

REVIEW ARTICLE

Theory and application of the material work function for chemical sensors based on the field effect principle

P Bergveld, J Hendrikse and W Olthuis

MESA Research Institute, University of Twente, Box 217, 7500 AE Enschede, The Netherlands

Received 1 May 1998, in final form and accepted for publication 21 August 1998

Abstract. This paper gives first of all the definition of the work function of an electronic as well as an ionic conductor. FET-type gas sensors known from the literature are considered in view of this basic theory and the parameter of the work function which is responsible for the sensing properties is noticed. This appears not always to be possible for the various types of gas sensors, in this case the gas FET, the SGFET and the IGFET. In contrast to this ambiguity, the sensing parameters of the recently developed E MOSFET, meant for application in an electrolyte, can clearly be identified. As an example an oxygen sensor based on the developed E MOSFET is described.

Keywords: redox potential, work function, field effect transistor (FET), environmental monitoring, process control, biomedical measurements

1. Introduction

In order to understand the possibilities of using the work function of a material as the parameter that can be modulated by external species to be measured from a sample, the work function should be defined properly. This can be done in relation to the Fermi energy of a solid or the electrochemical potential of a liquid. The analysis thus should start with the definition of the Fermi energy and the derivation of the solid-state interface potential between two solids that are in contact with each other, followed by the equivalence concerning the electrochemical potential of a liquid and the liquid/liquid interface potential.

1.1. The Fermi energy and the contact potential

Beside the mathematical description of the Fermi energy introduced by Fermi and Dirac, a more physical description concerns the experimental consideration of moving a unit charge (one electron) in vacuum from infinity to a solid phase surface. In this consideration the electrical potential at infinity in vacuum is assumed to be zero. Furthermore, one makes the assumption that, at a distance of approximately 10 nm from the solid's surface, the solid just does not influence the moving charge. The electrical potential at this small distance from the surface is called the Volta potential θ and consequently the energy of the unit charge has increased by an energy $E = -q\theta$, the Volta

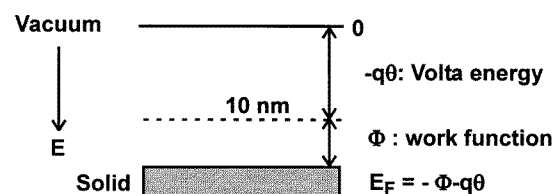


Figure 1. Definitions of energy levels and differences at a solid/vacuum interface.

energy. In the solid phase the mean electron energy is defined as the Fermi energy E_F and the difference between the Volta energy and the Fermi energy is defined as the work function Φ , a material constant for a certain solid. Consequently

$$E_F = -\Phi - q\theta. \quad (1)$$

This is schematically illustrated in figure 1. Note that a solid material is thus defined by its work function Φ rather than by its Fermi energy, because E_F can be influenced by external means, such as a connected voltage source with respect to 0 V.

When two electronic conductors, 1 and 2, for instance two metals, are connected (making contact) the electrons will distribute themselves in such a way that equilibrium is established, which means that

$$E_{F1} = E_{F2}. \quad (2)$$

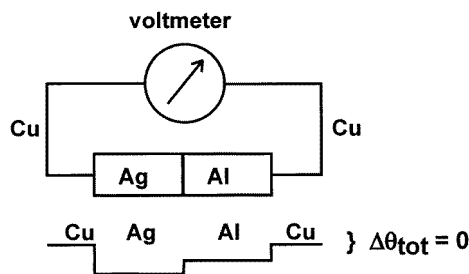


Figure 2. Direct measurement of the contact potential, resulting in $\Delta\theta_{tot} = 0$.

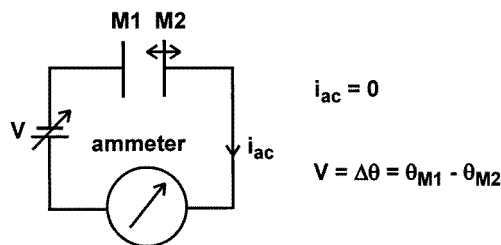


Figure 3. The principle of the Kelvin probe.

