

THE HISTORY OF CHEMICALLY SENSITIVE SEMICONDUCTOR DEVICES

P BERGVELD and N F DE ROOIJ

*Technische Hogeschool Twente, Afdeling der Elektrotechniek, Enschede, P O B 217
(The Netherlands)*

1. Introduction

The idea of modulating the conductance of a small semiconductor slice by means of an electric field perpendicular to the semiconductor surface dates from the late twenties [1] The proposed configuration, which shows a similarity with a simple capacitor, is represented in Fig 1

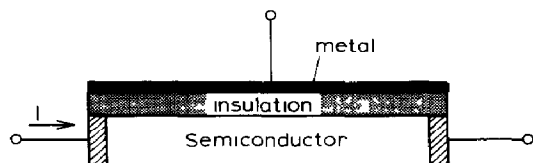


Fig 1 Original idea of a solid-state active component

A positive charge on the metal plate induces a negative charge in the semiconductor, resulting in a change in the lateral conductance The importance of this idea is the possibility of controlling a current without using input energy in a very small, solid device As is often the case with a first idea, it took many years before what was later called the "field effect", could be confirmed experimentally Intensive research by Brattain and Bardeen [2] at Bell Laboratories in the period 1947 - 1950 with germanium monocrystals did not result in any provable field effect The reason was that the induced negative charges were immobile, because they were trapped in free valence bonds, the so-called "surface states" These surface states can, however, be neutralised, as proved some years later by Garrett and Brattain, who experimented with a germanium-electrolyte cell [3] They showed that the semiconductor conductance could be considerably influenced by applying a negative potential to the germanium with respect to the electrolyte These experiments did not have any practical significance at that time, although later the influence of electrolyte concentration on the semiconductor conductance was also investigated and proved under certain conditions [4] In principle, the development of chemical sensitive semiconductor devices could have started at that moment at Bell Laboratories This was, however, not the aim of the investigators and the practical result was the invention of

the p-n junction (1949) by Shockley, and in relation to this, the junction transistor (1950) [5]

In the following decennium, the junction transistor was further developed, although the surface state phenomenon still appeared to be a large problem. To solve this many investigations were started, while, in the meantime, silicon was ousting germanium as the basic semiconductor material. The decisive step was proposed and proved to be realistic by Atalla in 1958 [6], who discovered that a thermal oxidation of the silicon surface resulted in neutralisation of most of the surface states. Soon after, the basic principles of the planar silicon technology were consolidated, followed by the invention of the Metal Oxide Semiconductor Field Effect Transistor (MOSFET) by Kahng and Atalla (1960) [7]. So, after approximately 40 years, the original idea of modulating the conductance of a semiconductor slice by means of an electric field was proved to be realistic and converted into a working device. The insulating layer of the device shown in Fig 1 is now the thermally grown silicon dioxide, while the surface conductance is measured by diffusions which are added to the structure and called source and drain, respectively, as shown in Fig 2. A mean value of the oxide thickness is $0.1 \mu\text{m}$. The metal plate of the capacitive structure is called the gate of the device. Due to the insulating properties of the input of the device, it is also called Insulated Gate Field Effect Transistor (IGFET)

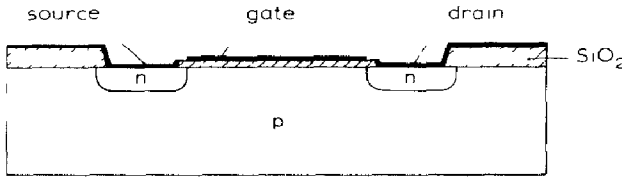


Fig 2 Cross section of a MOSFET

2. The operation of a MOSFET

The MOSFET shown in Fig 2 is an n-channel device, because a positive potential on the gate with respect to the substrate results in an inversion of the type of conductance of the p-type substrate at the surface between the n^+ diffusions of source and drain. The conductance of this inversion channel is given by

$$g = \mu C_{\text{ox}} \frac{W}{L} (V_g - V_t) \quad (1)$$

where μ is the mobility of electrons, C_{ox} the MOS capacity per unit area, L the length of the channel between source and drain, W the channel width, V_g the gate voltage and V_t the threshold voltage. The expression for the conductance, g , given by eqn (1) is only valid for $V_g > V_t$, resulting in a surface which is in strong inversion.

