CHAPTER 2

REVIEW OF LITERATURE

Chapter 2

Review of Literature

2.1. Theory of MOSFETs

MOSFET is a transistor in which the MOS (metal oxide semiconductor) part is related to the structure of the transistor, while the FET (field effect transistor) part is related to how it works. It consists of an oxide acting as gate insulator sandwiched between metal gate and a semiconductor. The structure of n-channel MOSFET is shown in Fig. 2.1.

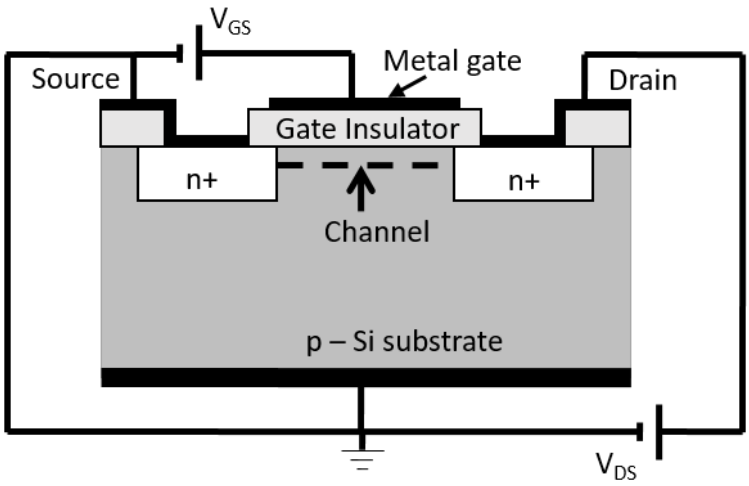

Fig. 2.1. Structure of n-channel MOSFET

The drain current equations for MOSFET in non-saturated or linear ($I_{DS,lin}$) and saturated region ($I_{DS,sat}$) are given by Eqs. (2.1) and (2.2) respectively [27].

$$I_{DS,lin} = \mu \frac{W}{L} V_{DS} C_{ox} \{(V_{GS} - V_{TH,MOSFET}) - 0.5\, V_{DS}\} \tag{2.1}$$

$$I_{DS,sat} = \mu \frac{W}{2L} C_{ox} (V_{GS} - V_{TH,MOSFET})^2 \tag{2.2}$$

where, μ is the electron mobility of the channel, W is the width and L is the length of the channel, C_{ox} is the oxide capacitance, V_{GS} is the gate to source voltage, V_{DS} is the drain to source voltage and $V_{TH,MOSFET}$ is the threshold voltage of MOSFET.

The threshold voltage of MOSFET is given by

$$V_{TH,MOSFET} = \frac{\Phi_M - \Phi_S}{q} - \frac{Q_{ox} + Q_{ss} + Q_B}{C_{ox}} + 2\Phi_f \quad (2.3)$$

where, Φ_S is the work function of semiconductor, q is the elementary charge, Q_{ox} is the gate insulator charge, Q_{ss} is the insulator semiconductor interface charge, Q_B is the charge present in the bulk semiconductor, Φ_M is the work function of gate material and Φ_f is the Fermi potential of the bulk semiconductor. The factor $2\Phi_f$ is termed as the inversion potential, which is required to form the inversion channel.

2.2. Problems in scaling of traditional MOSFETs

In the process of scaling, device miniaturization is done by conserving or improving the electrical characteristics of the device. That means a device of smaller size must give similar or better performance [54]. Reducing device dimensions means overall scaling which includes reduction in gate length, gate width, oxide thickness, depletion layer width, etc. The first complete scaling method was known as constant electric field scaling and was given by Dennard *et al.* in 1974 [23]. Later, another method came, which is known as constant voltage scaling. These are discussed in later sub-sections.

2.2.1. Constant Field Scaling

In this method of scaling, the device parameters are scaled in such a way as to keep the electric field constant. The scaling is done by a dimensionless factor λ. Table 2.1 shows the scaling of parameters using constant field scaling. It is seen that using this method many parameters can be scaled considerably giving better performance. The dimensions of the device have been trimmed down along with decrease in supply and threshold voltage, which have improved device performance along with increasing the circuit density. It also gives constant power density and highest reduction in power-delay product of a single transistor. Apart from these, when it comes to practical and experimental applications, this scaling method has many drawbacks. For instance,

when the supply voltage (V_{DD}) becomes very low (say about 1 V), further reduction in threshold voltage (V_{TH}) creates problem. This problem arises as thermal voltage ($V_T = kT/q$) cannot be scaled which is related to the subthreshold swing (SS) and subthreshold leakage current, discussed in later sections.

a. Subthreshold leakage current

The drain current that flows through a MOSFET device when the gate to source voltage is lower than the threshold voltage of the device is called subthreshold current. It leads to leakage power dissipation when the MOSFET is in standby mode. This power dissipation is also known as static power dissipation, which should have a very low value. Its higher value leads to battery draining of any device even in standby mode.

b. Subthreshold swing

The subthreshold current (I_{SubTH}) increases exponentially with increase in V_{GS} and decrease in V_{TH} as given by Eq. (2.4) [54].

$$I_{SubTH} \propto \exp\left(\frac{qV_{GS}}{nkT}\right) \tag{2.4}$$

The slope of the graph between $\log_{10} I_{SubTH}$ and V_{GS} is known as subthreshold slope. The inverse of subthreshold slope is known as subthreshold swing (SS)[54].

$$SS = \frac{\partial V_{GS}}{\partial \log_{10} I_{SubTH}} = \frac{2.303nkT}{q} \tag{2.5}$$

If $n = 1$, the value of SS is 60 mV/decade at room temperature. This is the best value of SS. Its value is generally greater than its best value at room temperature. However, the transistor should be designed in such a way that SS has a small value because its smaller value improves the ON/OFF current ratio of the device [54].

$$ON/OFF\ current\ ratio = \frac{I_{DS}|_{V_{GS}=0}}{I_{DS}|_{V_{GS}=V_{TH}}} \tag{2.6}$$

The value of V_{TH} has certain limitation, say it must be greater than about 0.4 V. For a device to have a very high performance, there must be sufficient gap between V_{TH} and V_{DD}. So, V_{DD} cannot be reduced to a very low value as V_{TH} cannot be downscaled. This

problem is reduced to a great extent with the use of nanomaterial based MOSFETs as V_{TH} variation can be done in varied range. Nanomaterial based MOSFETs also gives small SS and high ON/OFF current ratio.

2.2.2. Constant Voltage Scaling

Different power supplies need to be manufactured if supply voltages are scaled along with device dimensions. This results in wastage of resources, time and money. So, constant voltage scaling method was developed so, that standard voltage supplies can be used. In this method the supply voltage is kept constant and the device dimensions are reduced along with other parameter variations as shown in Table 2.1. The low V_{TH} problem does not arise in this case. Still it has many drawbacks. The main disadvantage of this method is that with reduced device dimensions, I_{DS} increases along with increase in current density, power and power density. The huge increase in power causes heating problems. Many other phenomena like electromigration in metal contacts, velocity saturation, hot carrier injection, punch through, high leakage current, etc. cannot be neglected.

The discussions above show that device scaling has its own limitations in various respects that are very difficult to overcome with the use traditional MOSFET devices. Even if the devices are scaled to some extent with certain limitations, for the small devices the short channel effects come into effect.

2.3. Short Channel Effects in traditional MOSFETs

The channel of a MOSFET (L) is said to be short when its length approximately equal to the space charge regions of source and drain junctions with the substrate. If x_{dD} and x_{dS} are the width of the depletion regions of drain and source respectively, then, a short channel MOSFET has $L \approx (x_{dD}, x_{dS})$ whereas a long channel MOSFET has $L \gg (x_{dD}, x_{dS})$. When the channel becomes short, the device has to face a lot of issues such as polysilicon gate depletion effect, threshold voltage roll-off, drain-induced barrier

lowering (DIBL), velocity saturation, reverse leakage current rise, mobility reduction, hot carrier effects, etc. These are known as short channel effects.

Table 2.1: Parameters obtained after constant field and constant voltage scaling

Sl. No.	Parameters	Constant field scaling	Constant voltage scaling
1.	Gate length (L_g)	L_g/λ	L_g/λ
2.	Gate width (W_g)	W_g/λ	W_g/λ
3.	Electric field (E)	E	λE
4.	Junction depth (x_j)	x_j/λ	x_j/λ
5.	Gate oxide thickness (t_{ox})	t_{ox}/λ	t_{ox}/λ
6.	Substrate doping density (N_A or N_D)	λN_A or λN_D	$\lambda^2 N_A$ or $\lambda^2 N_D$
7.	Gate oxide capacitance (C_{ox})	λC_{ox}	λC_{ox}
8.	Gate capacitance (C_{gate})	C_{gate}/λ	C_{gate}/λ
9.	Threshold voltage (V_{TH})	V_{TH}/λ	1
10.	Drain-source voltage (V_{DS})	V_{DS}/λ	1
11.	Drain-source current (I_{DS})	I_{DS}/λ	λI_{DS}
12.	Transit time (t_{tr})	t_{tr}/λ	t_{tr}/λ^2
13.	Transit frequency (f_T)	λf_T	$\lambda^2 f_T$
14.	Power dissipation (P)	P/λ^2	λP
15.	Power-delay product ($P\Delta t$)	$P\Delta t/\lambda^3$	$P\Delta t/\lambda$

As the channel length decreases, the bulk charge present near the gate decreases resulting in reduction in V_{TH}, which is known as V_{TH} roll-off. If the drain voltage becomes very high in short channel MOSFET, then channel depletion region and source/drain depletion region starts sharing charge easily. So, even at a lower value of V_{GS}, the transistor turns on resulting in decrease in V_{TH}. This is known as drain-induced barrier lowering (DIBL). The threshold voltage change (ΔV_{TH}) due to DIBL with

respect to the threshold voltage of a long channel device ($V_{TH,longchannel}$) is given by Eq. (2.7) [54].

$$\Delta V_{TH,DIBL} \equiv |V_{TH}| - |V_{TH,longchannel}| = -\left(\frac{qN_A W_d}{C_{ox}}\right)\left(\frac{r_j}{L_t}\right)\left(\sqrt{1+\frac{2W_d}{r_j}}-1\right) \quad (2.7)$$

where, N_A is the acceptor doping concentration, W_d is the width of the depletion region, C_{ox} is the oxide capacitance, r_j is the source/drain region's radius of curvature and L_t is the length of the channel region in contact with the gate oxide. The $\Delta V_{TH,DIBL}$ can be decreased by decreasing N_A and r_j and increasing C_{ox}.