The result is, using equation (1), an interfacial contact potential

$$\theta_1 - \theta_2 = \Delta\theta = \frac{1}{q}(\Phi_2 - \Phi_1). \quad (3)$$

An interesting question is that of how contact potentials can be measured in an experimental set-up and from this difference in work functions. Figure 2 shows how, in a direct measurement, necessary additional contact potentials will always exist, resulting ultimately in a measured voltage of zero.

Note that the inability to measure contact potentials between two metals in a direct way is in fact the result of the fundamental source of the contact potential, namely a redistribution of electrons. From this purely physical phenomenon the energy necessary for the measurement can never be withdrawn. Only if external energy can be supplied can a contact potential be measured, which is for instance the case with a thermocouple.

Another way of energy supply occurs in the so-called Kelvin probe procedure shown in figure 3. In this case a plate of metal M_2 is vibrating with respect to a fixed plate of metal M_1 . Owing to the difference in contact potential between M_1 and M_2 the capacitor formed by the two plates will be charged and discharged with the frequency of the vibration. The resulting alternating current i_{ac} will become zero as soon as the dc series voltage is equal to the contact potential θ_{M_1/M_2} but with opposite sign. Under this condition $V = -\theta_{M_1/M_2}$. The energy for the measurement is supplied by the mechanically induced vibration. The measurement set-up is also called a vibrating reed system.

A more modern alternative which does not need external energy in the measurement circuit makes use of the field-effect concept on which a MOSFET device is based. The gate input of a MOSFET (schematically shown in figure 4) does not take any input current in the equilibrium state (the stable dc bias condition).

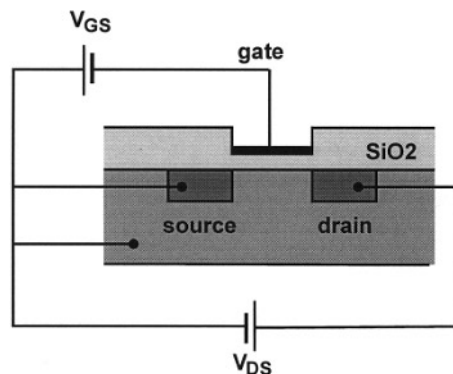


Figure 4. A representation of a MOSFET and supplied electrical voltages.

The dc drain current, I_D , of a MOSFET is determined not only by the externally applied voltages V_{GS} and V_{DS} and the sensitivity parameter β determined by the gate's dimensions[†] but also by the threshold voltage V_T , as can be seen in equation (4):

$$I_D = \beta(V_{GS} - V_T - \frac{1}{2}V_{DS})V_{DS}. \quad (4)$$

The value of V_T is determined, among other factors, by the difference between the work functions of the gate material and the silicon bulk:

$$V_T = \frac{\Phi_{gate}}{q} - \frac{\Phi_{Si}}{q} - \text{constant}. \quad (5)$$

Using two MOSFETs with different gate materials, M_1 and M_2 , in a differential measurement set-up thus gives the contact potential between the two gate materials:

$$\Delta V_T = \frac{\Phi_{M_1}}{q} - \frac{\Phi_{M_2}}{q} = \Delta\theta_{M_1/M_2} \quad (6)$$

assuming that the constant terms for the two MOSFETs are equal. Note that, because the absolute value of the constant term in equation (5) depends on process parameters and thus is unknown, the work function of a gate material with respect to that of silicon cannot be determined in this way.

1.2. The electrochemical potential and the liquid/liquid potential

For the description of interfaces of which one or both sides consist of a liquid, some basic knowledge of electrochemistry might be necessary but it can also be described in an analogous way to that given above. Also for a liquid phase one can consider a charge, being moved in vacuum from infinity to about 10 nm outside the liquid's surface. In this case the charge consists, however, of 1 mole of ions (Avogadro's number $N_A = 6.02 \times 10^{23} \text{ mol}^{-1}$) having a charge of F coulombs (Faraday's constant = qN_A) possibly with a higher charge per ion than the unit q ,

[†] The sensitivity parameter $\beta = \mu C_{ox} W/L$, where μ is the mobility of electrons in the channel, C_{ox} the gate oxide capacitance per unit area, W the channel's width and L the channel's length.

