have a strong influence on the sensor response this apparent reversal of the sensor response to different gases is due to the pre-processing (here: standardization). In addition, the sensor response is also influenced by the gate bias which is evident when comparing interval 2 with intervals 3 (negative bias at 260 °C) and 12 (negative bias at 200 °C).

3.2. Feature extraction

Evaluation of the obtained multi-dimensional sensor raw data, each individual cycle with 900 data points, is performed by pattern recognition tools. In order to reduce the dimension of the data matrix, significant features for determining gas type and concentration have to be identified and used for the discrimination later on [23]. Previous studies [7, 14, 20] have shown that in particular features describing the shape of the sensor response e.g. slopes (here: secants, sec) and mean values (mv) - so called standard features - are very effective for discrimination of different gases. In fact it was reported that these features outperform features typically extracted from periodic signals, i.e. Fourier and Wavelet transformation coefficients [24].

Since the aim of this work is to show that a combination of temperature cycled and gate bias cycled operation can enhance the selectivity of the sensor, the selected cycle (cf. Fig. 2a) is divided into several intervals in which the standard features are extracted. Figure 2b shows the standardized dynamic sensor response (one cycle per gas) where 18 intervals for feature extraction are marked. The combined cycle was divided into four relevant parts reflecting temperature (A, C) and gate bias changes (B, D). Each temperature part is divided into two segments resulting in features 1, 2 and 10, 11. The gate bias cycles at constant temperature are divided into seven segments each resulting in features 3 to 9 and 12 to 18.

The sensor response towards the different test gases (CO, NO₂ and NH₃) is small compared to the changes induced by temperature and bias (cf. Fig. 2b) and the change in the sensor response towards different concentrations of a test gas is even smaller. However, there is still a significant difference in the sensor response for each gas and concentration which is represented by the extracted features. Figure 4 shows exemplarily some mean values (mv) and their change due to exposure to the test gases (breaks between different gas concentrations are not shown as they are not used for evaluation). It can be observed that e.g. NH₃ identification

can be based on mv 2 (part A of the combined cycle), but this feature does not reflect changes in the concentration. Instead, NH_3 quantification can be based on mv 11 (part C) which, however, also shows a reaction for CO and NO_2 . Similarly, mv 17 (part D) provides information on different CO concentrations. Thus, combined discrimination and quantification of the three test gases with four concentrations each is not possible using individual features. Thus, pattern recognition tools like LDA are required to identify suitable (linear) combinations of significant features.

3.3. Linear Discriminant Analysis

The extracted features are used for pattern recognition. Linear Discriminant Analysis (LDA) [21, 23, 25, 26] was used as a powerful tool to evaluate multidimensional data. LDA is a supervised learning methods meaning that the correct classification is known for each object (here: cycle). The aim of the algorithm is to maximize the ratio (Γ : discrimination criterion) of the between-class scatter B (sum of the squared distances between the various classes) to the within-class scatter W (sum of squared distances of the group elements from the class mean):

$$\Gamma = \frac{B}{W} = \max \quad (3)$$

The eigenvectors v of the resulting eigenvalue problem

$$Bv = \lambda Wv \quad (4)$$

are the coefficients $c_{k,i}$ of the discriminant functions DF which are basically a linear combination of the extracted features x_i . The eigenvectors are sorted in descending order according to their corresponding eigenvalues λ_k .

$$DF_k = \sum_i c_{k,i} \cdot x_i = c_{k,0} + c_{k,1}x_1 + c_{k,2}x_2 + \dots + c_{k,m}x_m \quad (5)$$

The dimension of the space spanned by the DFs is $K = \min\{n - 1; m\}$ with n number of groups and m number of features. For the combined cycles (cf. Fig. 2a) $m = 36$ different features were extracted as shown in table 1 and the tested gas profile includes three gases (CO , NO_2 and NH_3) with four concentration each. Usually, m is much larger than n so that $n-1$ discriminant functions (linear combinations) are determined achieving a dimensionality reduction. Often, the first two discriminant functions represent more than 95 % of the total information contained in the data set. In this aspect, LDA is similar to the often used Principal Component Analysis (PCA) [21]. However, PCA is a non-supervised technique, i.e. the

information on the correct classification, which is available when using calibration data obtained in the lab, is not used for the dimensionality reduction. Further details and a comparison of both methods can be found elsewhere [23, 27, 28].