If V_{DS} is further increased, then the source and drain depletion regions merge with each other and a very high current flows between source and drain having no control by the gate. This phenomena is known as punch-through. The punch through voltage (V_{PT}), given by Eq. (2.8) varies directly with the square of channel length (L) [54]. Thus, as L decreases V_{PT} also, decreases.

$$V_{PT} = \frac{qN_A L^2}{2\varepsilon_0 \varepsilon_{Si}} \quad (2.8)$$

Carrier velocity is proportional to low values of longitudinal electric field. In short channel MOSFETs, the longitudinal electric field increases abruptly and loses its relationship with carrier velocity resulting in carrier velocity saturation (v_{sat}). In this phenomenon, the carrier scattering increases and the saturation current linearly varies with $V_{GS} - V_{TH}$.

$$I_{DS,sat} = v_{sat} W C_{ox} (V_{GS} - V_{TH})^2 \quad (2.9)$$

Also, the saturation current becomes independent of channel length and the device saturates at a lower value of V_{DS}. Due to this carrier velocity saturation, carrier mobility also degrades. Moreover, the high velocity electrons hit the Si atoms releasing electrons from the valence shell, which also acquire high velocity are this process further continues resulting in generation of electron-hole pairs with huge number of free carriers. This phenomenon is known as impact ionization.

The electrons or holes which acquire a very high kinetic energy due to high electric field are known as hot carriers. Such carriers can enter the areas such as gate dielectric and substrate region resulting in variation in V_{TH}. This phenomenon is known as hot carrier injection, which is more prominent in N-channel devices due to high mobility of electrons over holes.

The short channel effects, DIBL and gate leakage current can be reduced to a great extent by reducing the thickness of oxide layer. This can be done by using high-κ dielectric materials, discussed in later sections. To enhance the carrier mobility, Si is replaced by high mobility semiconductors such as GaAs, InP, etc. But using such materials for large scale production is lower. Nanomaterials such as graphene and CNT can be used instead of these, as they have much higher mobility than the former materials.

2.4. High-κ dielectrics

With the increase in the demand by IC industry to obtain greater performance at lower cost, the idea to increase the circuit density came into light. Increase in circuit density corresponds to more number of transistors on wafer. To accommodate more number of transistors, the transistor size should be reduced which corresponds to the decrease in channel length and thickness of dielectric material used. The IC technology is very much dependent on Si-based devices, which has SiO_2 as the most suitable dielectric. SiO_2 is a low-κ having κ value equal to 3.9. The scaling of SiO_2 can be done to at least 13 Å, which is possible under some critical conditions. The direct tunneling limit for SiO_2 is about 3 nm, beyond which high tunneling leakage arises. Moreover, the distance between Si atoms is about 0.3 nm, which is the ultimate limit for reduction in thickness.

The improvement of device performance by scaling is associated with the drain current equation of MOSFET as given by Eqs. (2.1) and (2.2). From both the current equations, it is seen that the drain current is directly proportional to oxide capacitance and inversely proportional to channel length. Considering the oxide or gate capacitance, its value can be calculated using the parallel plate capacitor equation given below:

$$C_{ox} = \frac{\kappa \varepsilon_0 A}{t_{ox}} \tag{2.10}$$

where, ε_0 is the permittivity of free space, t_{ox} is the oxide thickness, A is the area of capacitor and κ is the dielectric constant or relative permittivity.

The term C_{ox} can be written in terms of t_{SiO_2} with $\kappa_{SiO_2} = 3.9$ for SiO_2 as $t_{SiO_2} = 3.9\ \varepsilon_0(A/C_{ox})$. For a fixed value of capacitance density (C_{ox}/A), we can estimate the thickness of another dielectric, $t_{high-\kappa}$ having dielectric constant, $\kappa_{high-\kappa}$ using the relation,

$$\frac{t_{SiO_2}}{\kappa_{SiO_2}} = \frac{t_{high-\kappa}}{\kappa_{high-\kappa}} \Rightarrow t_{high-\kappa} = \frac{\kappa_{high-\kappa}}{3.9} t_{SiO_2} \tag{2.11}$$

Let us assume, $t_{SiO_2} = 13$ Å and the other high-κ dielectric material be ZrO_2 having $\kappa_{high-\kappa} = 25$. Therefore, the $t_{high-\kappa}$ value for ZrO_2 to maintain the same capacitance density will be about 83 Å. So, if the thickness of ZrO_2 is further reduced, the capacitance density will increase resulting in higher device performance at reduced device dimensions.

Only using high-κ dielectric instead of SiO_2 doesn't solve the problem. There are many other factors, which are to be analyzed before selecting an alternative dielectric such as band gap, band alignment to Si, thermal stability, film morphology, interface quality, compatibility with the current or expected materials to be used in processing for MOS devices, process compatibility and reliability [65]. One of the most important factor among these is the formation of interface charges. At Si- SiO_2 interface too trap charges are found but they become noticeable only when the oxide thickness is reduced to a great extent. Using high-κ dielectric on top of Si generates huge number of oxide trap and oxide-semiconductor interface trap charges resulting in degraded device characteristics. Because of this weak interface, researches tried to search for new materials having good compatibility with high-κ dielectrics. In this direction, nanomaterials came into light. Deposition of high-κ gate insulators does not degrade the carrier mobility of nanomaterials such as graphene and CNT because the topological structure results in an absence of dangling bonds. Moreover, steep switching between ON and OFF states for graphene or CNT transistors can be achieved by the integration

of thin high-κ gate dielectrics. Thus, the good interface between high-κ dielectrics and graphene or CNT makes them promising for use as semiconductor substrate.

2.5. Theory of ISFETs

ISFET was first introduced by Bergveld in 1970 [11] and called as the first miniaturized silicon-based chemical sensor. It is a modification of MOSFET with the metal gate connection replaced by a reference electrode inserted in an electrolyte, which is in contact with the gate oxide. So, the theory of ISFET can be obtained in relation with MOSFET theory [12]. For instance, let us consider an n-channel structure of ISFET as depicted in Fig. 2.2. It comprises of a p-doped silicon substrate and n-doped source and drain regions. These two regions are separated by a short channel, which has a gate insulator on its top. Generally, the gate insulator is made up of SiO_2. Also, this SiO_2 layer has on top another insulator layer formed of materials like Si_3N_4, Al_2O_3, Ta_2O_5 which are typically pH sensitive materials. This forms a double layer gate insulator and is used for pH sensitive ISFETs. An ISFET starts operating when the gate voltage (E_{ref}) is applied. This voltage is provided by a reference electrode, generally Ag/AgCl electrode is used. This electrode gives a closed circuit and bias to the analyte. A positive bias voltage with respect to the bulk, results in formation of an n-type inversion layer in the channel between source and drain. The drain current depends on the potential variations at the various interfaces and is given by the MOSFET drain current equations.

In both the drain current equations of MOSFET, we see that the drain current varies with V_{GS} when other parameters ($\mu\frac{W}{L}C_{ox}, V_{DS}$ and V_{TH}) are constant. $\mu\frac{W}{L}C_{ox}$ is a design constant, V_{DS} can be kept constant and V_{TH} is kept constant by controlled fabrication. So, V_{GS} is the only variable input parameter for MOSFET. Looking at ISFET, the reference electrode potential acts as the gate potential. Apart from this the some interfacial potential variations occurs at the ISFET's surface. These variations are incorporated with the threshold voltage considering threshold voltage of ISFET as the input variable.

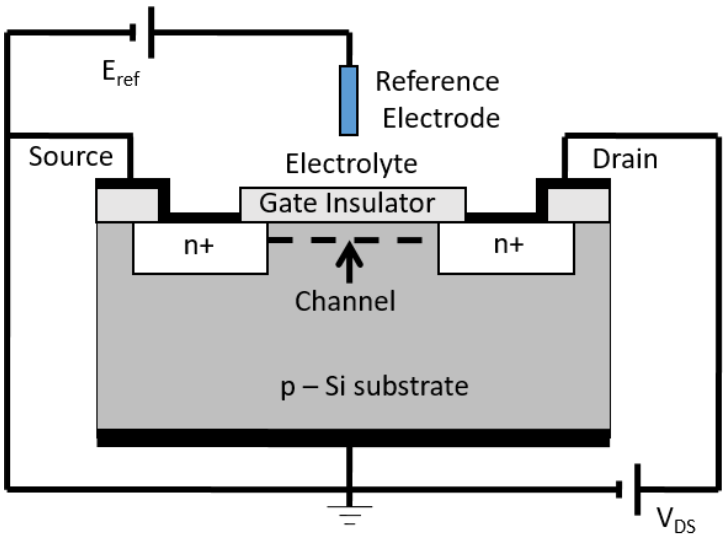

Fig. 2.2. The n-channel structure of an ISFET

The fabrication process of ISFET is similar to MOSFET, so, the constant physical parameters in the MOSFET threshold voltage equations are also included in ISFET threshold voltage replacing the metal gate work-function by a constant reference electrode potential, E_{ref} and the interfacial potential, $\psi_0 + \chi_{sol}$ at the oxide electrolyte interface. χ_{sol} is known as the surface dipole potential of the solvent, which is constant. The variable parameter is ψ_0, which is the surface potential at the oxide electrolyte interface and it varies with pH of the solution. Thus, the expression for threshold voltage of ISFET is given by [27]:

$$V_{TH,IS} = E_{ref} - \psi_0 + \chi_{sol} - \frac{\Phi_s}{q} - \frac{Q_{ox} + Q_{ss} + Q_B}{C_{ox}} + 2\Phi_f \tag{2.12}$$

The potential ψ_0 is mostly pH dependent and can be calculated by the equation given by Bousse [14] on the basis of site binding theory [104]:

$$\psi_0 = \left(\frac{2.303kT}{q}\right)\left(\frac{\beta}{\beta+1}\right)(pH_{pzc} - pH) \tag{2.13}$$

where, pH_{pzc} is called as point of zero charge. It is the value of pH for which $\psi_0 = 0$ i.e. oxide surface in electrically neutral; k is the Boltzmann constant; T is the absolute

temperature and β is the final sensitivity parameter. The derivation of this expression has been shown in later sections.