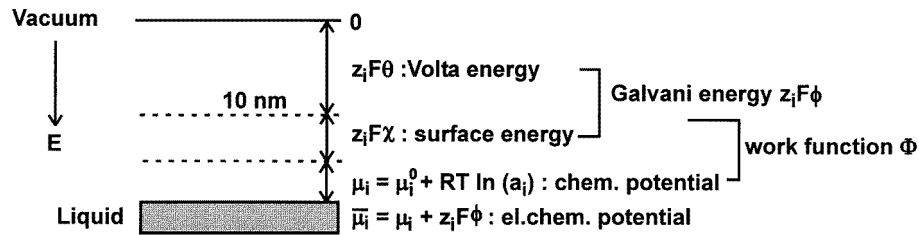


Figure 5. Definitions of energy levels and differences at a liquid/vacuum interface in analogy with figure 1.

due to the valence z_i , the charge on an ion in signed units of electronic charge. Therefore the Volta energy is in this case $z_i F \theta$, with θ again the Volta potential. See figure 5.

The work function Φ is in this case divided into two components. In the first place energy of electrostatic nature has to be gained, due to the orientation of molecular dipoles. This component is called the surface energy $z_i F \chi$ with χ the surface potential. The sum of the Volta potential and the surface potential is called the Galvani potential:

$$\phi = \theta + \chi. \quad (7)$$

Note that, in the previous section, the χ term is omitted, which is true for a metal, as described there, but not for all electronic conductors. For instance in the case of a semiconductor the χ term has to be added, where $q\chi$ is called the electron affinity, of which the value depends on the type and level of the doping.

Entering the liquid phase with the test charge will lead to a reorientation of the liquid's structure. This interaction has a chemical nature and the corresponding energy for the 1 mol test charge is called the chemical potential of the liquid μ_i . Thus

$$\Phi = \mu_i + z_i F \chi. \quad (8)$$

The chemical potential is a function of the concentration of the species involved, according to

$$\mu_i = \mu_i^\circ + RT \ln(f_i c_i) = \mu_i^\circ + RT \ln a_i \quad (9)$$

where R is the gas constant ($R = kN_A$, with k the Boltzmann constant), T the absolute temperature and a_i the activity of the particular ions. The relation between the concentration c_i and the activity a_i is given by $a_i = f_i c_i$, with f_i the activity coefficient. In dilute electrolyte solutions $f_i = 1$, but in more concentrated solutions $f_i < 1$. μ_i° is the standard chemical potential for $a_i = 1$.

The total energy required in order to move the 1 mol test charge in vacuum from infinity to inside the liquid is thus

$$\mu_i + z_i F \phi = \bar{\mu}_i \quad (10)$$

where $\bar{\mu}_i$ is called the electrochemical potential, which is completely analogous to the Fermi energy described by equation (1) and elucidated in figure 1.

If two electrolyte solutions are brought into contact in such a way that they cannot mix, for instance due to a mechanical obstruction like a membrane which is permeable to small ions only, the electrochemical potentials

$\bar{\mu}_i$ of the ions i will become equal ($\bar{\mu}_{i1} = \bar{\mu}_{i2}$), resulting in an interfacial potential

$$\phi_1 - \phi_2 = \frac{RT}{z_i F} \ln \left(\frac{a_{i2}}{a_{i1}} \right) = \frac{RT}{z_i F} \ln \left(\frac{f_2 c_{i2}}{f_1 c_{i1}} \right) \quad (11)$$

(the standard chemical potentials of the two are equal ($\mu_{i1}^\circ = \mu_{i2}^\circ$)).

Note that equation (11), basically the Nernst equation, forms the basis of many potentiometric sensors such as the ISE and the ISFET. In that case one of the concentrations, for instance c_{i1} , is kept constant in a membrane (ISE) or at a surface (ISFET), resulting in 59 mV/dec at room temperature for the monovalent ions to which the sensor has been made selective.

It should be kept in mind, however, that the origin of equation (11) comes from the thermodynamic equilibrium across the phases on both sides of the interface and the work function is

$$\Phi = \mu_i^\circ + RT \ln(f_i c_i) + z_i F \chi \quad (12)$$

according to equations (8) and (9). From equation (12) it can be seen that the work function depends not only on $\ln c_i$ but also on $\ln f_i$ and χ . In cases in which the latter values are also functions of c_i , a deviation from 59 mV/dec will be observed and in cases in which f_i and χ are functions of another external parameter interference will be experienced. On the other hand the dependences of f_i and χ on an external parameter might also be exploited as possibilities for controlling the work function Φ , leading to a new sensing concept, namely work-function modulation. Note that this can be established either by bulk modulation (the first two terms of equation (12)) or by the surface energy (the last term in equation (12)). For bulk modulation it is essential that species to be measured should penetrate the whole interior of the material, whereas for surface modulation only the surface of a sensing material need be reached. Because work functions can most easily be measured by means of a MOSFET structure (see figure 4), it will be no surprise that sensors based on the concept of work-function modulation will all be of the FET type.