The threshold voltage, V_t , can be written in distinguishable terms as follows

$$V_t = \Phi_M - \Phi_{Si} - (Q_{ss} + Q_{ox})/C_{ox} + 2\Phi_f - Q_B/C_{ox} \quad (2)$$

where Φ_M and Φ_{Si} are the work functions of the metal and the silicon, respectively, $(Q_{ss} + Q_{ox})/C_{ox}$ is the contribution to V_t caused by the charge of the interface states Q_{ss} and the charge in the oxide Q_{ox} , also assumed to be located at the silicon interface. The first three terms of eqn (2) are called the flat band voltage V_f , according to

$$V_f = \Phi_M - \Phi_{Si} - (Q_{ss} + Q_{ox})/C_{ox} \quad (3)$$

The flat band voltage is equal to that value of the gate voltage which results in flat electron energy bands in the silicon up to and including the surface.

The last two terms of eqn (2) reflect the additional gate voltage necessary to invert the type of surface conduction. Φ_f is the Fermi potential, which is a function of the doping concentration of the silicon substrate and Q_B , the charge per unit area in the depletion region underneath the surface, also depending on the doping concentration of the substrate.

In practical measurements, the channel conductance is measured by means of applying a drain source voltage V_d to the device resulting in a drain current I_d which can be measured. Under the condition $V_d < V_g - V_t$ eqn (1) gives the following equation for I_d

$$I_d = \mu C_{ox} \frac{W}{L} [(V_g - V_t) V_d - \frac{1}{2} V_d^2] \quad (4)$$

where the last term is the result of the gradual potential along the channel from source to drain [8].

In 1970 Bergveld [9] proposed a modification of the MOSFET concept by replacing the gate metal by an aqueous solution, resulting in a device in which the channel conductance also appeared to be a function of the ion concentration of the solution. He called this device an Ion Sensitive Field Effect Transistor, ISFET. In fact, he showed that the flat band voltage can be chemically influenced by ionic interaction with the silicon dioxide. In 1975 Lundstrom [10] showed that this is also true for interaction with gases, thus introducing the GASFET. Since then many papers have appeared on this matter. In the following Sections these papers will be analysed in respect of the fundamental approach to the theoretical description of the device operation.

3 The operation of an ISFET

In Bergveld's second paper [11] the operation of the ISFET was described as a modulation of the threshold potential according to eqn (2)

$$V_t = \text{Const} + \frac{RT}{F} \ln(a_{H^+_{sol}} + K a_{Na^+_{sol}}) \quad (5)$$

The term const was not further specified, while the second term was introduced as the junction potential at the solution-oxide interface due to diffusion and boundary processes at this interface in accordance with the phenomena described for glass membrane electrodes. Bergveld suggested that this action could also operate without a reference electrode in the solution, a suggestion which resulted in many discussions in later literature. To analyse the literature we will therefore now distinguish between the two possibilities — operation with and without a reference electrode — starting with the first mentioned situation.

The literature concerning the ISFET operation with a reference electrode in the solution, mainly focusses on the description of the term “const” in eqn (5) in significant parts. One of these parts is the contribution of the voltage of the reference electrode, which is sometimes described, as part of the gate voltage V_g [12] and sometimes as part of the threshold voltage V_t or the flat band voltage V_f , but, in this case, not always correctly. This is very confusing and this point will therefore be carefully analysed in this Section. This is necessary with regard to the application of an ISFET as sensor for absolute values of ion concentrations where the values of standard electrode potentials play an important rôle. The various descriptions given in the literature can best be understood by surveying the ISFET representation as given in Fig. 3.

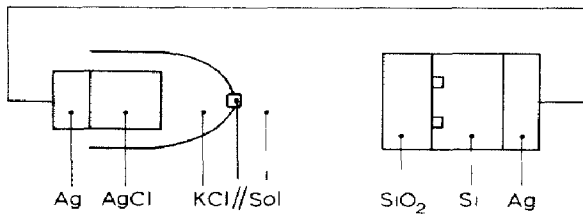


Fig. 3 Representation of ISFET connected to a reference electrode

The potential distribution can be described with the help of the Poisson-Boltzmann theory in relation to a thermodynamical analysis of the distinguishable interfaces throughout the whole measuring circuit, including the reference electrode and the connecting leads.