2.6. Theory of ENFETs

ENFET is a bioelectronic device, which is formed by incorporating an enzyme layer with an ISFET. The working principle of ENFET is based on drain current controlling as in ISFETs. Similar to ISFETs, the drain current of ENFETs depends on an interfacial potential generated between sensing (solid) layer and electrolyte solution (liquid) layer as explained by Nernst equation. But, in ENFETs on top of the gate oxide another sensing layer is deposited to facilitate immobilization of enzymes. Diffusion of biomolecules occur in the enzyme sensing layer following the Fick's law of diffusion. The enzymes catalyzes formation of products from the substrate biomolecules thereby altering the pH value at the gate surface. The product biomolecules forms surface complexes at the oxide surface as per the site binding theory. The binding of a charged biomolecules results in depletion or accumulation of carriers (Si–ENFET) [56, 74]. This results in change in the electrostatic potential of the semiconductor. As a result, the conductance of the transporting layer changes. This change of conductance is directly yields the concentration of biomolecules. This conductance contributes drain current flow between the source and drain. Hence, from signal transduction point of view, an ENFET can be termed as a bioelectronic device. Its main purpose is to convert a biochemical or biological signal into an electrical signal.

The idea of ENFET was first given by Janata and Moss in 1976 [44]. Taking this idea, Caras and Janata realized the first ENFET in 1980 [17], which was a penicillin sensitive biosensor as shown in Fig. 2.3 [86]. In this biosensor, the hydrolysis of penicillin was used as the basic enzymatic reaction catalyzed by penicillinase enzyme. This resulted in production of penicilloic acid as product, which thereby became the cause of pH variation of the ISFET. Ta_2O_5 was used as pH-sensitive membrane on top of gate insulator. The enzyme penicillinase was immobilized on top of Ta_2O_5. The output directly varied with the concentration penicillin in the given solution. In the similar way, various other enzymes can be utilized for ENFET development. Few enzyme-

analyte combinations, which are mostly used for development of biosensors has been shown in Table 2.2.

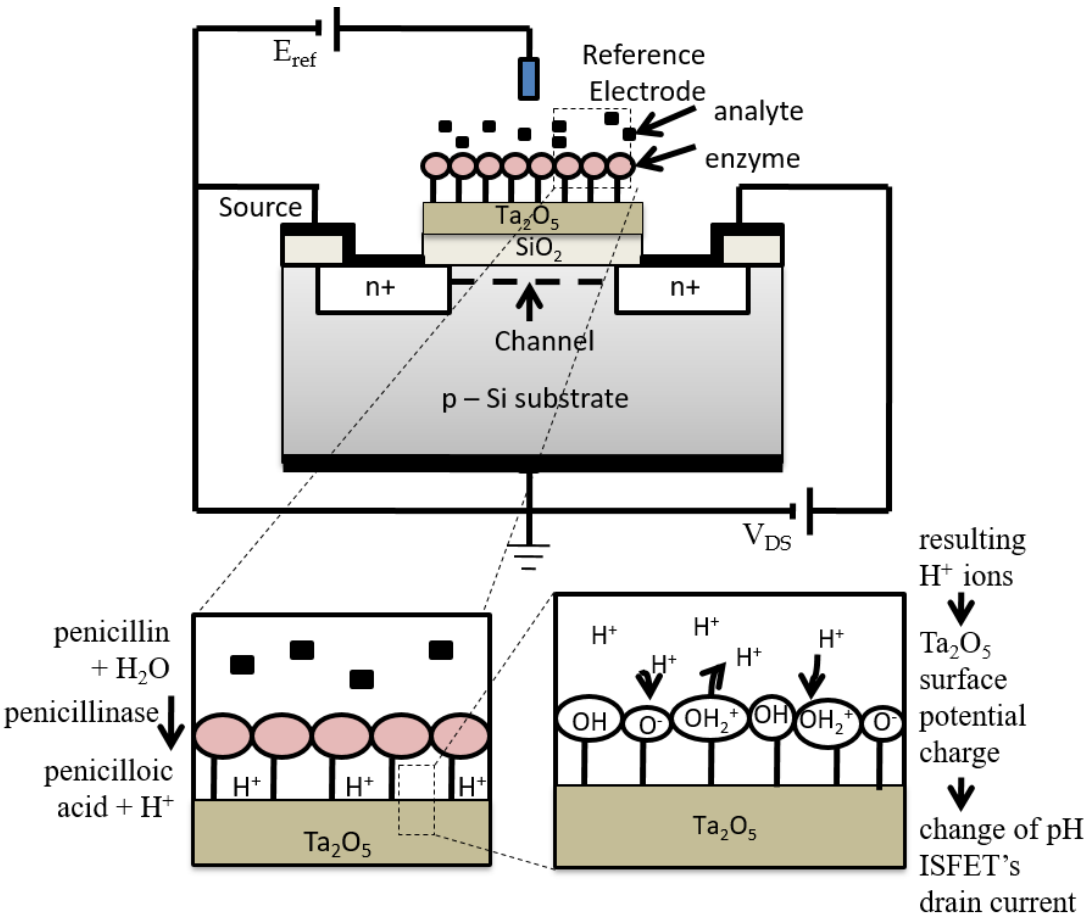

Fig. 2.3. PENFET (penicillin sensitive ENFET) structure and functional principle [86]

2.7. Fabrication Technology

2.7.1. Conventional MOSFET fabrication technology

The conventional MOSFET fabrication technology follows a list of complex processes such as wafer fabrication, oxidation, mask generation, photolithography, diffusion, deposition, etc. [79]. A huge set of laboratory equipment is required for the processes involved in fabrication. Conventional MOSFETs are mostly made up of silicon. Si wafer is obtained by processing sand or SiO_2. Sand highly purified by reacting with carbon and using epitaxial growth process for crystallization. This purified Si is kept in molten state and doped with desired impurity material. A crystal seed of Si is kept in a

Table 2.2: The enzymes used as bioreceptors and analyte used for detection

Sl. No.	Analyte	Enzymes used as Bioreceptors
1.	Urea	Urease, Penicillin G acylase, Penicillin penicillinase
2.	Glucose	Glucose dehydrogenase, Glucose oxidase/MnO_2 powder, Glucose oxidase
3.	Sucrose	Invertase/glucose dehydrogenase, Invertase/glucose oxidase/mutarotase
4.	Ethanol	Aldehyde dehydrogenase, Alcohol dehydrogenase
5.	Maltose	Maltase, Glucose dehydrogenase
6.	Lactose	β-Galactosidase, glucose dehydrogenase
7.	Ascorbic acid	Peroxidase
8.	Acetylcholine	Acetylcholinesterase
9.	Creatinine	Creatinine deiminase
10.	Formaldehyde	Alcohol oxidase
11.	Fluorine containing organophosphates	Organophophorus acid anhydrolase
12.	Organophosphate compound (paraoxon)	Organophosphate hydrolase

setup and heated to obtain a Si ingot, which can be cut in the dimensions required. On top of the wafer, SiO_2 layer is deposited by using wet oxidation technique in a high temperature furnace. Some regions of SiO_2 layer are to be removed to make the transistor. For that masking (generally electron beam masking) is done and then by using photolithography, the surface of the wafer is exposed is the regions where diffusion and channel formation is to be done. Another thin SiO_2 layer is deposited on the wafer followed by deposition of poly-silicon layer using chemical vapor deposition (CVD) technique. Then, by using another mask and photolithography, the poly-gate structures and interconnections are formed. The thin layer of SiO_2 is removed where source and drain regions are to be formed and diffusion is carried out. Again, a thick

layer of SiO_2 is deposited leaving the contact regions, where Aluminium metal is deposited by CVD process, masking and photolithography.

2.7.2. Conventional ISFET and ENFET fabrication technology

ISFET is formed by few modifications over MOSFET. If the metal gate of a MOSFET is replaced by electrolyte solution and a reference electrode, we get ISFET. So, the conventional ISFET fabrication technology follows the MOSFET fabrication steps excluding gate metallization. To make the ISFET pH sensitive, another insulator layer formed of materials like Si_3N_4, Al_2O_3, Ta_2O_5 (which are typically pH sensitive materials) is deposited on top of SiO_2 layer [55]. The gate of ISFET is now exposed to electrolyte solution. The pH variations in the electrolyte serves as input for the device.

In case of ENFET, an enzyme sensing layer (e.g. PVA [15]) is deposited on the gate oxide layer using techniques such as dip-coating or spin-coating, photolithography, etc. This sensing layer is required to hold the enzymes. The enzymes are then immobilized on to the enzyme sensing layer. In the method of immobilization, the enzymes are fixed on solid supports so that they do not move about in the electrolyte and retain their activity for longer duration. There are many different methods of immobilization such as physical adsorption [3, 88], entrapment [32, 89], covalent binding [99] and cross-linking [92].

2.7.3. ISFET and ENFET fabrication using chemical solution process

The conventional technique used for ISFET fabrication requires huge set of laboratory equipment making the process complex and expensive. To minimize the problems of traditional fabrication technology, the chemical solution process can be used [6-8]. This process uses the electrochemical deposition (ECD) technique for deposition of different layers of ISFET from bottom to top offering numerous advantages such as easy detection, low cost, low–power requirements, inherent miniaturization, portability, and high degree of compatibility with advanced micromachining technologies

The setup for deposition is a three-electrode system as shown in Fig. 2.4**,** which consists of counter electrode (CE), reference electrode (RE) and working electrode (WE). A voltage is applied between the WE and CE. The potential difference between RE and WE is fixed to a specified value and monitored. The role of RE is to act as reference in measuring and controlling the working electrode's potential and at no point does it pass any current. The CE passes all the current needed to balance the current observed at the WE. The deposition occurs on the WE.