3.4. Feature Selection

In order to select features and therewith reduce the dimension further, a measure of quality of the discrimination is necessary to proof that the selected features are indeed significant for the separation. A measure for the quality of the discrimination is the multivariate Wilks' Lambda

$$\Lambda = \prod_{k=1}^K \frac{1}{1+\lambda_k} \quad (6)$$

Where K is the number of theoretically possible discriminant functions and λ_k is the eigenvalue corresponding to the discriminant function.

Since the first two discriminant functions usually represent more than 95 % of the contained information a more suitable definition for the multivariate Wilks' Lambda is

$$\bar{\Lambda} = \prod_{k=1}^2 \frac{1}{1+\lambda_k} \quad (7)$$

Although Wilks' Lambda is scaled to 0 to 1 it is often difficult to interpret as it does not give an absolute measure of quality (because it depends on the specific problem, for example on the number of classes). Therefore, cross-validation (cf. section 3.5) is a more rigorous method for evaluating the performance of data processing.

Feature selection can be performed in various ways both before and after solving the eigenvalue problem. One approach for selecting the most significant features before solving the eigenvalue problem is based on the univariate Wilks' Lambda which is scaled from zero to one. Small values of Wilks' Lambda represent significant features.

$$\Lambda_i = \frac{W_{ii}}{B_{ii}+W_{ii}} = \frac{\text{diag}_i(\mathbf{W})}{\text{diag}_i(\mathbf{B})+\text{diag}_i(\mathbf{W})} \quad (8)$$

With Λ_i the Wilks' Lambda corresponding to feature i . Feature selection based on this parameter can be automated in a so-called stepwise LDA method with the F-value criteria which is a combination of both the forward selection and backward elimination techniques in order to find the most important features [25]. However, this is a kind of *black-box approach* and the results are sometimes not easy to comprehend and interpret.

Feature selection after solving the eigenvalue problem has the drawback that in a first step the calculation is based on all extracted features which can lead to over-fitting problems when

using small data sets. However, the selection is then easier and can be based on the average coefficient $\bar{c}_i$, respectively the loadings plot.

$$\bar{c}_i = \sum_{k=1}^K |c_{k,i}^*| \cdot \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_K} \quad (9)$$

Here, $c_{k,i}^*$ is the standardized coefficient of discriminant function k and feature i .

In general, feature selection is often an iterative process to identify the best features. In addition to the discrimination performance, stability can also be taken into account.

3.5. Validation

Cross-validation is an important step in the signal processing in order to proof the obtained results. When working with pattern recognition algorithms, in particular with supervised methods like LDA, the algorithms tend to over-fit (over estimate) the data. There are several different algorithms to validate the results [23]. The most well-known validation algorithms are *leave-one-out* and *k-fold cross validation*, *hold-out* and *bootstrapping*. The basic idea of the hold-out method is to split the data set in two parts: one for training and one for evaluation. The obtained LDA coefficients from the training set are then used to project the “unknown” data from the second part (evaluation data) into the LDA space with the training data. K-fold cross-validation on the other hand not only splits the data set in two parts but rather in k parts which are successively used as evaluation data set with the rest of the data used for training. If k is equal to the number of data points (here: cycles) the algorithm is identical to leave-one-out cross validation, which basically checks if each feature vector is correctly classified if the data evaluation is trained with all vectors. For all validation algorithms a classifier is required, e.g. k-nearest-neighbors (kNN) or Mahalanobis distance [25] which classify new vectors by comparison with the given training data. The classification rate is the result of the validation and can be also be used as a measure of quality (performance) of the discrimination. Additionally, validation can be used to study the stability of the discrimination to e.g. drift and noise.

4. Results

4.1. Quantification of carbon monoxide

The influence of the different features on the discrimination is studied using the example of quantification of carbon monoxide. In a first step, only features representing a temperature change, i.e. features 1, 2, 10 and 11, (compare Fig. 2b) are used for calculating the LDA. The result is shown in Fig. 5a. Discrimination is possible; however, the groups are quite close to each other which can lead to misclassification of unknown data with the obtained LDA coefficients. Similar results are obtained when using features from part D of the combined cycle only, i.e. features depending only on gate bias changes at 200 °C. However, the result of the discrimination can be enhanced significantly when combining both types of features. Especially the separation of the two highest concentrations, i.e. 600 and 800 ppm, is significantly better. All three LDA show increasing concentration mainly along the first discriminant function (more than 99 % of the information is represented by DF1). On the other hand, the second discriminant function represents only sensor drift and is not used for separation.