2. The gasFET

Shortly after the first publication on ISFETs [1], operating according to equation (11), Lundström *et al* [2] published the first results on the hydrogen sensitivity of a FET device having a palladium gate. Actually the device is a normal

MOSFET with the usual aluminium gate illustrated in figure 4 replaced by a palladium gate. For a measurement the gate should be connected to the source plus bulk of the device. Palladium was chosen because of its catalytic properties and its permeability to H atoms. Hydrogen in the ambient dissociatively adsorbs at the outer Pd surface and the resulting hydrogen atoms diffuse into the thin layer. This process occurs relatively fast at elevated temperatures (about 150 °C). Lundström *et al* showed that the amount of hydrogen inside the palladium is too low to influence the bulk part of its work function, but that on the other hand the hydrogen atoms are adsorbed at the Pd/SiO₂ interface, forming a dipole layer that thus influences the surface energy part $q\chi$ of the work function (see equation (12)). The magnitude of the modulation depends on the coverage θ with H atoms of the available number of adsorption sites per unit area at the Pd/SiO₂ interface. The observed modulation of the threshold voltage ΔV_T of the Pd-gate MOSFET was modelled with respect to the maximum achievable shift of the dipole-layer voltage $\Delta\chi_{max}$ at maximum coverage:

$$\Delta V_T \approx \Delta\chi \approx \Delta\chi_{max}\theta \quad (13)$$

where the coverage $0 < \theta < 1$ was modelled as a function of the partial hydrogen pressure of the ambient, P_{H_2} , according to

$$\theta = \frac{\alpha\sqrt{P_{H_2}}}{1 + \alpha\sqrt{P_{H_2}}} \quad (14)$$

where α is an overall equilibrium constant which appears to be oxygen and temperature dependent.

In principle all gases that are able to dissociate at the Pd surface, like H₂, H₂O, H₂S and a number of hydrocarbons, will result in a transfer of H atoms to the Pd/SiO₂ interface and will thus interfere with the original hydrogen sensitivity. Because the dehydrogenation is in general temperature dependent with different ‘threshold’ temperatures for different gases, some selectivity can be achieved by applying different operational temperatures.

Note that the principle of modulation of the surface part of the work function at the gate material/insulator interface of a MOSFET can be exploited only in cases in which the gate material is electronically conducting and penetrable by the species to be measured. Up to now the only material which combines these two requirements is palladium with respect to hydrogen-bearing species. In an attempt to use the same principle for other gases, MOSFET devices have been constructed with a discontinuous gate: figure 4 with (small) holes or slits in the gate which consists in this case of a catalytic metal such as Pt. In this way Dobos *et al* [3] illustrated a gasFET sensitive to CO. If the gas in the holes can be considered to be an extension of the gate insulator, the metal/gas interface can thus be seen as the Pd/SiO₂ interface in the original gasFET of Lundström and the operational mechanism is thus again the modulation of the surface part of the work function of the gate metal. Optimization of this construction is rather difficult and therefore a much simpler and more straightforward approach is to use a suspended gate, of which the elaboration will be described in the next section.

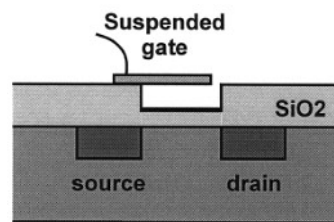


Figure 6. A schematic representation of a MOSFET with a suspended gate.

3. The suspended gate FET

In the suspended gate FET (SGFET), first described by Blackburn *et al* [4], the gate metal is suspended above the gate insulator (SiO₂) by a narrow gap with access holes for a gas, as schematically shown in figure 6. This construction makes it possible to develop gas sensors sensitive to other gases than hydrogen because the metal/gaseous-insulator interface is now freely accessible to the gas instead of the gas having to diffuse through the gate material as in the case of the palladium gate gasFET.