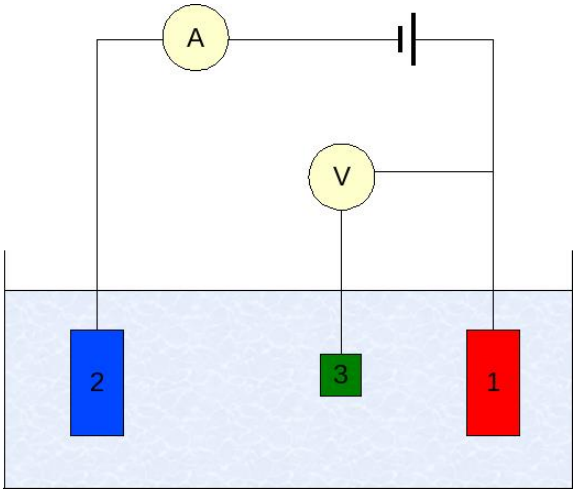

Fig. 2.4. Basic setup for three-electrode system (1) working electrode; (2) counter electrode; (3) reference electrode

2.8. Theories and measurement techniques used in ENFET fabrication

2.8.1. Metal semiconductor junctions and fermi level depinning

While forming junctions between metal and semiconductor, their work functions plays a major role. Depending on whether the work function of metal (ϕ_m) is greater or lesser than the work function of semiconductor (ϕ_s), two types of junction are formed.

a. Schottky junction

A Schottky junction or contact is formed when the barrier height between a metal and semiconductor is very large. Also, it is formed when the doping concentration is very

low, even less than the density of states in the conduction band or valence band. In Schottky contacts, $\phi_m > \phi_s$ for n-type semiconductor and $\phi_m < \phi_s$ for p-type semiconductor. Such contacts are known as rectifying contacts because it conducts when it is forward biased and acts as insulator in reverse biased state. There are different conduction mechanisms in Schottky contacts depending on the level of doping i.e. thermionic emission, thermionic-field emission, field emission and leakage current.

b. Ohmic junction

The Ohmic junction is the reverse of Schottky junction. It is formed when the barrier height between a metal and semiconductor is very small. Electrons from conduction band can move very easily to the metal and vice versa by tunneling. Here, $\phi_m < \phi_s$ for n-type semiconductor and $\phi_m > \phi_s$ for p-type semiconductor.

The ohmic junction behaves like a resistor and the semiconductor in both forward and reverse bias regions gives its resistance. Therefore, this junction conducts in both the regions. The metal-semiconductor junctions having average barrier height are known as quasi-ohmic contacts. Such contacts can be made ohmic by increasing doping level of the semiconductor. By doping the fermi level of the metal and semiconductor can be matched. This method is known as fermi level depinning [62].

2.8.2. Gravimetric Analysis

The thickness of the films can be measured by using gravimetric analysis, which is the method used in analytical chemistry for the quantitative determination of an analyte based on the mass of a solid. This method provides for exceedingly precise analysis. It provides very little instrumental error and does not require a series of standards for calculation of an unknown. This method does not require expensive equipment. Gravimetric analysis method has high degree of accuracy, when performed correctly. To measure thickness (δ) of thin film, the mathematical equation shown below can be used [42, 93].

$$\delta = \frac{M_s}{A_s \times \rho_s} \tag{2.14}$$

where, M_s is the mass and ρ_s is the density a particular substance, A_s is the area of the matrix on which thin film is deposited. Mass is be measured using microbalance.

2.8.3. Michaelis–Menten constant

In enzyme kinetics, Michaelis–Menten kinetics is one of the best–known models. It's name was given by Leonor Michaelis and Maud Menten. The model describes rate of enzymatic reactions of a substrate. Thus, rate of velocity is given by Eq. (2.15) [83].

$$v = v_{max} \frac{[S]}{[S] + K_M} \tag{2.15}$$

This equation is called Michaelis–Menten equation. Here, v_{max} represents the maximum rate achieved by the system at maximum saturation of the substrate concentration ($[S]$). The K_M is known as Michaelis Menten constant, which is the substrate concentration at which the reaction rate is half of v_{max}. It tells about activity of an enzyme. Low value of K_M means the enzyme has high activity and high value of K_M means the enzyme has low activity. K_M can be calculated by different ways such as using Lineweaver–Burk Plot [28]. It is the plot of reciprocal of concentration of analytes and reciprocal of response as shown in Fig. 2.5.

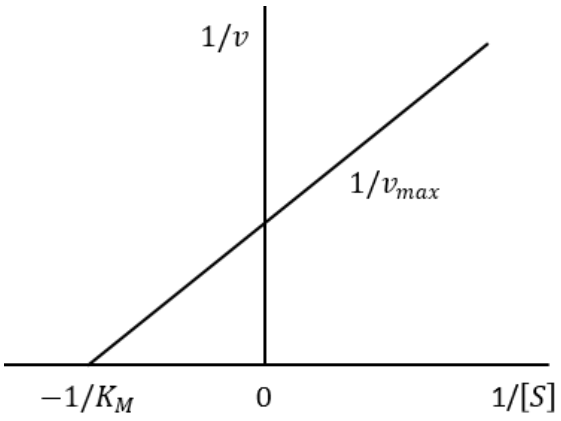

Fig. 2.5. Lineweaver-Burk Plot

2.8.4. Extrapolation in Linear Region (ELR) technique for determination of threshold voltage

There are many methods of determination of MOSFET threshold voltage such as constant current (CC) method, ELR method, transconductance extrapolation method in the linear region (GMLE), etc. [78]. The ELR method is mostly used for threshold voltage extraction. It consists of finding the gate-voltage axis intercept (i.e. $I_D = 0$) of the linear extrapolation of the $I_D - V_G$ curve at its maximum first derivative (slope) point (i.e. the point of maximum transconductance, g_m), as illustrated in Fig. 2.6. The value of V_T is given by the gate voltage axis intercept, which for the device at hand happens to be 0.51 V.

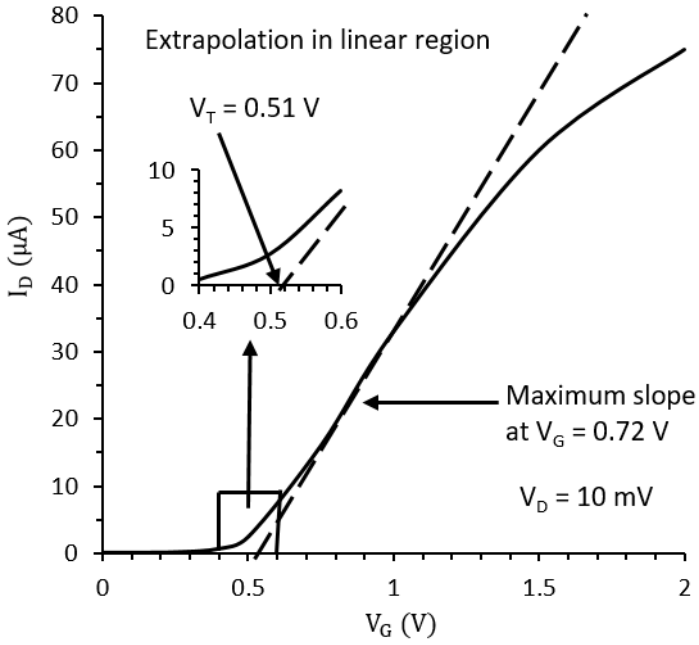

Fig. 2.6. $I_D - V_G$ curve showing ELR technique [78]

2.9. An overview on ISFET and ENFET models

Massobrio *et al*. developed the first ISFET SPICE model in 1991 [71]. They developed the model of ISFET and implemented in SPICE (Simulation Program with Integrated Circuit Emphasis). The first SPICE 1 was released in 1973 by Laurence Nagel and it was written in Fortran. SPICE is an open source analog electronic circuit simulator,

which is used for electronic circuit simulation and analysis. This program is utilized for board-level or IC design in order to check the circuit integration and also to predict the circuit behavior.

They inserted new parameters to the .MODEL card of the SPICE software for the simulation of ISFET behavior. The ISFET is similar to MOSFET except the gate material is replaced by electrolyte and reference electrode. Because of this some parameters of the SPICE model needs to be changed. The changes were found out on the basis of site binding theory [104] and Gouy- Chapman-Stern theory [14]. The equations affecting the change in parameters were inserted in the MOSFET subroutine of SPICE. This results in the introduction of many new parameters in the .MODEL card. Using the developed physico-chemical model in SPICE, the H^+ sensitive ISFET has been described. The model is suitable for analysis of the behavior of ISFET on the basis of fundamental physical parameters like binding site density and temperature, but for electrochemical analysis much more parameters need to be added. This model shows the efficiency of the use of physical models over circuital models. Another study was made to analyze the effect of bias potential on the pH sensitivity of the ISFET [69]. It was found that the influence of bias potential on pH sensitivity of the device is quite versatile. Also, it was seen that by modifying the set of equations describing the model, more complex models can be made at the physical level.

Another group of people tried to simulate an ISFET with oxynitride gate insulator on the basis of site binding theory [73]. They found that pH sensitivity of the device that depends on the surface composition and surface charge density can be calculated with a good efficiency. It showed that the ISFET works as a good pH sensor. They also tried to detect protein charges to check the performance of the device as immune-sensor but the response was worse.

It is known that ISFET's pH sensitivity is due to the proton interactions at its gate surface. van Hal *et al*. [101] tried to describe this sensitivity factor in a different way by forming a new and simplified model. This model was based on the differential capacitance and intrinsic buffer capacity of the device as the fundamental factors [100]. Results showed exclusion of the chemical and charge related factors. The general sensitivity expression was based on the Gouy-Chapman-Stern and site binding theory as usual. Simulation of the model was done for three different oxides SiO_2, Al_2O_3 and

Ta_2O_5. The intrinsic buffer capacity and differential capacitance was calculated for all the different oxides. Large variations were seen in intrinsic buffer capacity among the oxides but the variations in differential capacitance was very less specifically closer to the point of zero charge. Hence, intrinsic buffer capacity was found to be the major factor affecting the sensitivity of the device. To get high sensitivity approximately closer to the maximum theoretical value (59.2 mV pH^{-1}), a high intrinsic buffer capacity is required which can be obtained with higher number of surface sites or smaller ($pK_a - pK_b$) value. This high value of intrinsic buffer capacity also reduces the effect of electrolyte concentration on the ISFET's surface potential. Considering these factors, the relative sensitivity of Ta_2O_5 was found to be the highest with negligible influence of electrolyte concentration.