In general, the electromotive force of an electrochemical cell, formed by two metals, M_1 and M_2 , in an electrolyte can be described as consisting of three contributions

$$E = \Delta\phi_{M_1}^{el} - \Delta\phi_{M_2}^{el} - \Delta\phi_{M_1}^{M_2} \quad (6)$$

where

$$\Delta\phi_{M_1}^{el} = \frac{\mu_{M_1}^{+el} - \mu_{M_1}^{+met}}{F}$$

is the metal-electrolyte interfacial potential at M_1 ,

$$\Delta\phi_{M_2}^{el} = \frac{\mu_{M_2^+}^{el} - \mu_{M_2^+}^{met}}{F}$$

is the metal–electrolyte interfacial potential at M_2 ,

$$\Delta\phi_{M_1M_2} = \frac{\mu_e^{M_1} - \mu_e^{M_2}}{F}$$

is the contact potential between M_1 and M_2

For a metal–metal ion reaction,



it yields that

$$\mu_M^{met} = \mu_{M^+}^{met} + \mu_e^{met} \quad (7)$$

For certain redox couples the corresponding electrode potential is tabulated with regard to the hydrogen electrode and, in this way, the tabulated values of a reference electrode ($E_{ref} = \mu_{M^+}^{el} - \mu_M^{met}$) have to be interpreted. If in the example mentioned above M_1 is a reference electrode, substitution of eqn (7) in eqn (6) results in

$$E = E_{ref} - \Delta\phi_{M_2}^{el} + \frac{\mu_e^{M_2}}{F} \quad (8)$$

For the ISFET configuration as given in Fig 3, a similar description can be given in such a way that due to the polarisable oxide–silicon interface no emf is developed by the cell, but instead of this a voltage V_g is applied to the cell. With regard to eqn (8), M_2 now represents the silicon, S_1 . Further the term $\Delta\phi_{M_2}^{el}$ must now be written as follows

$$\Delta\phi_{M_2}^{el} = \left(\phi_o + \frac{RT}{F} \ln a_1 \right)_{ox}^{el} + \phi_{ox} + \chi_{ox}^{S_1} + \phi_{S_1} \quad (9)$$

where the first term represents the potential difference across the oxide–electrolyte interface, assuming that the oxide behaves in a Nernstian way. Further, ϕ_{ox} represents the voltage across the oxide, $\chi_{ox}^{S_1}$ the silicon surface electric potential and ϕ_{S_1} the potential across the space charge region in the silicon. The expression for the flat band voltage of an ISFET follows from eqns (8) and (9), considering that under flat band conditions $\phi_{ox} + \phi_{S_1} = 0$

$$V_f = E_{ref} - \left(\phi_o + \frac{RT}{F} \ln a_1 \right)_{ox}^{el} + \chi_{ox}^{S_1} + \frac{\mu_e^{M_2}}{F} \quad (10)$$

The last two terms of eqn (10) represent the silicon work function Φ_{S_1} (in electrochemistry called α [13]), as mentioned in the expression for the flat band voltage of an MOS transistor (eqn (3)). Adding the effect of surface state- and oxide charge, as well as a correction for the reference electrode with the liquid junction potential, the ISFET, including the reference electrode, can best be described as a modification of eqn (3) as follows

$$V_f = E_{ref} - \Delta\phi_j - \left(\phi_o + \frac{RT}{F} \ln a_1 \right) - \Phi_{S_1} - (Q_{ss} + Q_{ox})/C_{ox} \quad (11)$$

An MOS structure (eqn (3)) made on the same chip as an EOS structure (eqn (11)) makes it possible to measure a difference in the flat band voltages

$$\Delta V_f = E_{ref} - \Delta \phi_j - \left(\phi_o + \frac{RT}{F} \ln a_1 \right) - \Phi_M \quad (12)$$