The electrical sensitivity of the FET structure is less than that of a MOSFET with only SiO₂ as a dielectric layer (figure 4), because the parameter β (see equation (4)) is smaller due to the lower gate capacitance of the SiO₂/gas insulator sandwich. In principle different permittivities of different gases may modulate the drain current (equation (4)) beside possible dipole-layer formation at the gate metal/gas and gas/SiO₂ interfaces. It could be argued that the main effect is caused by the modulation of the surface energy at the metal/gas interface due to an adsorption process, which means that the gas species itself need not have a dipole moment. (H has no dipole moment.)

The first SGFET devices contained a Pt suspended gate and appeared to be sensitive to methanol and methanol chloride vapours [4]. In order to attain selectivity and extend the number of gases to which the device should be sensitive, the Pt gate was coated with a polymer, such as polypyrrole, in later constructions (Fig. 6 with a thin layer of polypyrrole around the suspended Pt gate). Indeed, devices with electrodeposited polypyrrole on the Pt gate (decreasing also the gap distance as a secondary effect) appeared to be sensitive to a number of other gases, such as ethanol, 1-propanol and 1-butanol [5]. Laser-induced chemical vapour deposition of polypyridine on the suspended Pt gate resulted in a reversible sensitivity to propylamine vapour [6].

The explanation of the above-mentioned results of polymer-coated suspended Pt-gate FETs is more difficult than that of the bare platinum devices. In the latter case it will be clear that only the surface part of the Pt work function can be modulated, whereas, in the case of a polymer coating, the surface as well as the bulk part of the polymer work function may be modulated because polymers are permeable to gases. The gas may oxidize or reduce the polymer as a whole, thus affecting surely the work function of its bulk. Unfortunately, from a single measurement no information regarding whether the work function's modulation is caused by a change in the bulk

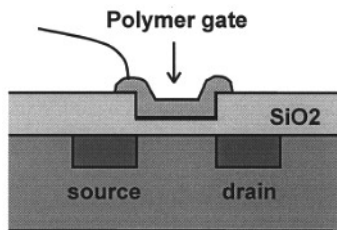


Figure 7. A schematic representation of an IGFET with an organic semiconductor gate.

part, a change in the surface part or both can be obtained. The observed slope of the V_T /methanol concentration of 60 mV/dec may indicate that the bulk term is involved. However, in cases in which the adsorption is logarithmic such a sensitivity would also result from a purely surface effect. Anyway the fact that the (overall) sensitivity appears to depend on the chemical nature of the original monomer as well as on the deposition parameters opens the way to the development of a large variety of gas sensors with different sensitivities and selectivities.

4. The insulated gate FET

For the actual operation of the SGFET it makes no difference whether the polymer is deposited onto the suspended gate material or onto the oxide of the device. Sensors based on this approach are called IGFETs, a name that is a little confusing because it has been used in the past also for the normal MOSFET. In cases in which the polymer is electrically conducting, or, in other words, an organic semiconductor, it can serve as the gate of the FET structure, so a suspended gate is no longer necessary. The resulting device is schematically represented in figure 7.

Because the gate material is permeable to gases the device can operate like the Pd MOSFET by the formation of a surface dipole layer at the polymer/SiO₂ interface, thus modulating the surface-energy part of the work function of the gate material. However, contrary to the Pd MOSFET, also the bulk term of the work function may be modulated due to the possible formation of charge-transfer complexes between the polymer and the gas molecules. Such a partial transfer of electrons is in fact a redox mechanism. If electrons are transferred to the gas molecules, the polymer is reduced; if electrons are transferred to the polymer, it is oxidized. Janata [7] described this phenomenon in terms of donor and acceptor behaviours respectively of the polymer in analogy with the usual semiconductor description in which donor and acceptor concentrations determine the value of the Fermi level and thus of the work function. Because the dissolved gas molecules and the polymer molecules may share an electron, the charge transfer is only partial. This effect can be considered to be a modulation of the activity coefficient f_i in the bulk term of the work function (equation (12)). Note that the effect is the result of neutral (doping) molecules entering the polymer phase, which does not give a Nernstian interfacial potential like that described by equation (11) which explicitly originates from interfacial transfer of charges (ions or electrons).