The ISFET model simulated in SPICE earlier was basically a physical model. Daniel *et al.* [21] tried to develop, a SPICE compatible electrochemical model of ISFET. They tried to modify some parts of the work done by Grattarola *et al.* [34, 35] to full their objective. They incorporated a variable carrier mobility model along with a modified electrolyte double layer model with more advancements. This helped in development of a more accurate model for ISFET. The set of equation used by this model revolved round different phenomena of the device i.e. chemical, electrical and thermal. The model appeared to be a bit complex but the number of parameters used in it were quite reduced. It was found that the temperature had much influence on the ISFET response which could be controlled by selection of certain factors like, gate dielectric material, substrate doping or device operating point to some extent [45].

The earlier ISFET models developed in SPICE version 2G had certain drawbacks though some BIOSPICE built-in models were developed. Whenever any new model needs to be made or some further modifications needs to be done in those models, it was a rigorous task. Because the whole program source files should be available, a good knowledge of the programming language used is required and also, the whole program should be compiled again and again. So, looking at the drawbacks, the researchers tried to develop a generalized model which is more user-friendly. Martinoia and Massobrio [87] developed a behavioral macromodel which is compatible with the commercially available HSPICE [41]. HSPICE was used because it has variety of MOSFET models available as compared to other SPICE versions such as PSPICE or SPICE3. But it can

also be used with the latter versions. In this model, two different parts were formed which were fully uncoupled: electronic part and electrochemical part. The electronic part was formed of normal MOSFET that is the basic part of ISFET and the electrochemical part was formed of the electrolyte-insulator interface. The electrochemical part consisted of the diffuse layer capacitance, Helmholtz capacitance and electrolyte-insulator interface potential. Later both the independent sections were combined to get the behavioral macromodel implemented in HSPICE. The whole model was implemented as a sub-circuit block in HSPICE, which can be easily accessed by user.

After incorporating ISFET macromodels in SPICE, researchers tried to develop CAD based microsystems with the ISFET models. This helped in testing the possibility of on-chip integration of circuits formed of chemically sensitive devices. Different types of microsystems were formed using ISFET. One of such kind consisted of three sections i.e. the sensor part, signal conditioning part and the testing part [68]. The pH sensitive ISFET was present in the sensor part of the system along with two MOSFETs of same structure. The signal conditioning part had a current source and current mirror, which helped in forming a differential stage between the ISFET and MOSFETs. The system also had three different output stages on the same chip. The testing part was formed of 47 different test structures like resistors, capacitors, diodes, MOSFETs, etc. This formed the platform for parameter extraction and technological evaluation for SPICE so that the device can be fabricated. Another type of microsystem was formed of cell based ISFETs [70]. It was formed of 12 ISFETs, 2 temperature sensitive devices and 1 conductivity sensor. It was designed to monitor the metabolic activities of cellular populations. Later, multi domain ISFET model was developed for simulations at device, circuit or system level to give more accurate models [46].

Another model was developed for multi-floating gate pH sensor using the Gauss theorem and charge neutrality concept [107]. Due to charge trapped in the multi-floating gate, the threshold voltage of the devices was found to be very high. In this model, the threshold voltage was reduced to a great extent by removing the charge on multi-floating gate. The results for different solutions showed correctness of the model towards solving the large threshold voltage problem.

Most of the ISFET models developed earlier were of Si based devices. The Si based ISFET had lots of advantages and disadvantages too. Though the use of Si based ISFETs paved the path towards chip integrability but when it came to nanoscale devices, lots of drawbacks were seen with these devices such as: technology used for fabrication is IC technology which requires huge number of sophisticated instruments and therefore cannot be easily set up in ordinary laboratory; have several junctions resulting into large internal contact resistance; high threshold voltage; small on-off current ratio; low sensitivity and scaling limitations for which it can't be easily pushed to nanotechnology domain. Then, Si nanowires came which had many improvements. SiNW based ISFET models both single [77] and dual [33] gated were developed using various modeling techniques. Dual gated models had much improved sensitivity [58]. Also, graphene based ISFET models were developed [57].

Then, came the carbon nanotube (CNT) based ISFETs with much improvement over Si based devices and lots of advantages such as high carrier mobility (due to ballistic transport), high chemical stability and robustness, high ON/OFF current ratios, steep switching and compatibility with high dielectric materials facilitating device miniaturization. An analytical model for CNT based glucose biosensor was developed with improved performance [81]. Also, a numerical simulation tool [1] for nanoscale ISFET is also developed by merging nanoscale ballistic MOSFET analytical equations with the Gouy–Chapman–Stern model equations of ISFET to form a system of nonlinear equations that can be solved iteratively to yield ISFET output current. The tool has its use to optimize the sensitivity and linearity of nanoscale ISFETs, and study their dependence on reference voltage, drain current level and gate-insulator thickness.

An ISFET having an enzyme layer immobilized on its pH sensitive surface is called enzyme FET (ENFET). Modeling of ENFET's is done on the basis of ISFET's pH detection principle in addition with enzymatic reactions involved in the process. Very few Si based ENFET electrochemical models are found in literature [5, 15, 16]. However, high-κ dielectric based graphene and CNT ENFET electrochemical models are not found till date.

2.10. Theories used for ISFET and ENFET modeling

2.10.1. Site binding theory and electrical double layer theory

The well-known site binding theory was given by Yates *et al.* in 1973 [104]. It explains the changes in the surface charge that occurs at the oxide-electrolyte interface. This theory was later generalized by Fung *et al.* [31] in 1986 for ISFETs. This theory says that the oxide surface contains binding sites which should be amphoteric. These amphoteric binding sites are formed of hydroxyl (OH) groups, which can accept or donate ions. The protonation or deprotonation at the binding sites depends on the concentration of hydrogen ions present in the electrolyte solution. The oxides such as SiO_2, Al_2O_3, etc. have only one type of site present, i.e. $A\text{-}OH$, where A represents Si, Al, etc. As per colloid chemical studies, the site should be amphoteric. Therefore, it is assumed that the oxide surface has three different sites, i.e. $A\text{-}OH$, $A\text{-}O^-$ and $A\text{-}OH_2^+$. The ionization reactions for SiO_2 are:

$$SiOH \rightleftharpoons SiO^- + H^+ \tag{2.16}$$

$$SiOH + H^+ \rightleftharpoons SiOH_2^+ \tag{2.17}$$

From the ionization reactions, it is understood that protonation or deprotonation at the oxide surface leads to positively or negatively charged surface. The resultant surface charge depends on the number of positive and negative sites available. The charged site, which dominates the other gives the charge of the surface depending on pH of the electrolyte solution. The H^+ and OH^- ions are known as potential determining ions for the oxide-electrolyte interface. Apart from these, more ions are present in the electrolyte i.e. cations and anions, which are referred as electrolyte ions. These cations and anions have a tendency to combine with the oppositely charged surface sites. The cations forms pair with the negatively charged site and the anions with the positively charged sites forming physical bond resulting in ion pairs as shown in Fig. 2.7. This process is known as surface complexation. It results in variation in the pH and charge at the oxide surface.

Many different types of ions are found in the electrolyte and near the oxide surface. The distribution of the ions in the electrolyte and in contact with oxide surface is given by Gouy-Chapman-Stern theory. This theory is also known as double layer theory because the ions are distributed in two layers in the electrolyte. The two layers are:

a. Stern inner layer

This layer is further divided into two planes known as Inner Helmholtz Plane (IHP) and Outer Helmholtz Plane (OHP). The IHP is the one closer to the oxide surface. In this plane, the locus of the centers of those adsorbed electrolyte ions are found, which forms surface complexes at the charged oxide surface sites. In the electrolyte, some ions hold several water dipoles around them in circular area. These are known as hydrated ions. In the OHP, the locus of the centers of those hydrated ions are found which are closest to the oxide surface.

b. Diffuse layer

The region beyond the OHP, extending deep into the bulk electrolyte solution is called the diffuse layer. This layer consists of the remaining hydrated ions which behaves as ionic cloud. These ions are balanced by the uncompensated sites present at the oxide surface. The distribution of ions using the electrical double layer model has been shown in Fig. 2.7. The electrical double layer can be represented in terms of two capacitors connected in series with each other. Each layer behaves as a capacitor. C_H is the Helmholtz capacitance for the Stern inner layer and C_D is the diffuse layer capacitance.

2.10.2. Diffuse layer charge density using Gouy-Chapman theory

The relationship between the charge and potential at the interface is given by the Grahame equation. The charge density is related to the count of charged groups present at the interface. To derive this relationship, Gouy-Chapman model can be used. Electroneutrality is assumed at the charged interface i.e. the total charge present in the unit area of diffuse layer should be equal to charge of opposite polarity in the unit cross-section of the interface.