Combination of both operating modes, i.e. using all features extracted from the combined cycle, therefore reduces the uncertainty for determining the CO concentration.

4.2. Discrimination of carbon monoxide, ammonia and nitrogen dioxide

Figure 6a shows the result of the discrimination of carbon monoxide, ammonia and nitrogen dioxide with four concentrations each. For the calculation all extracted feature have been used. Discrimination of the three gases independent of their concentration is easily possible, even though the first discriminant function mainly separates ammonia from the other gases which is due to the fact that the sensor has the highest sensitivity towards ammonia. The location of the groups in the scatterplot, however, is hard to interpret. LDA as a supervised learning algorithm, i.e. correct classification is known for each object aims to maximize the distance between different groups while minimizing the scatter within each group thus optimizing group separation. The scale of the LDA plot reflects the standard deviation, i.e. noise, of the data along the discriminant functions; thus, high values indicate a good signal to noise ratio. Additionally, the same LDA plot also allows direct determination of the concentration for carbon monoxide and ammonia. Leave-one-out cross-validation shows that the gas type is correctly identified in all cases and the concentration is classified correctly for almost 95 % of the data points. Only the groups for different concentrations of nitrogen

dioxide overlap and have a lower cross-validation rate. Although the dimension of the data set is reduced by feature extraction, two features each (mean value and slope) are calculated for 18 intervals resulting in a 36-dimensional feature space with a high level of redundancy and also irrelevant features. To reduce the number of features, feature selection was performed based on the loadings plot or using the stepwise LDA algorithm as described in section 3.4.

Using the loadings plot is more intuitive and also provides information on how each feature contributes to the discrimination. The loadings plot for the discrimination of the three gases is shown in Fig. 6b (only features with an average coefficient greater than 0.1 are shown). When comparing the scatterplot in Fig. 6a with the loadings plot in Fig. 6b it can be observed that features from intervals with high gate bias (mv 5, 6, 7, 15, 16, 17) contribute mainly to the separation of the three test gases, whereas features from the negative bias regime (mv 3 and 18) contribute to the quantification of NO₂. Quantification of CO can mainly be based on mv 17 (200 °C and slightly positive bias). Feature mv 2 (260 °C, zero bias) separates ammonia from the other gases and quantification can then be based on mv 11 (200 °C and zero bias). Based on the loadings plot, ten features were manually selected (cf. Table 1) and used for calculating a modified LDA. It can be observed that especially features at high gate bias values give the best discrimination performance but also features from the temperature cycle (part A and C, mv 2, mv 11 and sec 1) are important. This again proves the concept of a combined temperature and gate bias cycle. In general, slopes are less significant than mean values. Projecting the data with the modified LDA based on the manually selected ten features leads to a similar result as when using all 36 features (compare Fig. 6a and Fig. 7a). Although the groups are closer to each other and show more scatter, especially for carbon monoxide at 200, 400 and 600 ppm, leave-one-out cross-validation still shows a very good discrimination.

Note that most of the discriminating power (>98 %) is represented by the first discriminant function (DF1) which is also indicated by its significantly larger scale compared to DF2. The first discriminant function primarily separates ammonia due to the fact that the sensor is more sensitive to ammonia (see also Fig. 4) than to the other two gases. Nevertheless, discrimination of all three gases is possible by only evaluating the DF1 value even though this is hardly visible on the large scale in Fig. 6a. However, discrimination is more stable when additionally using the second discriminant function.

Using the stepwise linear discriminant analysis based on the (univariate) Wilks' Lambda, respectively the F-value, to identify the most significant features [25, 26], more features are selected (cf. Table 1). The discrimination performance remains very high (Fig. 7b) and is

comparable with the LDA based on features selected from the loadings plot (Fig. 7a). Again, features from all parts of the combined cycle have been selected proving the improved performance of the combined TC and GBC approach.