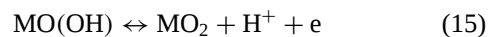
Therefore Janata and Josowicz [8] call the bulk modulation due to neutral oxidizing or reducing species non-Nernstian behaviour.

Attempts to construct an IGFET type of sensor have up to now produced hardly any experimental results which could be explained unambiguously with the modulation of the f_i or χ term of the work function (equation (12)) [9]. The obtained sensitivity to gases does not clearly show whether the bulk term, the surface term or both are involved. This is, however, no surprise because the contributions of the two terms can in principle neither be distinguished nor be controlled separately. This is mainly due to the fact that no independent measurements are available for this type of gas sensor. If the same type of sensor could be used in an electrolyte, this implicit problem would be solved, because in this case the modulation of the work function Φ (the modulation of V_T , equation (5)) and that of the electrode potential of the gate material could be measured independently. Such a device was recently developed by our group and is called the ^EMOSFET.

5. The ^EMOSFET

The ^EMOSFET contains a gate of sputtered iridium oxide. The n-channel device, which is connected to a source and drain follower circuit and which is normally in the 'on' state, results in a stable adjustment of the source voltage with respect to ground when the gate is grounded. The value of the source voltage is, up to a constant, equal to the threshold voltage V_T . This holds for dry conditions as well as for cases when the device is immersed in an electrolyte. In the latter case the grounded iridium oxide gate is in fact pushing the electrolyte to a voltage with respect to ground that is equal to the electrode potential E . Thus measuring the electrolyte with a high-ohmic amplifier via a reference electrode delivers simultaneously with the V_T measurement information about the electrode potential with respect to the potential of the reference electrode ($E - E_{ref}$). These independent measurements may allow one to gain insight into the operational mechanism of the device as a function of the oxidation or reduction of the iridium oxide in relation to electrolyte parameters. The circuit of the simultaneous measurement is shown in figure 8.

It is well known that electrodes made of hydrated metal oxides MO(OH), where M can be Ir, W, Pt and so on, take part in redox reactions of the type



For the electrode potential E with respect to a reference electrode it is easily derived that

$$E - E_{ref} = E^\circ - \frac{RT}{F} \ln \left(\frac{f_{Re}[\text{Re}]}{f_{Ox}[\text{Ox}]} \right) + \frac{RT}{F} \ln(a_{\text{H}^+}) \quad (16)$$

with $[\text{Re}] = [\text{MO(OH)}]$, $[\text{Ox}] = [\text{MO}_2]$ and f_{Re} and f_{Ox} the related activity coefficients.

From practical measurements with electrodes of this type, used as pH-sensitive sensors, based on the last term of equation (16), it is known that often a deviation is observed from the 59 mV/dec at room temperature, as the

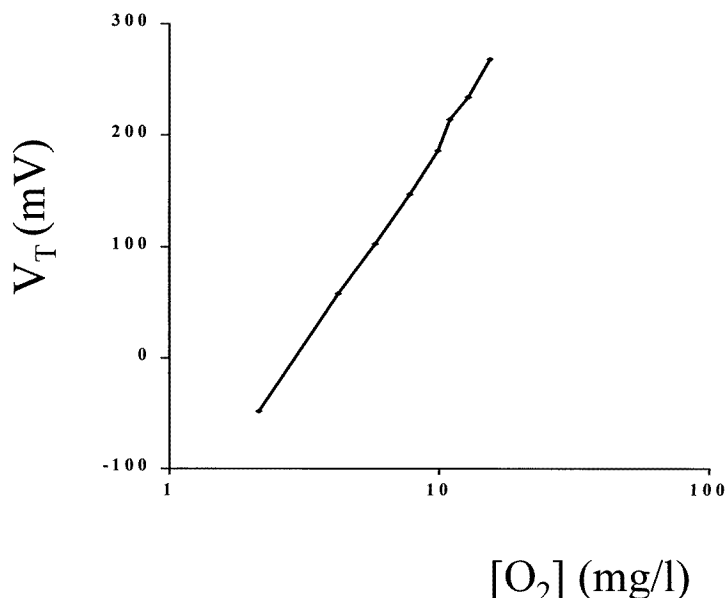


Figure 10. The threshold voltage V_T as a function of $[O_2]$ at 13 nA bias current.

caused by the bulk term of the work function of the gate material, in this case controlled by the redox mechanism.