As per Poisson equation, the charge density function is given as

$$\rho(x) = -\epsilon_r \epsilon_0 \frac{d^2\psi}{dx^2} \tag{2.18}$$

If σ_d is the charge density of the diffuse layer extending from OHP $(x = d)$ to $x = \infty$, then

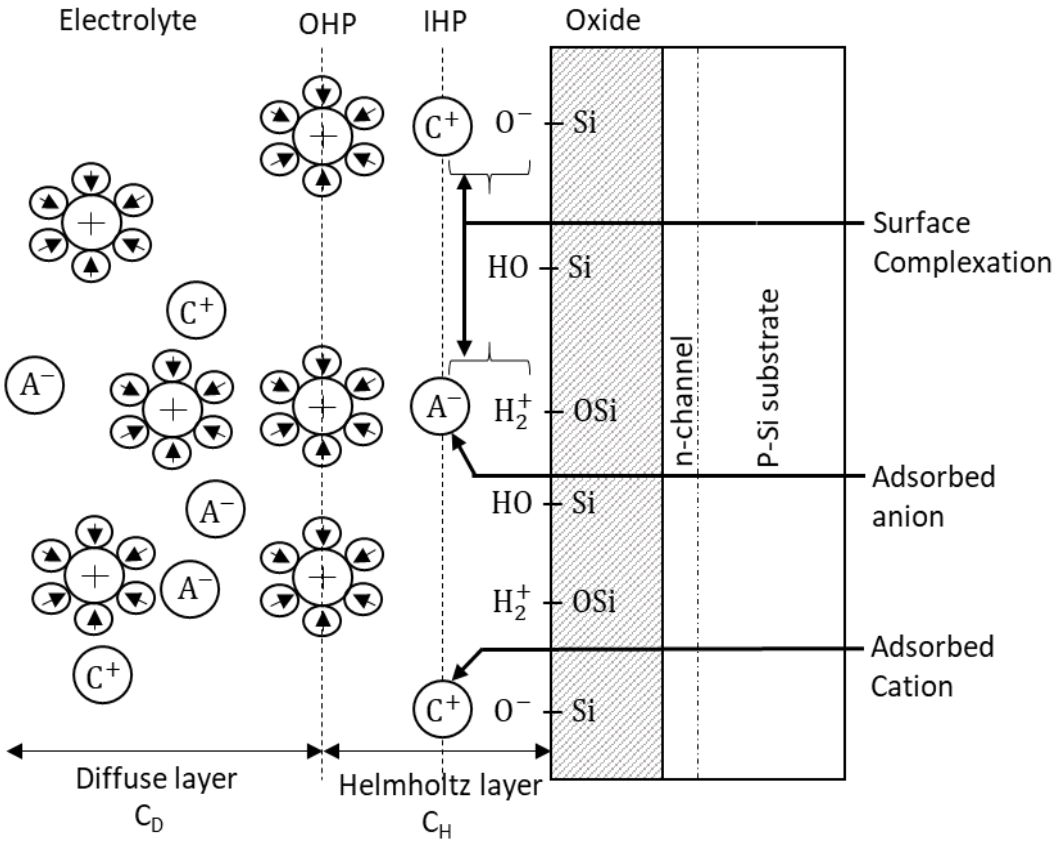

Fig. 2.7. Site binding and electrical double layer model of ISFET [27]

$$\sigma_d = -\int_d^\infty \rho dx = -\epsilon_r \epsilon_0 \int_d^\infty \left(\frac{d^2\psi}{dx^2}\right) dx$$

$$= \epsilon_r \epsilon_0 \left[\left.\frac{d\psi}{dx}\right|_\infty - \left.\frac{d\psi}{dx}\right|_d\right] = -\epsilon_r \epsilon_0 \left.\frac{d\psi}{dx}\right|_d \quad (2.19)$$

It is observed that, outside the OHP oppositely charged ions are accumulated. $e\psi$ amount of work is required to be done in order to bring an ion from ∞ to ψ potential region. The concentration of ions near the OHP is defined by the Boltzmann distribution, given as,

$$C_i = C_i^0 \exp\left(-\frac{z_i e\psi}{kT}\right) \quad (2.20)$$

The charge density ρ is related to the ion concentration, as given below,

$$\rho = \sum z_i e C_i = \sum_i z_i e C_i^0 \exp\left(-\frac{z_i e\psi}{kT}\right)$$

$$\Rightarrow -\epsilon_r \epsilon_0 \frac{d^2\psi}{dx^2} = e \sum_i z_i C_i^0 \exp\left(-\frac{z_i e\psi}{kT}\right)$$

$$\Rightarrow \frac{d^2\psi}{dx^2} = -\frac{e}{\epsilon_r \epsilon_0} \sum_i z_i C_i^0 \exp\left(-\frac{z_i e\psi}{kT}\right) \quad (2.21)$$

The above equation gives the Poisson-Boltzmann equation, whose solution gives the potential ψ at distance x.

Let us consider an auxiliary variable p, such that,

$$p = \frac{d\psi}{dx} \Rightarrow \frac{d^2\psi}{dx^2} = \frac{dp}{dx} = \frac{dp}{d\psi} \cdot \frac{d\psi}{dx} = p\frac{dp}{d\psi}$$

$$\int p dp = -\frac{e}{\epsilon_r \epsilon_0} \int \sum_i z_i C_i^0 \exp\left(-\frac{z_i e\psi}{kT}\right) d\psi$$

$$\Rightarrow \frac{p^2}{2} = \frac{kT}{\epsilon_r \epsilon_0} \sum_i C_i^0 \exp\left(-\frac{z_i e\psi}{kT}\right) + I \quad (2.22)$$

Using the boundary conditions, $x \to \infty, \psi = 0$ & $\frac{d\psi}{dx} = 0$, the constant of integration is calculated as

$$I = -\frac{2kT}{\epsilon_r \epsilon_0} \sum_i C_i^0 \quad (2.23)$$

Using Eq. (2.23) in Eq. (2.22), we get,

$$\left(\frac{d\psi}{dx}\right)^2 = \frac{2kT}{\epsilon_r \epsilon_0} \sum_i C_i^0 \left[\exp\left(-\frac{z_i e\psi}{kT}\right) - 1\right] \quad (2.24)$$

For the systems having symmetric electrolyte solution, we get,

$$\sum_i C_i^0 \left[\exp\left(-\frac{z_i e\psi}{kT}\right) - 1\right] = C^0 \left[\exp\left(-\frac{ze\psi}{kT}\right) + \exp\left(\frac{ze\psi}{kT}\right) - 2\right] \quad (2.25)$$

Using the boundary condition $\psi = \psi_d \; at \; x = d$, in Eq. (2.25), we get,

$$\left[\exp\left(-\frac{ze\psi}{kT}\right) + \exp\left(\frac{ze\psi}{kT}\right) - 2\right] = \left[\exp\left(-\frac{ze\psi}{2kT}\right) - \exp\left(\frac{ze\psi}{2kT}\right)\right]^2 \tag{2.26}$$

$$\Rightarrow \left(\frac{d\psi}{dx}\right)^2 = \frac{2kTC^0}{\epsilon_r\epsilon_0}\left[\exp\left(-\frac{ze\psi}{2kT}\right) - \exp\left(\frac{ze\psi}{2kT}\right)\right]^2$$

$$\Rightarrow \left.\frac{d\psi}{dx}\right|_d = \left(\frac{2kTC^0}{\epsilon_r\epsilon_0}\right)^{\frac{1}{2}}\left[\exp\left(-\frac{ze\psi}{2kT}\right) - \exp\left(\frac{ze\psi}{2kT}\right)\right] \tag{2.27}$$

Now, Eq. (2.19) implies

$$\sigma_d = -\epsilon_r\epsilon_0 \left.\frac{d\psi}{dx}\right|_d = (2kT\epsilon_r\epsilon_0 C^0)^{\frac{1}{2}}\left[\exp\left(\frac{ze\psi}{2kT}\right) - \exp\left(-\frac{ze\psi}{2kT}\right)\right]$$

$$\sigma_d = (8kT\epsilon_r\epsilon_0 C^0)^{\frac{1}{2}} \sinh\left(\frac{ze\psi}{2kT}\right) = (8kT\epsilon_r\epsilon_0 C^0)^{\frac{1}{2}} \sinh\left(\frac{q\psi}{2kT}\right), [q = ze] \tag{2.28}$$

This is known as Grahame equation.

2.10.3. Bousse's Model

In the year 1983, Bousse developed a model for determination of the voltage drop at the insulator-electrolyte interface of an ISFET device, known as Bousse's Model [14]. According to this model, the potential at the insulator-electrolyte interface depends on the pH of the electrolyte. It can be determined by two parameters, the pH at the point of zero charge, and a sensitivity parameter. These parameters are obtained from the site-dissociation model of the insulator-electrolyte interface, combined with the Gouy-Chapmann-Stern theory of the electrical double layer at this interface.

The potential ψ_0 at the insulator-electrolyte interface arises due to the interaction of the ions in the electrolyte with the insulator surface. The oxides such as SiO_2, Al_2O_3, etc. have only one type of site present, i.e. $A\text{-}OH$, where A represents Si, Al, etc. As per colloid chemical studies, the site should be amphoteric. Therefore, it is assumed that the oxide surface has three different sites, i.e. $A\text{-}OH$, $A\text{-}O^-$ and $A\text{-}OH_2^+$. The acidic and

basic equilibrium constants for the neutral site $A\text{-}OH$ are given by K_a and K_b respectively as given by Eqs. (2.29) and (2.30).

$$A\text{-}OH \rightleftharpoons A\text{-}O^- + H_S^+, with\ K_a = \frac{[A\text{-}O^-][H^+]_S}{[A\text{-}OH]} \tag{2.29}$$

$$A\text{-}OH + H_S^+ \rightleftharpoons A\text{-}OH_2^+, with\ K_b = \frac{[A\text{-}OH_2^+]}{[A\text{-}OH][H^+]_S} \tag{2.30}$$

Following the Boltzmann statistics, the concentration of H^+ ions at the surface, $[H^+]_S$ is related to the bulk concentration, $[H^+]$ as given by Eq. (2.31).

$$[H^+]_S = [H^+]\exp\left(-\frac{q\psi_0}{kT}\right) \tag{2.31}$$

At the surface, the number of sites are conserved as given by Eq. (2.32).

$$N_S = [A\text{-}OH] + [A\text{-}O^-] + [A\text{-}OH_2^+] \tag{2.32}$$

Again, the surface charge per unit area is given by Eq. (2.33)

$$\sigma_0 = q([A\text{-}OH_2^+] - [A\text{-}O^-]) \tag{2.33}$$

Solving the Eqs. (2.31) to (2.33), Eq. (2.34) is obtained

$$[H^+] = \left(\frac{K_a}{K_b}\right)^{\frac{1}{2}} \exp(y_0) \frac{\left[\frac{\alpha_0}{\delta} + 1 + \left(\frac{\alpha_0}{\delta}\right)^2 (1-\delta^2)\right]^{\frac{1}{2}}}{1-\alpha_0} \tag{2.34}$$

where, $y_0 = \frac{q\psi_0}{kT}, \alpha_0 = \frac{\sigma_0}{qN_S}$ & $\delta = 2(K_aK_b)^{\frac{1}{2}}$