Because the first two terms are tabulated with respect to the hydrogen electrode, Φ_M is a known value, being the work function of the metal, and a_1 is the known ion activity of the electrolyte (for instance a_{H^+} for an H^+ sensitive oxide), ϕ_o can now be calculated, being the standard potential of the oxide-electrolyte interface

For fundamental research on the behaviour of V_f , CV measurements on the system Electrolyte-Oxide-Semiconductor (EOS) are very suitable, because this EOS system is a simpler configuration than an ISFET, and is not affected by the possible non-ideal behaviour associated with the transistor structure

The most extended paper with respect to this is published by Siu and Cobbold [14] They distinguish between two limiting cases, namely, a blocked and an unblocked electrolyte-oxide interface For the unblocked interface the description of eqn (11) for the flat band voltage appears to be valid, while for the blocked interface, where any interfacial reactions on the oxide surface are assumed to be absent, the flat band voltage appears to be a function of the ionic concentration of the electrolyte according to eqn (13) in the low ionic concentration region

$$V_f = E_{ref} - \Delta \phi_j + \frac{2kT}{q} \sinh^{-1} [\sigma_d (8\epsilon_e kT n^o)^{-1/2}] - \Phi_{Si} - (Q_{ss} + Q_{ox})/C_{ox} \quad (13)$$

In addition to the terms already mentioned, σ_d denotes the charge in the Gouy-Chapman layer, ϵ_e the electrolyte permittivity, and n^o the ionic density The "ionic" term differs greatly from that mentioned in eqn (11) and is, in this case of low ionic concentration, the result of an appreciable potential drop across the Gouy-Chapman layer in the electrolyte, thus influencing the pure electrostatic condition The authors assume that any realistic EOS or ISFET structure will behave in a manner somewhere in between the characteristics described in the two limiting cases, probably resulting in a sensitivity of the various devices described in the literature, which is not always pure Nernstian It is essential that both limiting cases describe pure surface effects, resulting in a shift of the CV curves along the voltage axis as a function of pH, from which a linear behaviour between V_f and pH can be expected, although limited to low ionic concentrations for the case of a blocked interface However, de Rooij and Bergveld also observed a change in the shape of the CV curves, an effect that they ascribe to a chemical interaction with the surface states, thus influencing the surface-state density [15] The original curves are shown in Fig 4

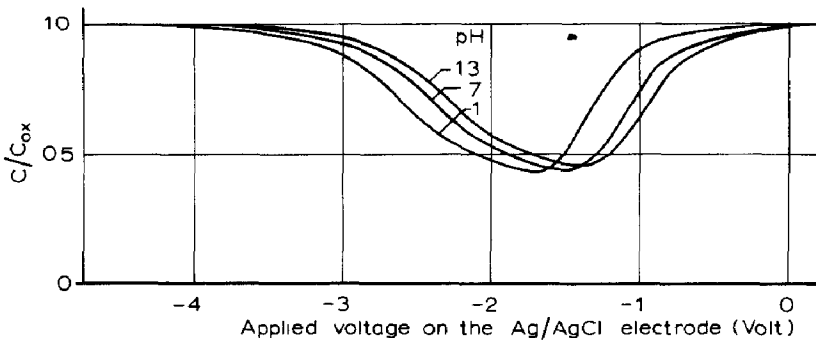


Fig 4 Measured quasi-static CV curves at different pH values on electrolyte/SiO₂/Si structures

Recent investigations show that the flat band voltage has a plateau around pH = 6. This is correlated with an observed time drift of the ISFETs, produced by the same process as the EOS structures, which changes its direction around pH = 6, as shown in Fig 5 [16].

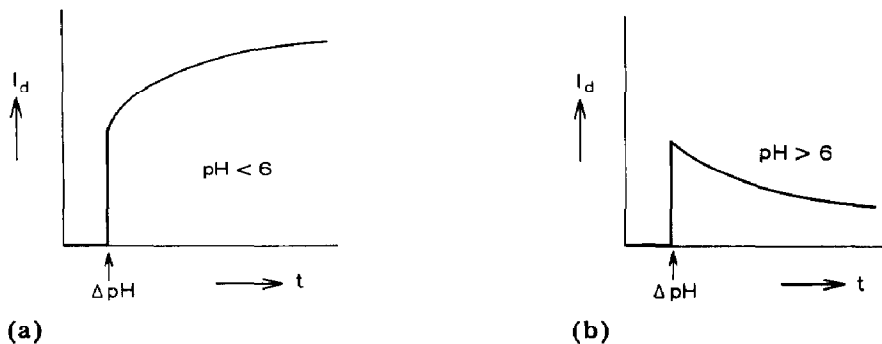


Fig 5 Transient behaviour of controlled $V_g(t)$ after a stepwise change in pH in two pH ranges (a) pH < 6, (b) pH > 6