The validity of equations (16) and (20) for the electrode potential and the work function, respectively, can be proven by measuring both simultaneously as functions of the pH of the electrolyte and subtracting both signals: $(E - E_{ref})$ and V_T . Subtracting equations (16) and (20) predicts a pure Nernstian behaviour (59 mV/dec at room temperature, the last term of equation (16)) and this is exactly what is observed, as illustrated in figure 9.

In addition it was shown that interference by oxygen due to intermittent purging with O_2 and N_2 at constant pH occurred in $E - E_{ref}$ as well as in V_T but not in the difference between the two signals. This proves that the oxidized: reduced ratio is responsible for the interference in the bulk term of the work function [10].

Now it is proven that an E MOSFET is in fact a redox sensor it can thus be applied as a potentiometric oxygen sensor as an alternative to the well-known amperometric Clark cell. Because at zero current it takes a long time for the oxygen in the solution and the iridium oxide electrode to reach equilibrium due to the very low exchange-current density, a small externally supplied current is used to force the redox reaction to adjust. So in this case the V_T of the E MOSFET is measured with a small current applied through the oxide/electrolyte interface with respect to an arbitrary counter electrode. Figure 10 shows the oxygen dependence of V_T at a reducing current of 13 nA [11].

The sensitivity is 300 mV/dec, a value which can also be calculated from the related theory. In order to prevent interference with other redox-active species the E MOSFET should be covered with a thin membrane. Note that this membrane serves only the selectivity of the sensor and is not part of the sensor's operational mechanism as is the case with the Clark cell.

At present the E MOSFET's development is being extended with the use of conducting polymers instead of

the iridium oxide. Doping of the polymer and incorporation of enzymes will make the development of a large variety of sensors possible.

6. Conclusion

In this paper a series of sensors has been described and discussed, of which the operational mechanism is based on the modulation of the work function, which is measured by a FET structure, thus with respect to the work function of silicon. The most well-known work-function sensors are gas sensors of which it could not always be proven whether the bulk term or the surface term of the work function was the actual sensing parameter.

The recently developed E MOSFET is meant to be used in electrolytes, which is not only a different application but also makes it possible to identify the actual parameter which gives the sensor its sensing properties. This makes it easier to optimize its operation and to develop an entirely new class of sensors which are based on work-function modulation.

Acknowledgments

The authors thank the technician G Bultstra for making the illustrations.

References

- [1] Bergveld P 1970 Development of an ion-sensitive solid state device for neurophysiological measurements *IEEE Trans. Biomed. Eng.* **17** 70–1
- [2] Lundström I, Shivaraman A S, Svensson C M and Lundkvist L 1975 Hydrogen sensitive MOS field effect transistor *Appl. Phys. Lett.* **26** 55–7
- [3] Krey D, Dobos K and Zimmer G 1982 *An Integrated Co-sensitive MOS Transistor, Sensors and Actuators* vol 3, pp 169–77

- [4] Blackburn G F, Levy M and Janata J 1983 Field-effect transistor sensitive to dipolar molecules *Appl. Phys. Lett.* **43** 700–1
- [5] Josowicz M and Janata J 1986 Suspended gate field effect transistor modified with polypyrrole as alcohol sensor *Anal. Chem.* **58** 514–7
- [6] Papez V and Brodská S 1997 Deposition of the chemically sensitive polymer layer on SGFET gate by laser induced chemical-vapor deposition *Sensors Actuators B* **40** 143–5
- [7] Janata J 1991 Chemical modulation of the electron work function *Anal. Chem.* **63** 2546–50
- [8] Janata J and Josowicz M 1997 Nernstian and non-nernstian potentiometry *Solid State Ionics* **94** 209–15
- [9] Liess M, Chinn D, Petelenz D and Janata J 1996 Properties of insulated gate field-effect transistors with a polyaniline gate electrode *Thin Solid Films* **286** 252–5
- [10] Hendrikse J, Olthuis W and Bergveld P 1998 A method of reducing oxygen induced drift in iridium/oxide pH sensors *Sensors Actuators B* to appear
- [11] Hendrikse J, Olthuis W and Bergveld P 1998 The E^+ MOSFET as a potentiometric transducer in an oxygen sensor *Sensors Actuators B* **47** 1–8