At the surface to be electrically neutral, when $y_0 = 0$ & $\alpha_0 = 0$, the H^+ ion concentration at the point of zero charge is given by Eq. (2.35).

$$[H^+] = \left(\frac{K_a}{K_b}\right)^{\frac{1}{2}} \tag{2.35}$$

The difference between the pH and pH at the point of zero charge, pH_{pzc} is obtained as

$$v = \ln[H^+] - \ln\left(\frac{K_a}{K_b}\right)^{\frac{1}{2}} = 2.303\left(pH_{pzc} - pH\right) \tag{2.36}$$

Taking logarithm on both sides of Eq. (2.37), we get,

$$\ln[H^+] = \ln\left(\frac{K_a}{K_b}\right)^{\frac{1}{2}} + \ln\exp(y_0) + \ln\left[\frac{\alpha_0}{\delta} + 1 + \left(\frac{\alpha_0}{\delta}\right)^2(1-\delta^2)\right]^{\frac{1}{2}} - \ln(1-\alpha_0) \tag{2.37}$$

Now, neglecting higher order terms by putting $1 - \delta^2 \approx 1$, we get,

$$v = y_0 + \sinh^{-1}\left(\frac{\alpha_0}{\delta}\right) - \ln(1-\alpha_0) \tag{2.38}$$

To find the relationship between ψ_0 and pH, the Gouy-Chapman-Stern theory is used as given by Eq. (2.39).

$$-\psi_0 = 2\left(\frac{kT}{q}\right)\sinh^{-1}\left[\frac{\sigma_d}{(8\epsilon_w kTC)^{\frac{1}{2}}}\right] + \frac{\sigma_d}{C_{stern}} \tag{2.39}$$

where, ϵ_w is water dielectric constant, C is the electrolyte concentration, σ_d is the diffuse layer charge and C_{stern} is the Stern capacitance.

$$\psi_0 = -\frac{\sigma_d}{C_{eq}} \tag{2.40}$$

with, $C_{eq}^{-1} = 2\left(\frac{kT}{q}\right)(8\epsilon_w kTC)^{-\frac{1}{2}} + C_{stern}^{-1}$

For an EIS structure,

$$\sigma_0 + \sigma_d = \Delta\sigma = -(Q_{tot} + Q_s) \tag{2.41}$$

where, Q_s is the substrate charge and Q_{tot} is the total insulator charge.

The last term in Eq. (2.38), $\ln(1-\alpha_0)$, is negligible w.r.t y_0 and can be neglected. Using Eqs. (2.40) and (2.41) in Eq. (2.38), we get,

$$v = y_0 + \sinh^{-1}\left(\frac{y_0}{\beta} + \frac{\Delta\sigma}{qN_S\delta}\right) \tag{2.42}$$

where, $\beta = \frac{q^2 N_S \delta}{C_{eq} kT}$, is a dimensionless sensitivity parameter.

Changing the variables $y_0' = y_0 + \frac{q\Delta\sigma}{kTC_{eq}}$ and $v' = v + \frac{q\Delta\sigma}{kTC_{eq}}$ in Eq. (2.42), we get

$$v' = y_0' + \sinh^{-1}\left(\frac{y_0'}{\beta}\right) \tag{2.43}$$

For small values of the argument, $\sinh^{-1}$ behaves like a linear function resulting in a slope $\frac{\beta}{\beta+1}$.

$$\therefore v = y_0 + \frac{y_0}{\beta} = y_0\left(\frac{\beta+1}{\beta}\right)$$

$$\Rightarrow 2.303\left(pH_{pzc} - pH\right) = \frac{q\psi_0}{kT}\left[\frac{\beta+1}{\beta}\right]$$

$$\Rightarrow \psi_0 = \frac{2.203kT}{q}\frac{\beta}{\beta+1}\left(pH_{pzc} - pH\right) \tag{2.44}$$

The above equation gives the Bousse's model.

2.10.4. Fick's Laws of Diffusion

Adolf Fick described the diffusion phenomena in the simplest way by giving the famous Fick's laws of diffusion in the year 1855 [30]. According to him, the molar flux due to diffusion is proportional to the concentration gradient and the rate of change of concentration at a point in space is proportional to the second derivative of concentration with space. Fick's first law was given under steady state and second law under non-steady state conditions.

Fick's first law of diffusion states that the local rate of transfer of solute per unit area per unit time is proportional to the concentration gradient of the solute.

$$J = -D\frac{\partial C(x,t)}{\partial x} \tag{2.45}$$

where, J is the rate of transfer of solute per unit area or the diffusion flux, C is the concentration of solute, x is the co-ordinate axis in the direction of the solute flow, t is the diffusion time and D is diffusivity. The negative sign on the right hand side of the Eq. (2.45) states that the matter flows in the direction of decreasing solute concentration i.e. gradient is negative.

Fick's second law of diffusion explains the variation in concentration of solute with time. Following the mass conservation law, if there is no source or sink present, then the change in concentration of solute with time must be the same as the local decrease of the diffusion flux, which is

$$\frac{\partial C(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x} \tag{2.46}$$

Substituting the value of J from Fick's 1st law we get

$$\frac{\partial C(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[D\frac{\partial C(x,t)}{\partial x}\right] \tag{2.47}$$

Under low solute concentration and a given temperature, the diffusivity can be considered as a constant and we get,

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2} \tag{2.48}$$

Units of D (Diffusion Coefficient) are cm^2/sec those of $C(x,t)$ atoms/cm^3

2.11. An overview on nanomaterials: Graphene and CNT

Nanomaterials are an emerging class of materials formed of nanoparticles having size ranging from 1 to 100 nm (at least in one dimension). Miniaturization of devices with enhanced performance led to the uplift of nanomaterials. Such materials occupied a good place in the modern technological market starting from huge automobile engineering to bioengineering applications.

The nanoparticles have a wide range of unique properties whether it be chemical, physical or electronic domains, which make them superior to the traditional materials.

Few important characteristics exhibited by nanomaterials that make them promising tools for biosensors are large surface-to-volume ratio, strong adsorption ability, high catalytic efficiency and high surface reaction activity.

2.11.1. Graphene

Graphene is a two-dimensional single carbon layer of the graphite structure. It is a semi-metal and a zero bandgap semiconductor. To induce a bandgap in graphene the lateral lattice symmetry should be broken by some structural or chemical modifications [38]. Structurally modifying graphene by using quantum confinement effect produces semiconducting graphene structures such graphene nano-ribbons (GNRs), graphene nanomesh (GNM) and graphene quantum dots (GQDs). Chemically the hexagonal lattice symmetry is broken and a bandgap is induced by replacing some carbon atoms with some other atom (e.g. N), doping or surface functionalization. The structure of N-doped and B-doped graphene has been shown in Fig. 2.8 and 2.9 respectively. Few other methods that are used are stacking of graphene in layers (e.g. AB stacking) and mechanically deforming graphene, say by applying strain. This tunable bandgap of graphene provides flexibility in designing electronic devices and a lot of scope for new research.

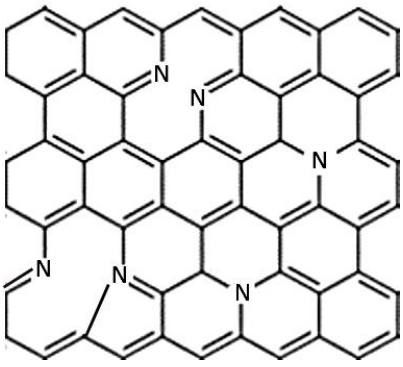

Fig. 2.8. N-doped graphene

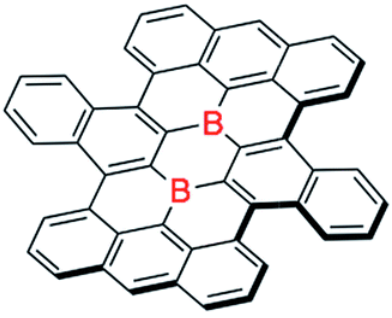

Fig. 2.9. B-doped graphene

Apart from these graphene have some excellent thermal, mechanical and electrical properties. It has a very high electron mobility of about 15,000 $cm^2 V^{-1} s^{-1}$ under

ambient conditions, which can increase up to 100,000 $cm^2 V^{-1} s^{-1}$ [13]. It given ballistic transport up to 0.3 μm at room temperature. Moreover, thermal conductivity is very high in graphene as low energy phonons are involved in heat transfer. The manufacturing cost of graphene is also very low, which makes it promising for industrial applications mostly in semiconductor industry.

2.11.2. Carbon Nanotubes

With the discovery of CNT in the soot of arc discharge at NEC by Iijima in 1991, researchers started analyzing its potential to enter the semiconductor market. CNTs have huge promising properties such as ballistic transport, very high potential for signal amplification, fast electron-transfer capabilities, high surface area-to-weight ratio, selectively bind with biomolecules after functionalization, high chemical stability and mechanical strength. Also, it changes conductance upon binding with charged macromolecules.

CNTs are formed by rolling the graphene sheets into cylindrical structures. The rolling of graphene sheets is done considering a chiral vector C_h corresponding to graphene vectors a_1 and a_2 with two integers (*n*, *m*) as shown in Fig. 2.10. The indices (*n*, *m*) determine the diameter and chirality. The (*n*, *0*) structure is called zigzag and the structure (*n* = *m*) is called armchair. Another type of CNT construction is called chiral where ($n > m > 0$). The chirality determines the mechanical, electronics and optical properties of CNTs. The diameter of CNTs varies from a few nanometers (in the case of SWCNT) to several tens of nanometers (in the case of MWCNT). The lengths of the CNT are usually found in the micrometer range. It has been reported that CNT has electron mobility ~10,000 cm^2/ Vs at room temperature for high carrier density [106].

Diameter (d) of CNTs can be calculated by Eq. (2.49)

$$d = \frac{a}{\pi}\sqrt{n^2 + nm + m^2} \tag{2.49}$$

where, a is the lattice distance ~2.46 nm.