4.3. Stability/Repeatability

In order to validate the results of separation of carbon monoxide, nitrogen dioxide and ammonia, an LDA was calculated where each group of the mentioned gases contains four different concentrations: carbon monoxide 200, 400, 600 and 800 ppm, nitrogen dioxide 20, 30, 40 and 50 ppm and ammonia 30, 45, 60 and 75 ppm. The result is depicted in Fig. 8 with the training data marked with open symbols. Features from all parts of the combined cycled were used for determination of this LDA. In a second step, the obtained LDA coefficients are applied to project “unknown” (semi-solid symbols in Fig. 8) data from a measurement which was performed a few days after the initial (training) measurement. The “unknown data” are CO (800 ppm), NO₂ (50 ppm) and NH₃ (75 ppm). Fig. 8 shows that the unknown data groups are located fairly close to the corresponding training group. Only a small drift along DF2 (positive for background, NO₂ and CO, negative for NH₃) is evident which can be compensated by rotating the entire coordinate system in counter clockwise direction, based on a single re-calibration measurement with one gas [29]. However, even without drift compensation all unknown data points of the test gases are classified correctly. The classification rate of unknown data points from the carrier gas group is only 38 % which is due to sensor drift.

Nevertheless, the sensor drift can be compensated when not only one measurement series is used for training but a combination of two (or more) measurements which were over a certain period of time. Again, each group of the mentioned gases contains four concentrations from two different measurements: carbon monoxide 200, 400, 600 and 800 ppm, nitrogen dioxide 20, 30, 40 and 50 ppm and ammonia 30, 45, 60 and 75 ppm. A third data set is then used for evaluation, respectively projection of “unknown data”. Figure 9 shows the separation of the training groups marked with open symbols (for each group/gas four concentrations were used) and the projection of the “unknown” groups marked with semi-solid symbols. Evidently, the projected data groups are nearly ideally overlapping with their corresponding training groups. A Mahalanobis distance classifier correctly identifies each “unknown” data point (here: cycle) incl. the data for background resulting in a perfect classification rate of 100 %. This proves that sensor drift can be effectively compensated by using larger data sets for training, in

particular a combination of two or more data sets which were recorded at different points in time and hence, include sensor drift. Thus, classification of new (“unknown”) data is much more stable.

5. Conclusion and Outlook

We have shown that by combining temperature cycled operation (TCO) and gate bias cycled operation (GBCO) the selectivity of gas sensitive SiC-FETs can be enhanced further. Quantification of gases is possible with each method; however the resolution is improved when using both methods together. Identification of carbon monoxide, ammonia and nitrogen dioxide is possible independent of their concentration over a wide range. Different feature selection methods have proven that features obtained from both methods are important for the identification. In addition to an evaluation of the performance using leave-one-out cross-validation, the stability was proven by projecting data from a different measurement performed several days later into the LDA plot obtained from the original training data. Correct classification is achieved although a small systematic drift is evident for all gases. However, the drift can be compensated when using two data sets for training so that the drift of the sensor is included in the training.

The long-term stability of sensors operated in combined TCO and GBCO mode needs to be investigated further as well as the optimum design of the cycles to achieve the best performance with respect to selectivity and stability. Additionally, future work will address the mechanism occurring when changing the bias in order to better understand and further enhance the improved selectivity. Temperature modulation is an effective method since the operating temperature of the sensor in combination with the catalytic material influences the gas interaction with the sensor surface and chemical reactions, respectively, and thereby the selectivity of the sensor. However, the effect of changing the gate bias is as yet less well understood and thus the reason for the enhanced selectivity remains unclear. It was reported [12, 13] that the hysteresis of the sensors when changing the bias is caused by loading and unloading of trapping states in the insulator (here: SiO₂). However, this is probably not the dominant cause for the observed sensor behavior, as changes in the trapping states show much longer time constants [11]. Since platinum was used as a catalytic layer in these experiments we think that platinum oxide is partly created, i.e. the balance reaction between platinum to platinum oxide in oxygen rich atmospheres is influenced by the applied bias. This phenomenon was previously observed for yttria-stabilized-zirconia (YSZ) electrodes when applying a bias [30]. We also observed that gate bias changes lead to additional stress for the sensor. The sensors required long run-in times in GBCO before the sensor baseline stabilizes in air, especially for high gate bias values. After a few hours the bias cycling did no longer affect the stability of the baseline markedly.

However, the overall mechanism taking place when applying a gate bias to a gas sensitive field effect transistor needs to be investigated in greater detail, in order to understand both short-term and long-term effects.

Acknowledgment

The authors would like to thank SenSiC AB, Kista, Sweden, for providing the sensors and 3S - Sensors, Signal Processing, Systems GmbH, Saarbrücken, Germany, for providing the hardware for sensor operation and read-out.

Table 1 Intervals of the combined temperature and gate bias cycle with description and features selected based on the loadings plot and on the stepwise LDA algorithm.