The time constant of the drift phenomena is of the order of 4 min. It is assumed that the initial fast response is a primary effect as described by eqn (11) but that the long drift is associated with a diffusion of hydrogen bearing species through the oxide, resulting in an interaction with surface states. This entails that eqn (11) has to be modified as follows

$$V_t = E_{ref} - \Delta\phi_1 - \underbrace{\left(\phi_o + \frac{RT}{F} \ln a_1 \right)}_{\text{fast}} - \Phi_{S1} - \underbrace{(Q_{ss}f(\text{pH}) + Q_{ox})/C_{ox}}_{\text{slow}} \quad (14)$$

The secondary effect of slow drift is especially observed with thermal silicon dioxide, pretreated with aluminum, probably resulting in the formation of aluminosilicate groups. In this context, the papers of Revesz [17] have to be interpreted, where the possibility of diffusion of hydrogen or hydrogen bearing species in this type of oxide is made evident.

Note that the chemical control of the surface state density has, in fact, already been studied for many many years, even from the beginning of the development of surface devices as mentioned in the Introduction, but focussed on the minimization of the surface state effect and not on the use of it

Bergveld's original statement, that it should be possible to use ISFETs without a reference electrode will now also be clear. A direct modulation of the interface state density (the last term of eqn (14)) will also be noticeable if no reference electrode is inserted in the electrolyte. Strictly speaking, a device operating in this way, by modulating the surface state density by chemical interaction, is not a field effect transistor, because the operation is not a function of an externally applied electric field as mentioned in the Introduction. The response will be slow, although it can be speeded up for thinner oxide layers. The device will, however, remain sensitive for any electrostatic external field of which the effect limits its application possibility.

4. The application of an ISFET as sensor for absolute values of ionic activities

Because, in practice, ISFETs will mostly be used with a reference electrode and with an ion selective layer on top of an insulating layer, we will focus our attention further on a device operating according to eqn (11). It is useful to compare the operation of this device with that of a glass membrane electrode, which is a well-known device for measuring absolute values of ionic activity, a property with which the ISFET has to compete.

The measuring circuit of a pH glass membrane electrode is schematically represented in Fig 6. As can be seen from Fig 6, the e m f E_m , measured between the internal and the external reference electrode equals the sum of all voltages connected in series.

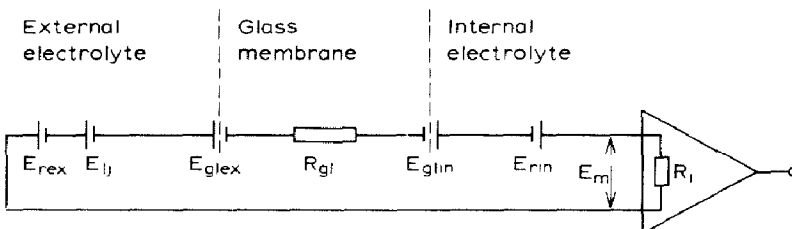


Fig 6 Schematic diagram of the measuring circuit of a pH electrode

$$E_m = E_{rex} + E_{lj} + E_{glex} + iR_{gl} - E_{glin} - E_{rin} \quad (15)$$

where E_{rex} and E_{rin} are the voltages of the external and internal reference electrodes respectively, E_{lj} is the liquid junction potential of the external reference electrode, E_{glex} and E_{glin} are the voltages at the external and

internal glass-electrolyte interfaces, while R_{gl} is the internal resistance of the glass membrane. If the internal resistance, R_1 , of the measuring amplifier is very much larger than R_{gl} , the contribution of iR_{gl} , where i is the input current of the amplifier, can be neglected. Further, it appears that reference electrodes can be constructed so accurately that