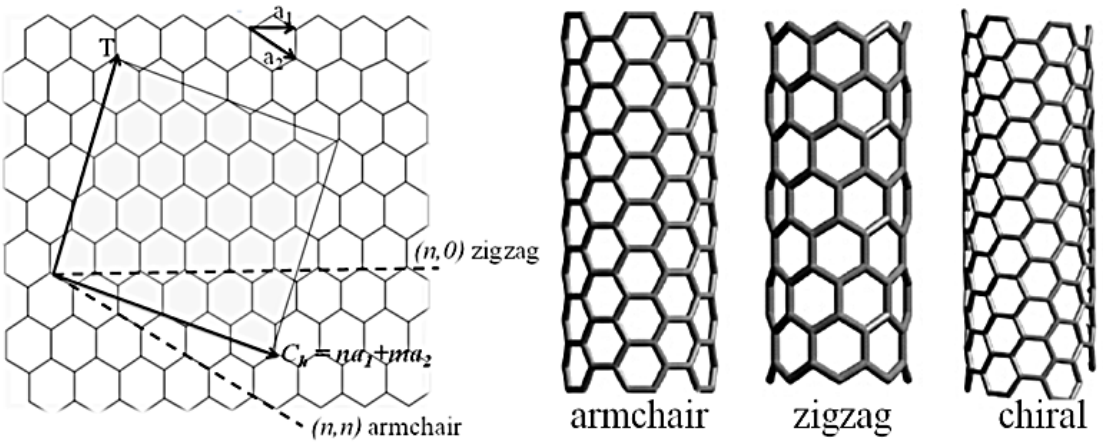

Fig. 2.10. Showing formation of different CNT structures using graphene sheets

Carbon nanotubes can be metallic or semiconductor depending on their chirality.

• If $(n - m)$ is divisible by 3, the tube is metallic.

• If $(n - m)$ is not divisible by 3, the tube is semiconducting.

The band gap of semiconducting CNTs depend on its diameter as shown in Eq. (2.50) with values ranging from ~ 0.5 eV (for large diameter of 1.5 nm) to ~10 meV (for small diameter).

$$E_g = \frac{2V_{pp\pi}a_{c-c}}{d} \tag{2.50}$$

where, $V_{pp\pi}$ is the C-C tight binding overlap energy (~ 2.5 eV) and a_{c-c} is the nearest neighbor distance between C-C bonds (~ 0.144 nm).

CNTs are attractive for nanoelectronic applications due to their excellent properties such as high electrical conductivity, high chemical and thermal stability and large surface area. In a nanotube, low bias transport can be nearly ballistic across distances of several hundred nanometers.

The CNTs are found to be p-type under ambient conditions with majority carriers being the holes. The holes are produced because under such conditions oxygen molecules are adsorbed on the CNTs by withdrawing electrons. So, CNTs become deficient of electrons making them p-type. To make the CNTs from p-type to n-type, it is doped

with alkali metals such as K, etc. In case of K doping, K releases the outer most electrons to CNT. This extra electron in CNT contributes charge and thus, CNT becomes n–type. The structure of K doped CNT has been shown in Fig. 2.11. The K ions are non-covalently bonded to the CNT. The problem with K doped CNTs is that they become unstable when exposed to air. So, they should be kept in vacuum or isolated from air. To reduce the instability with air contact, CNTs are doped with inorganic molecules such as NH_3, NO_2 and organic molecules such as amines. One such amine is polyethylenimine (PEI). Doping with PEI makes CNT n-type giving better and stable response. The structure of PEI-doped CNT has been shown in Fig. 2.12.

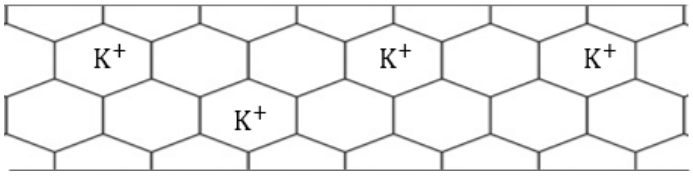

Fig. 2.11. K-doped CNT

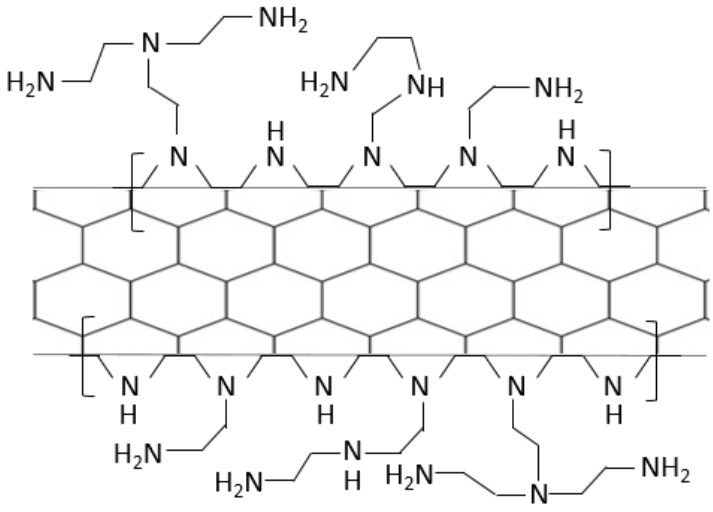

Fig. 2.12. PEI-doped CNT

2.12. Cholesterol and cholesterol oxidase

Cholesterol is a soft and waxy fat that our body needs to function properly [10, 103]. Almost every cell in our body consists of cholesterol. The body manufactures it by its own but it can also be taken externally in the form of food. It is oil-based and so does not mix with the water-based blood. So, it is carried around the body in the blood by lipoproteins. Two types of lipoprotein carry the parcels of cholesterol: low-density lipoprotein (LDL - cholesterol carried by this type is known as bad cholesterol) and high-density lipoprotein (HDL - cholesterol carried by this type is known as good cholesterol). Cholesterol has four main functions, without which one cannot survive. Those are that it contributes to the structure of cell walls, makes up digestive bile acids in the intestine, allows the body to produce vitamin D and enables the body to make certain hormones.

Cholesterol acts as both our friend and foe. At normal levels, it is an essential substance for the body's normal functioning, but if levels in the blood get too high, it becomes a silent danger that puts us at risk of a heart attack. It is a very important indicator of many diseases like myocardial infarction, hypertension, arteriosclerosis and dysfunctions in lipid metabolism [2]. Low levels of cholesterol does no good to the body. It may result in diseases such as hyperthyroidism, anaemia and mal adsorption. This makes cholesterol detection very significant. Its chemical formula is $C_{27}H_{46}O$ and its structure has been shown in Fig. 2.13.

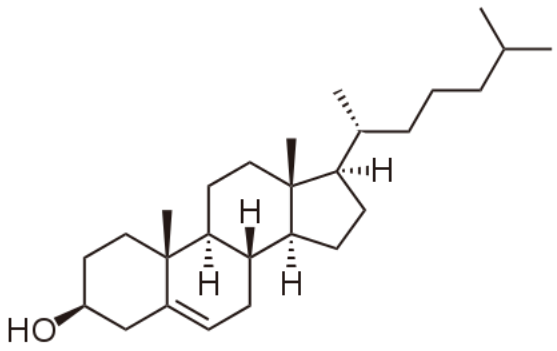

Fig. 2.13. Structure of cholesterol

3β–hydroxysteroid oxidase, popularly known as cholesterol oxidase (ChOx), is a bacterial enzyme or flavoenzyme that catalyzes the oxidation of cholesterol to cholest-4-en-3-one and H_2O_2. It is an extracellular enzyme that is found in various microorganisms, streptomyces, nocardia erythropolis, etc. It can be found in secreted form or may be associated with the surface of cell [59]. It is very specifically used for cholesterol detection in various biosensors.

2.13. Acetylcholine and acetylcholine esterase

Acetylcholine is an ester of acetic acid and choline found in the central and peripheral nervous system [40]. It acts as a neurotransmitter or messenger in the brain and body of animals and humans. A neurotransmitter is a chemical released by nerve cells to transmit signals to other cells. Its chemical formula is $CH_3CO–O(CH_2)_2N^+(CH_3)_3$ and its systematic name is *2–acetoxy–N,N,N–trimethylethanaminium*. The basic structure of acetylcholine has been shown in Fig. 2.14.

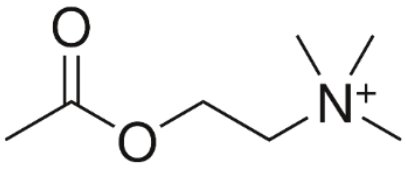

Fig. 2.14. Structure of acetylcholine

Acetylcholine has different functions in different parts of the body. In all autonomic ganglia of the body, it acts as the principal neurotransmitter. It also operates as neurotransmitter in the motor division of the somatic nervous system. Its activity produces an inhibitory effect in the body resulting in lowering the heart rate. Apart from this, it behaves as an excitatory neurotransmitter at neuromuscular junctions of the skeletal muscles. It performs the task of opening the ligand-gated sodium channels, which are present in cell membranes and then, places the sodium ions in the muscle cells. It also helps at the wake up time by enhancing the sensory perceptions and sustaining proper attention. The decrease in the levels of acetylcholine makes individuals prone to various nerve disorders such as Parkinson’s disease, Alzheimer

and multiple sclerosis [36]. People suffering from such diseases should take acetylcholine rich food such as milk and royal jelly.

Acetylcholinesterase is an enzyme, which converts the acetylcholine into choline and acetate. It is found in abundance in the synaptic cleft. Its role is to clear the free acetylcholine from the synapse, which is necessary for proper functioning of muscles. This enzyme gives a very speedy response. 1 molecule of acetylcholinesterase can hydrolyze about 10,000 molecules acetylcholine in 1 second.

2.14. Summary

In this chapter, the theoretical background of MOSFETs, ISFETs and ENFETs have been discussed with special emphasis on their mathematical modeling. Different ISFET and ENFET models available in literature have been shown with its advantages and disadvantages. The theoretical deductions of various models used for modeling such as Fick's law, Bousse's model, etc. have also been discussed in details. Moreover, the nanomaterials and biomolecules considered for modeling have been described with their significances.