Interval	Description	features	Features selected based on	
			loadings plot	stepwise LDA
1	Part A, temperature change (200 → 260 °C), zero bias	Mean value (mv)		x
		Secant (sec)	x	
2	Part A, temperature plateau (260 °C), zero bias	Mean value (mv)	x	x
		Secant (sec)		
3	Part B, bias plateau (-1 V) at 260 °C	Mean value (mv)	x	x
		Secant (sec)		
4	Part B, bias ramp from -1 to +1 V at 260 °C	Mean value (mv)		
		Secant (sec)		
5	Part B, bias ramp from +1 to +3 V at 260 °C	Mean value (mv)	x	x
		Secant (sec)		
6	Part B, bias plateau (+3 V) at 260 °C	Mean value (mv)		x
		Secant (sec)		
7	Part B, bias ramp from +3 to +1 V at 260 °C	Mean value (mv)	x	x
		Secant (sec)		
8	Part B, bias ramp from +1 to -1 V at 260 °C	Mean value (mv)		
		Secant (sec)		
9	Part B, bias plateau (-1 V) at 260 °C	Mean value (mv)		x
		Secant (sec)		
10	Part C, temperature change (260 → 200 °C), zero bias	Mean value (mv)		x
		Secant (sec)		
11	Part C, temperature plateau (200 °C), zero bias	Mean value (mv)	x	x
		Secant (sec)		
12	Part D, bias plateau (-1 V) at 200 °C	Mean value (mv)		x
		Secant (sec)		
13	Part D, bias ramp from -1 to +1 V at 200 °C	Mean value (mv)		x
		Secant (sec)		x
14	Part D, bias ramp from +1 to +3 V at 200 °C	Mean value (mv)	x	x
		Secant (sec)		
15	Part D, bias plateau (+3 V) at 200 °C	Mean value (mv)		x
		Secant (sec)		
16	Part D, bias ramp from +3 to +1 V at 200 °C	Mean value (mv)	x	x
		Secant (sec)		

17	Part D, bias ramp from +1 to -1 V at 200 °C	Mean value (mv)	x	x
		Secant (sec)		
18	Part D, bias plateau (-1 V) at 200 °C	Mean value (mv)	x	x
		Secant (sec)

Fig. 1

(a) Schematic cross-sectional view of the used depletion type SiC-FET, (b) photo of the SiC-FET die glued on a ceramic heater and mounted on a 16 pin TO-8 header together with a Pt-100 temperature sensor (c) top view of the sensor die holding two two-terminal devices (1) and two three-terminal devices (2) as well as a temperature sensor (3, not used in this work).

Fig. 2

(a) Combination of temperature cycle (parts A and C) consisting of two plateaus and gate bias cycle combining plateaus and ramps (parts B and D). (b) Standardized dynamic sensor response of the sensor is shown. In different colors the responses for carbon monoxide, nitrogen dioxide and ammonia are shown and different intervals for feature extraction are marked.

Fig. 3

Signal processing flow chart showing the applied signal processing steps for data evaluation.

Fig. 4

Exemplary features (here: mean values, mv) and their change over time due to exposure to the different test gas and different concentrations.

Fig. 5

LDA showing the quantification of carbon monoxide using features from different parts of the combined cycle. The results have been validated by leave-one-out cross-validation.

Fig. 6a

LDA showing the discrimination of CO, NO₂ and NH₃ with four concentrations each. For the calculation features from all parts of the combined cycle were used.

Fig. 6b

Loadings plot showing the significance of each feature (only features with average coefficient greater than 0.1 are shown; “mv”: mean values, “sec”: secants).

Fig. 7a

LDA showing the discrimination of CO, NO₂ and NH₃ with four concentrations each. For the calculation only features selected from the loadings plot were used (compare Table 1).

Fig. 7b

LDA showing the discrimination of CO, NO₂ and NH₃ with four concentrations each. For the calculation only features selected by the stepwise algorithm were used (compare Table 1).

Fig. 8

LDA showing the discrimination of CO, NO₂ and NH₃ independent of concentration. Training data are marked with open symbols. Each group consists of four concentrations of the corresponding gas. Test data (one concentration per gas) of a second measurement performed a few days after training are marked with semi-solid symbols.

Fig. 9

LDA showing the discrimination of CO, NO_2 and NH_3 independent of concentration. Training data are a combination of two measurements, marked with open symbols. Test data of a third measurement performed a few days after training are marked with semi-solid symbols.