$$E_{rex} - E_{rn} + E_{l1} = 0 \quad (16)$$

can be approached

This means that eqn (15) can be simplified to

$$E_m = E_{glex} - E_{gln} \quad (17)$$

Using the Nerst equation for E_{glex} and E_{gln} it follows for a pH selective glass membrane

$$E_m = E_{ex}^0(T) + \frac{RT}{F} \ln a_{H^+_{ex}} - (E_{in}^0(T) + \frac{RT}{F} \ln a_{H^+_{in}}) \quad (18)$$

where $E_{ex}^0(T)$ and $E_{in}^0(T)$ are the standard potentials at the external and internal glass membrane surfaces, and $a_{H^+_{ex}}$ and $a_{H^+_{in}}$ are the activities of the hydrogen ions in the external and internal solutions. It appears that $E_{ex}^0(T) = E_{in}^0(T)$, independent of pH. Because the internal electrolyte has a constant pH, the last term of eqn (18) forms an offset voltage, $E_{in}(T)$, which is a function of temperature. Thus

$$E_m = -E_{in}(T) + \frac{RT}{F} \ln a_{H^+_{ex}} \quad (19)$$

Further, it is obvious from eqn (19) that also, the sensitivity for changes in $a_{H^+_{ex}}$ (pH changes to be measured) is a function of temperature, but in such a way that the isotherms have one common point, as illustrated in Fig 7.

For practical reasons it is useful that this point of intersection coincides with the electrical zero of the measuring system, usually chosen at $\text{pH} = 7$.

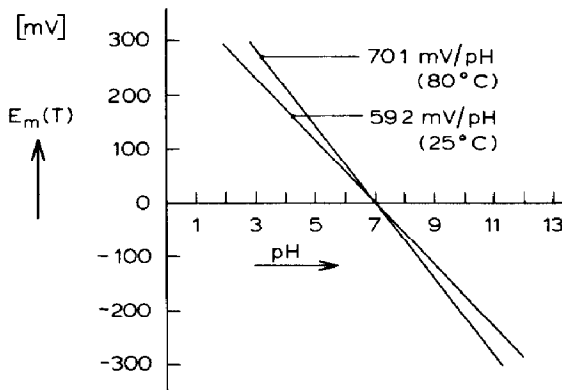


Fig 7 The characteristics of a pH glass membrane electrode for two temperatures T_1 and T_2

($a_{H^+_{ex}} = 10^{-7}$) as already shown in Fig 7 In this case the reference point of 0V on the voltage scale is insensitive for temperature variations, which is an essential condition for an absolute measurement

The isotherm interaction point coincides with the electrical zero if $dE_m/dT = 0$ From eqn (19) it can be seen that for pH = 7 this is the case if

$$\frac{d}{dT} (E_m(T)) = 1.39 \text{ mV}/^\circ\text{C} \quad (20)$$

It appears that this requirement can be met in practice, within the described accuracy, by a careful choice of the composition and buffering properties of the internal electrolyte of the electrode

Changes in temperature now only require a compensation for the sensitivity of the electrode, or, in other words, the slope of the curve given in Fig 7 In practice this can easily be achieved by an adjustment of the amplification of the measured voltage as a function of the temperature, which can be measured by a separate sensor

The glass membrane electrode is an excellent example where technological process control leads to a sensor with a built-in absolute reference point, which, unfortunately, cannot be said for many other sensors used for measuring physical or chemical quantities The question arises whether an ISFET can compete in this way with the glass membrane electrode as a replacement in the future

Combining eqns (4), (2), (3) and (11) results in the following measurable output quantity of an ISFET

$$I_d = \mu C_{ox} \frac{W}{L} \left[\left(V_g - E_{ref} + \Delta\phi_j + \phi_o + \frac{RT}{F} \ln a_1 + \Phi_{S1} + \frac{Q_{ss} + Q_{ox} + Q_B}{C_{ox}} - 2\Phi_f \right) V_d - \frac{1}{2} V_d^2 \right] \quad (21)$$

It is useful to find a condition in agreement with the considerations mentioned above for a glass membrane electrode where $I_d = \text{const}$ for pH = 7 ($a_1 = 10^{-7}$), independent of temperature variations Then again, only the sensitivity factor of the device has to be compensated as function of the measured temperature It is obvious from eqn (21) that much more temperature sensitive parameters are now involved, partly of a chemical and partly of a physical (solid state) character The contribution of the reference electrode is now part of the equation, in contrast to the glass membrane electrode where two reference electrodes are necessarily used and these contributions to the output voltage of the electrode compensate each other Also, the standard potential of the electrolyte-oxide interface is now part of the equation

Knowing the tolerances in solid-state technological processing it is unrealistic to expect that a similar solution as realized for glass membrane electrodes (eqn (20)) can be found for ISFETs by a proper choice of materials and processing Fortunately, the ISFET has one great advantage over the conventional glass membrane electrode and that is the fact that it is an elec-

tronic component, where certain properties can be adjusted electronically. It is therefore useful further to consider eqn (21) as a basic equation for various electronic application possibilities. This will be the subject of a following paper [18].

References

- 1 J T Wallmark, The field-effect transistor, an old device with new promise, *IEEE Spectrum*, 1 (3) (1964) 182 - 186
- 2 W H Brattain and J Bardeen, Surface properties of germanium, *Bell Syst Tech J*, 32 (1953) 1 - 41
- 3 W H Brattain and C G B Garrett, Experiments on the interface between germanium and an electrolyte, *Bell Syst Tech J*, 34 (1955) 129 - 175
- 4 G Brouwer, Control of the surface potential of germanium by means of a variable pH electrolyte containing hydrogen peroxide and potassium chlorides, *J Electrochem Soc, Solid State Sci*, (1967) 743 - 748
- 5 W Shockley, A unipolar field effect transistor, *Proc IRE* 40 (1952) 1365 - 1376
- 6 M M Atalla *et al*, Stabilization of silicon surface by thermally grown oxides, *Bell Syst Tech J*, 38 (1959) 749 - 783
- 7 D Kahng and M M Atalla, Silicon-silicon dioxide field induced surface devices, *IRE Solid State Device Res Conf, Pittsburg, 1960*
- 8 J A Van Nielen and O W Memelink, The influence of the substrate upon the d.c. characteristics of silicon MOS transistors, *Philips Res Rep*, 22 (1967) 55 - 71
- 9 P Bergveld, Development of an ion-sensitive solid-state device for neurophysiological measurements, *IEEE Trans Biomed Eng*, 17 (1970) 70 - 71
- 10 I Lundstrom *et al*, A hydrogen sensitive Pd-gate MOS transistor, *J Appl Phys*, 46 (1975) 3876 - 3881
- 11 P Bergveld, Development, operation and application of the ion sensitive field-effect transistor as a tool for electrophysiology, *IEEE Trans Biomed Eng*, 19 (1972) 342 - 351
- 12 S D Moss *et al*, pH Catheter tip semiconductor microelectrode, in *Theory, Design and Biomedical Applications of Solid State Chemical Sensors*, CRC Press, (1978) pp 119 - 131
- 13 R P Buck and D E Hackleman, Field effect potentiometric sensors, *Anal Chem*, 49 (14) (1977) 2315 - 2321
- 14 W M Siu and R S C Cobbold, Basic properties of the electrolyte-SiO₂-Si system physical and theoretical aspects, *IEEE Trans Electron Devices*, 26 (11) (1979) 1805 - 1855
- 15 N F De Rooij and P Bergveld, The influence of the pH on the surface state density at the SiO₂-Si interfaces, in S T Pantelides (ed), *The Physics of SiO₂ and its Interfaces*, Pergamon Press, New York, 1978, pp 433 - 437
- 16 N F De Rooij and P Bergveld, The influence of the pH on the electrolyte-SiO₂-Si system studied by ion-sensitive FET measurements and quasi-static C-V measurements, *Thin Solid Films*, 71 (1980) 327 - 331
- 17 A G Revesz, On the mechanism of the ion sensitive field effect transistor, *Thin Solid Films*, 41 (1977) 643 - 647
- 18 P Bergveld, The operation of an ISFET as an electronic device, *Sensors and Actuators*, 1 (1981) 17 - 29